The BIRO Consortium

Best Information through Regional Outcomes:

a Shared European Diabetes Information System for Policy and Practice

A public health project supported by the
European Commission

Perugia, Italy 2009

Best Information through Regional Outcomes:

a Shared European Diabetes Information System for Policy and Practice

A public health project supported by the
European Commission

Perugia, Italy 2009

This book includes a series of reports from the European project "Best Information through Regional Outcomes" (BIRO), co-funded by DG-SANCO, European Commission, 2005

A joint production of the BIRO Consortium:

Joanneum Research, Austria[1], Ministry of Health, Cyprus[2], NOKLUS, Norway[3]
Paulescu Institute, Romania[4], Serectrix snc, Italy[5], Telemedica Consulting, Romania[6],
University of Bergen, Norway[7], University of Dundee, Scotland[8]
University of Malta, Malta[9], University of Perugia, Italy[10]

Foreword

Nick Fahy, Head of the Health Information Unit, DG-SANCO, European Commission

Book Editor

Fabrizio Carinci [5,10], Technical Coordinator of the BIRO Project

Authors

Massimo Massi Benedetti[10], Scientific Coordinator of the BIRO Project
Joseph Azzopardi[9], Consultant Diabetologist
Valentina Baglioni[10], Software Engineer, Project Manager
Massimo Brillante[8], Software Programmer
Peter Beck[1], Software Engineer
Fabrizio Carinci [5,10], Senior Statistician
Scott Cunningham[8], Technical Consultant
Concetta Tania Di Iorio[5], Legal Expert
Andriana Evripidou[2], Statistician
Graham Leese[8], Consultant Diabetologist
Karianne Fjeld Looevaas[3,7], Registry Manager
George Olympios[2], Diabetologist
Pietro Palladino[10], Software Engineer
Philipp Perner[1], Software Engineer
Simion Pruna[4], Senior Scientific Researcher in Medical Informatics
Luca Rossi, Statistician
Svein Skeie[3,7], Consultant Diabetologist
Peter Taverner[3,7], Software Engineer
Vivie Traynor[2], Professional Nurse

Senior Consortium Representatives

Constantin Ionescu-Tirgoviste[4], Professor of Diabetic Medicine
Andrew Morris[8], Professor of Diabetic Medicine
Thomas Pieber[1], Professor of Medicine (Endocrinology)
Andreas Polynikis[2], Chief Medical Officer
Fausto Santeusanio[10], Professor of Diabetic Medicine
Sverre Sandberg[3], Professor

Acknowledgments

Annarita Ragni[10], Secretary, BIRO Coordinating Centre, University of Perugia, Italy
Amanda Adler, University of Oxford and University of Cambridge, United Kingdom
Xenios Anastassiades, Ministry of Health, Cyprus.
Douglas Boyle, A/Prof, University of Melbourne, Australia
Marco Orsini Federici, Clinical Expert, Medtronics, Italy
Ian Herbert, Consultant Health Informatician, United Kingdom
Ritchie Mac Alpine[8], Diabetes Regional Facilitator
Simon Scerri, University of Malta, Malta
Fred Storms, CBO, Consultant Diabetologist, Netherlands
David Whiting, Epidemiologist, IDF, Brussels, Belgium

Citation

The BIRO Consortium. Best Information through Regional Outcomes: a Shared European Diabetes Information System for Policy and Practice, University of Perugia, Perugia, 2009

Index

Foreword

We live in an information age, but good information is still scarce and hard to find. For the Commission, we work with a wide range of partners in order to generate and share relevant information that can help to guide policies to improve health at all levels in Europe.

Chronic conditions in general and diabetes in particular represent a challenge for good health in Europe that is already significant, and which we can expect to become greater in the years to come.

We know that action could be taken that would significantly reduce this burden, but that not enough is yet being done.

Good indicators to benchmark the problems we face and the steps being taken can be a powerful mechanism to help bring about improvements, as we have already seen in areas where these have been further developed at European level, such as cancer.

Our aim in the Health Information Unit of the European Commission is to provide information and analysis for evidence-based development, implementation and evaluation of action for health in the EU, at both Community level and within Member States.

European health information can provide added-value through information enabling comparisons in particular, which can in turn support identification, dissemination and application of best practice.

Providing such information is not easy. It requires a sustained effort across countries, across organisational and professional boundaries, and involving citizens and the wider community.

The BIRO project has worked to engage experts with different backgrounds (academic, policy making, clinical, engineering, statistical) to construct an innovative form of public health information system.

The solution that BIRO is producing is a valuable contribution to facing the health challenges of Europe. I welcome the contribution that BIRO has made, and offer my thanks to all those who have contributed to these results. I hope that continued work in this area can help us to build on these steps in order to help address these important European health challenges in the future.

Nick Fahy
Head of the Health Information Unit
Health and Consumers Directorate-General
European Commission

Introductory Remarks

Massimo Massi Benedetti

What is BIRO and the scope

The Best Information through Regional Outcomes (B.I.R.O.) Project is an initiative in line with the program issued by the European Commission to support the integration of efforts within and between current and prospective Member and Associated States (*Community action in the field of Public Health, Workplan 2003-2008)* responding to a call of the DG SANCO inviting projects to "*improve information to the public and formulate appropriate strategies, policies and actions*" and targeting "*appropriate sustainable coordination, in the area of health information…collection of data and information, comparability issues, exchange of data and information within and between Member States, continuing development of databases, analyses, and wider dissemination of information*".

Therefore the general objective of the project is to build a common European infrastructure for standardized information exchange in diabetes care, for the purpose of monitoring, updating and disseminating evidence on the application and clinical effectiveness of best practice guidelines on a regular basis.

Why B.I.R.O.

Diabetes is a progressively increasing heavy burden for the affected individuals and for the society as a whole due to short and long term negative effects on individuals' health and to their social and economical implications.

The now well known epidemic of the disease is related to a number of factors the most important of which are to be considered the relevant reduction of physical exercise and un-healthy nutrition characterising the model of evolution of the actual society which brings to obesity and as a consequence to diabetes.

It has been documented that such scenario in not restricted to specific countries or geographical areas, but has a global expression, interestingly enough, being in percentages more prominent in developing countries and in the most deprived social classes. The problem has reached such an extent that action has been taken by the most relevant international agencies like the WHO and the European Union to end with the United Nations which has produced a Resolution on Prevention of Diabetes and its complications approved by the UN General Assembly on December 2006 as a result of a worldwide awareness campaign led by the International Diabetes Federation in conjunction with a number of partners from international agencies and societies to national diabetes associations, Governments and industries.

The common target is to revert the Diabetes epidemic and to prevent the onset and the evolution of Diabetes complications in order to guarantee a full satisfaction of life expectancies for people with diabetes and to reduce the overall cost of the disease.

According to the widely accepted concept that *"it is not possible to manage what is not measurable",* relevant efforts are being made to evaluate the impact of diabetes in clinical, social and economic terms, and, given the global nature of the phenomenon, to develop strategies and methodologies for the production of globally useful international indicators.

The B.I.R.O. project has been designed for such purposes responding to a specific call of the European Union.

The roots of B.I.R.O.

Since the 1970s the alarm on the major relevance of non-communicable disease in respect to the prevailing attention given to the communicable diseases was launched by the scientific community and formalized in occasion of the WHO Conference in Alma Ata in 1978.

The perception of the need for action to be taken based on new and global models of intervention grew amongst a bunch of visionary members of the Diabetes community in Europe who joined forces with the WHO European Region and as a result "*Representatives of Governments Health Departments and patients organisations from all European countries met with diabetes experts under the aegis of the Regional Offices of the World Health Organisation and the International Diabetes Federation in St Vincent, Italy on October 10-12 1989"* giving birth to the so called St. Vincent Declaration (SVD).

The overall goals of the SVD for children and adults with diabetes were *"1)Sustained improvement in health experience and a life approaching normal expectations in quality and quantity, 2)Prevention and cure of diabetes and of its complications by intensifying research efforts".*

In order to reach the defined goals it was considered necessary to *"Establish monitoring and control systems using state of the art information technology (IT) for quality assurance of diabetes health care provision ... and to Promote European and international collaboration in programmes of diabetes research and development through national, regional and WHO agencies..."* The example of the SVD was then adopted in the following years in most parts of the world with the production of similar documents.

A quantity of initiatives flourished following up the SVD, not just because of it, but because the times were evidently mature for this to happen demonstrating that the European environment was the most advanced to feel the challenge.

The cross-fertilization of ideas and initiatives exploded involving national and international societies, national governments, the European Union institutions with a

relevant unrestricted support from diabetes related industries.

The DIABCARE initiative promoted by IDF Europe and the European region of WHO was the first attempt to aggregate diabetes care data for the production of international benchmarking based on locally produced information. It is to be considered that no internet existed at that time and the first attempt were based on advanced scanning technologies of fax transmitted information. However the indicators identified in the so called “Basic Information Sheet” do represent the platform for the actual evolution of the indicators selected afterwards. A strong support to such initiative was given by the EASD Study Group DOIT (Diabetes Optimisation through Information Technology) in a period when the now widely diffused e-technologies represented just hypothetical and uncertain evolutions. However topics like human interface, data protection confidentiality and security, data collection management, minimum indicator’s set, standardisation of procedures, interfacing, just to mention few, were deeply debated and heavily contributed to the development of the present systems.

The growing burden of diabetes progressively stimulated the attention of the European Union institutions also following a strong action of lobbying of the European diabetes community united under the flag of the SVD. A Diabetes Working Group of Members of the European Parliament (DMEPsWG) was institutionalised within the EU Parliament facilities to which IDF Europe served for the secretarial needs.

The joint action of IDF Europe, EASD and DMEPsWG heavily contributed to rise the attention of the EU institutions leading to the Parliament policy position in diabetes facilitating the investments of the EU Commission on diabetes research projects and diabetes care improvement initiatives avoiding that diabetes would be cancelled from the EU funding agenda as it had been planned.

A number of projects related to IT applications to diabetes care have been funded by the EU, the first, already on 1984, which was the EURODIABETA having the target to develop electronic medical records in diabetes care.

More projects were funded by the different EU institutions also as a result of the SVD awareness campaign, like the DIABCARE, DIABCARD, DIABSTYLE, EUDIP just to mention some of them. All these projects were definitely related to a number of activities promoted at national or regional level in many countries, Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Norway, Portugal, Rumania, Scotland, Spain, Sweden, The Netherlands, Turkey, UK. It is quite evident how the diabetes community has anticipated the EU in widening its frontiers. Something deriving from nearly all the EU funded and national projects can be found in the B.I.R.O. Project due to the direct contribution of the Partners of the Consortium and to the knowledge and ideas that have become a common and public European patrimony.

The already started EUBIROD Project whose Consortium reunite and represents a much wider number of partners in and outside Europe is the indicator of the usefulness of the methodologies developed within the B.I.R.O. Project and the guarantee that such a patrimony of ideas and knowledge will eventually represent a real advantage for the people with diabetes in the very Spirit of the St Vincent Declaration.

Some of the members of the Consortium have had the privilege to be part of this process since the beginning, they can be proud of it and it is their responsibility to transmit such Spirit to the newer generations.

The BIRO Project

Fabrizio Carinci

ABSTRACT

Introduction
The project “Best Information through Regional Outcomes” (BIRO) bears upon a systematic, evidence-based approach underpinned by well defined clinical guidelines widely used in diabetes. The scope was to improve management and prevention of diabetes complications in Europe through better information at all levels, fostered by the construction of a shared infrastructure for data processing and analysis. The proposal envisaged a new model for international benchmarking of quality and outcomes through a shared system delivering major indicators in a sustainable and automated fashion.

Objectives
To build a common European infrastructure for standardized information exchange in diabetes care, for the purpose of monitoring, updating and disseminating evidence on the application and clinical effectiveness of best practice guidelines on a regular basis. To implement targeted strategies, including: the adoption of a systems approach to make best use of different sources of information; the promotion of an efficient use of available resources, using systems in place at the regional level; and the implementation of technical solutions to build comprehensive reports on multiple outcomes.

Materials and Methods
A novel approach was defined through the adoption of definitions, including a particular notion of “region” in the BIRO framework and a target classification of BIRO users. The system data model was specified by means of a recursive structure that could be replicated across different national and sub-national levels. The BIRO process was split by design into “local” and “global” components distributing efforts among different sites, all contributing with structured data towards a central server. The workplan was designed to allocate different roles and responsibilities across the Consortium, integrated through strong involvement of all partners. A total of N=15 workpackages were in the program.

Results
The project took 40 months to be completed, with all workpackages delivered according to the plans. The clinical review specified parameters and indicators targeted by the system. A common dataset and data dictionary created schemas for the XML representation of all elements. A template was agreed for BIRO reports. Database tools were developed in Java to load local data formatted according to BIRO specifications in Postgres. The overall infrastructure was identified through a privacy impact assessment allowing specification of details for data transmission. A statistical engine has been developed using R and Latex to deliver local reports and aggregate tables. Secure protocols for data transmission have been developed through web services. Global statistical reports are made possible through a central engine that submits all results to an automated a web portal. A visual interface has been created to integrate all functions, with a setup program available to run BIRO on both Linux and Microsoft systems. Usability of the system has been tested in real conditions through technology transfer. Project website, newsletter, forum, reports, a series of meetings and seminars, and the present monograph were realised to disseminate results. All BIRO software has been released as open source under the GPL.

Discussion
Different users can take advantage from the availability of diabetes information for governance, research and health care. The deployment of a technological platform may help gathering data directly from sources to improve completeness of information, quality and outcomes of people with diabetes. Several unexpected results manifested the importance of genuine collaboration, well beyond the boundaries of statistical analysis, to act on diabetes in very practical terms.

Conclusions
The BIRO system can be now rolled out to a network of clinical units, regions, and Member States. Its development creates chances to operate in the direction of a European Diabetes Register, towards which the project has contributed with the realization of an innovative and compelling prototype. An expanding collaboration, under the banner of EUBIROD, will continue to operate to further develop the system and to apply it for the routine publication of diabetes reports.

1.2.1 Introduction

Health information is frequently hidden, fragmented, dispersed, under-utilised, poorly summarised, and undervalued. Few integrated approaches exist to fully exploit the advantages of linking and analysing health datasets straight from their point of collection, as a means to facilitate data exchange and information delivery.

Barriers to the use of health information for the rapid evaluation of health care include:
- insufficient utilization of information systems by clinicians and policy makers
- poor linkage between regional data sources and European statistical agencies
- limited application of sophisticated statistical routines in European health reports
- inadequacy of software available in the public domain
- insufficient use of medical records due to increasing privacy concerns
- lack of standardized approaches for secure data transmission

The project "Best Information through Regional Outcomes" (BIRO) aimed to resolve the above obstacles by using a systematic, evidence-based approach to guide the construction of a modern health information system.

In 2004, a visionary system was envisaged to link the existing knowledge base to specialised software, specifically tailored for routinely updating health reports to support the European strategy in diabetes[1,2].

The fundamental idea behind the project was that the various goals of health information were to be best realized through an alliance of regional initiatives that already have systems in place, or at least plan to implement new ones, to fulfil the same objectives.

A high quality network embraced the BIRO model, including renowned teams running diabetes registers in Scotland[3], Norway[4], Austria[5] and Italy[6] and a qualified group of partners from acceding and candidate countries (Malta[7], Cyprus, and Romania[8], now New Member States).

The main feature of BIRO was that it was specifically designed to share information routinely collected for clinical management. Its design was clearly based on the epidemiological results of ground-breaking studies e.g. the DCCT [9] and the UKPDS[10-12], which successfully identified the main risk factors existing in diabetes.

These studies clearly showed that concerted actions in prevention and in the organization of health services may reduce the risk of complications up to 30%. The BIRO proposal targeted better knowledge at the population level. By capturing data from different sources, a large proportion of subjects could be tracked, including people with diabetes that are not regularly followed up by diabetic clinics. The challenge was to rescue those "hard-to-reach", who frequently present most complicated (and costly) clinical conditions. A better, shared information system may also allow making economic analysis more comparable and consistent across Europe.

Reports continue to show that trends in diabetes-related expenditure increasingly threaten European health systems, taking into account demographic forecasts. Diabetes is the number one cause for admission to dialysis and a leading cause of blindness, accounting for 10 to 12% lower limb amputations, and 3-6% of total health care costs in Europe[13].

A substantial gap must be filled between what we know, and our capacity to know what we actually do. Despite of new evidence about the emergency of diabetes, systematic solutions informing professionals and the public in a timely manner are still lacking.

Modern reporting tools can go hand in hand with integrated disease management, for which better outcomes have been demonstrated on a solid scientific ground[14]. During the last years, researchers have used health information systems[15] and disease registers[16] to provide answers to specific questions, such as: which category of subjects needs more physician services? Are high-risk patients poorly served or do they have poor health outcomes despite being well served? Does high resources utilization represent overuse or is utilization related to high need?

Systematic solutions to follow up large cohorts of subjects with diabetes through the use of databases, preferably with a recognised, population-based denominator, have been generally referred to as "diabetes registers"[16].

Diabetes indicators e.g. those selected by the EUDIP project[2] constitute an essential element to assist different categories of users in evaluating the state of the art. Most indicators can be computed through the use of high quality registers collecting data on socio-economic factors, patterns of care, and a range of clinical/systems outcomes[17-19].

Electronic medical records can be linked to other sources, allowing a mix of strategies using methods from outcomes research[20-22], disease management[14] and health information systems[23-24].

BIRO addressed the opportunity to realize a relatively simple, unified protocol linking data and evidence. The project aimed to deliver a model for international benchmarking of quality and outcomes strongly advocated by previous proposals[16]. Through a shared system, it introduced an easier way to perform case-mix analysis[25-27], assessing variation in clinical practice[28], and evaluating adherence to clinical guidelines[29-31].

The general objective of the BIRO project was:

- *to build a common European infrastructure for standardized information exchange in diabetes care, for the purpose of monitoring, updating and disseminating evidence on the application and clinical effectiveness*

of best practice guidelines on a regular basis.

The strategy to realise it included the following points:

- *adoption of a systems approach to make best use of different sources of information*
- *promotion of an efficient use of available resources, using systems in place at the regional level*
- *implementation of technical solutions to build comprehensive reports on multiple outcomes*

Specific objectives of the project were:

- identification of a set of clinical guidelines based on the scientific literature
- selection of a European minimum dataset for international comparisons
- adoption of common set of indicators for routine monitoring of diabetes outcomes
- realization of a data dictionary to standardise definitions
- development of a targeted reports template
- design and implementation of a relational data model
- definition of the best architecture for privacy protection
- design and implementation of statistical methods for the production of health reports
- validation of a secure protocol for international communication and shared data analysis
- customisation and development of open source specialized software
- linkage of the different components in a user-friendly reporting facility
- dissemination of all results through a web portal.

This monograph presents the main achievements of the project, showing the various features of the system being realized and the steps followed for its development. It also provides general directions for the basic use of BIRO, indicating ideas for the future.

This chapter presents a summary of BIRO methods and results, providing appropriate pointers to all the other chapters of the monograph, where further details can be easily found.

1.2.2 Materials and Methods

The BIRO project targeted the realization of a "Shared Evidence-based Diabetes Information System" (SEDIS) as an efficient and sustainable solution to perform the following tasks:

- analysis of longitudinal trends and average outcomes in a diabetic population
- identification of patterns of care and prevention associated to positive outcomes
- identification of population strata and/or practices at risk of negative outcomes
- assessment of the level of adherence to best practice guidelines
- development and testing of a general model of shared information system

The plan for the BIRO submission started as early as March 2004, when a Consortium (see Box 1.2.1) was formed to apply to the European Commission for the "Call for Proposals Public Health - 2004 OJ 2004/ C52, 27/2/2004".
The proposal was favourably evaluated, and the grant agreement was signed by the Commission on 11/10/2005. More details on the application process can be found in Box 1.2.2.

The system was planned to be realized by means of:

- multidisciplinary collaboration
- protection of data ownership
- privacy protection
- shared information infrastructure
- high level database and statistical technology
- open source software

Briefly, the method implied a connection between databases maintained by regional networks through an agreed scheme not implying the exchange of individual records.

The construction of diabetes indicators was formulated as a result of a structured linkage between a common dictionary and aggregated tables contributed by members of the network, including various parameters and clinical characteristics[29-33].

Each of the specific objectives of the project was assigned a specific work package, for which a designated leader was given the goal of building a specific component. Integration was sought through strong involvement of partners in all aspects of the project, rather than a specific role.

Few assumptions had to be done at the beginning of the project to clarify some definitions regarding central aspects of the project.

The relation between a centre, region, or country partner of BIRO was one of the main topics subject of initial discussion.

It was agreed that "region" in the BIRO model should have been intended not as an administrative entity, but as a network of centres sharing a homogeneous set of organizational aspects, including the definition of individual data items and the way they measure and collect them.

BIRO sets its system at a higher level, defining common standards that do not imply necessarily a change in the way data is collected from the direct source. When the system would have been up and running, it would have been the responsibility of the regional level to map local definitions against the common format, produced independently.

As a result, the BIRO model can be applied recursively. Each country may have the need for national indicators that can be obtained in different ways. Some countries may be small enough to represent a single region, while others may be highly decentralised, and admit some heterogeneity in the definitions and the way data is gathered. In such situations (e.g. Italy or Spain) a common European standard may even facilitate national

Box 1.2.1 BIRO Consortium

Associated partners

Cyprus Ministry of Health (CYPRUS). The Ministry offers a growing experience in health promotion, with a particular interest in diabetes. Cyprus has recently emphasized its active cooperation at the European level, particularly with the International Diabetes Federation. For its geographical location, the partner can play a relevant in the defini-tion of best direction for technology transfer.

Joanneum Research (AUSTRIA). The Institute of Medical Technologies and Health Management at JRS combines health management experience with expertise in software engineering, offering services to public institutions, professional organisations, industry and science. The Institute aims to devise solutions to optimise health care delvery, and improving administrative and clinical processes. It develops applications for routine application in disease management, patient care and epidemiology.

Paulescu Institute (ROMANIA). The Institute of Diabetes "N. Paulescu" offers an important mix of experience in medical care and diabetes research. Paulescu is a public body offering multidisciplinary clinical care for the patients of all ages and with all type of diabetes. The institute performs basic and clinical research activities in the field of metabolic diseases. Major areas of interest are: neuro-electrophysiology, nutrition, genetics, epidemiology and immunology related to metabolic diseases.

University of Bergen (NORWAY). The Faculty of Medicine and Dentistry covers a broad spectrum of research fields in clinical medicine, biomedicine and health sciences, and has established research schools in experimental cancer studies, neurosciences, cardiology and circulation, epidemiology and vitamines, with focus on vitamine B and homocysteine. Highly competent research groups are also present in molecular biology, translational biologicall/clinical research and global health issues.

University of Malta (MALTA). The University of Malta offers an important experience in medical care and monitoring of diabetes in Malta, with active cooperation at the European level. It works closely with the Diabetes Clinic through which research into diabetes is coordinated. Research interests of the diabetes department include epidemiology, use of computers for diabetes research and management.

University of Perugia (ITALY). The Department of Internal Medicine of the Faculty of Medicine of the University of Perugia, with main interests in diabetes research and clinical activity, produces a large number of publications in the most important scientific journals, organizes series of international meetings, and coordinates diabetes health care in Regione Umbria. The more relevant areas of interest are pathophysiology, artificial systems for insulin delivery, IT and diabetes: quality development.

University of Dundee (UNITED KINGDOM). The UoD has been heavily involved in the creation and development of the Scottish SCI Diabetes Collaboration project, a system used by health care professionals across Scotland for the day-to-day management of patients with Diabetes. Considerable experience has been gained in creating automated links from Primary and Secondary care to provide a central repository that hosts the shared-electronic record for Diabetes in Scotland.

Supporting Institutions

Noklus (NORWAY). Noklus and the Norwegian Diabetes Registry for adults have a longstanding commitment in quality of care programmes and research. We have in depth interest in diabetes and we are establishing and running the Norwegian Diabetes Registry for adults. We have experience from earlier EU-projects like DiabCare (EU Framework IV and V) which are diabetes related and related to quality performance and care.

Regione Umbria (ITALY). Regione Umbria is located in Central Italy, has a surface of nearly 8,500 Km^2 and a population of 840.000. The health system is structured in 4 local health authorities, with nearly 140,000 hospital admissions per year. The Regional Health Care Administration has strongly supported the project as a springboard for the further development of the Regional Diabetic Register. The administration is very experienced in data linkage of routine health databases.

Subcontractors

Serectrix (ITALY). A consulting firm devoted to public health applications and evidence-based information networks. The company is highly specialised in the design and conduct of multidisciplinary collaborative projects for optimising the delivery of health services, particularly in chronic diseases. Specific areas of interest include: outcomes evaluation, privacy issues and health information systems.

Telemedica Consulting (ROMANIA). Telemedica Consulting is a software developer specialised in open source applications in the health sector. Team coordinator for the EU project "Black Sea TeleDiab (BSTD)", exhibited by DGXIII along with 10 others (chosen out of 180) at the Medinfo2001 exhibition in London. Telemedica is also the developer of SincroDiab, a software tool allowing central longitudinal management of diabetes episodes, shared across a range of providers in Romania.

Box 1.2.2 BIRO diary

The BIRO project was submitted to the "Call for Proposals Public Health – 2004 OJ 2004/ C52 , 27/2/2004, DG-SANCO, European Commission, Luxembourg" on 26/4/2004. The proposal no.790718 was favourably evaluated by the Commission Services, but ranked 2nd on the reserve list with decision taken on 8/7/2004, notified 22/7/2004. On 28/1/2005 the Coordinator was informed that additional funding had become available and negotiations could initiate. Partners of the European BIRO project met for the first time to discuss in person in Assisi (Italy) on 1/5/2005. Main topics were the compatibility of different registers and how they could talk each other through a common structure. The meeting allowed defining the content of work packages and assigning roles and responsibilities. The BIRO Consortium Agreement was signed by all partners on 11/10/2005. Negotiations successfully ended with co-financing of 714,675¤ (60% of total cost) officially granted with signature of grant agreement no.2004129, on 10/11/2005. The project officially started on 2/12/2005, ending date 1/12/2008. The kick off meeting took place in Perugia (Italy) on 5/12/2005. It allowed agreeing major steps to realize the objectives of the project. The first technical meeting was held on 6-7/3/2006 in Dundee (Scotland). A common XML dataset was drafted and basic information on EUCID indicators was acquired. The first BIRO investigator meeting took place in Malta, 30/5 to 2/6/2006. An initial draft of the clinical review was revised and an indicator short list was defined. The main categories of BIRO users were also specified, and the contents of the technology transfer agreed. Firm steps for the construction of the system were undertaken at the second technical meeting in Graz (Austria), 29/9 to 1/10/2006. A requirement analysis was conducted and major software specifications were sketched out. Different architectural alternatives were specified in relation to the initial progress of the privacy impact assessment. A third technical meeting was held again in Dundee on 26-27/3/2007, allowing to define a final draft of the BIRO XML Schema, the plan for the statistical engine and a structure of aggregate tables. A review of the privacy literature was also carried out. Colleagues from Cyprus introduced the successful development of the first diabetes register on the island, based upon BIRO recommendations. The favourable situation was examined in more detail during the 2nd BIRO Investigator Meeting held in Larnaca (Cyprus), 23-26/5/2007. The meeting was crucial to define the fundamental aspects of the project, supported by the Cyprus Diabetes Association. Partners from the Ministry of Health referred the strong impact of BIRO on local policy, with the start of the Cyprus Diabetes Register and the organization of the first diabetes clinic on the island. High-calibre evaluators Amanda Adler (Un.Cambridge/Oxford, UK) and Fred Storms (CBO, Netherlands) offered their valuable input to the Consortium. The data dictionary and the procedures for the creation of the exchange XML format and the BIRO database were agreed. The Privacy Impact Assessment Consensus Panel was successfully conducted through a special session that allowed definition of the best BIRO architecture. The reports template was finalised. Later on, an organizational meeting in Rome (Italy), 3/11/2007 led to the definition of the overall structure of the software, including communication software shown for the first time to the Consortium. A follow-up meeting held in Rome on 20/4/2008 provided a complete overview of the BIRO architecture, the meaning and definition of statistical objects, and the plan for the statistical engine. A structure for the web portal was also decided. Draft of an original paper on privacy impact assessment was presented. During 2008, a six months extension was requested to the Commission, finally approved to set new end date at 1/6/2009. A meeting in Brussels on 29/11/2008 allowed examining the state of the art and the major pitfalls/delays in the development of the software. The final "integration" meeting took place in Bergen (Norway), on 15-17/1/2009. Technical referents analysed all unresolved aspects in the development of the software, debugging it while loading different data sources and running its components. Deliverables were analysed to fit with the overall design, particularly data schemas. Different presentations of the results were considered, and the plan for the final delivery of the project agreed among partners. Furthermore, an e-learning platform was presented as a means to disseminate results more widely; contents of the web portal were agreed, and a plan for the final monograph was made. The Bergen meeting resulted to be the most practical session ever made, integrating all efforts towards the deployment of a compact product. The BIRO system was prepared for the BIRO Academy meeting in Kuwait City (Kuwait), 2-4/5/2009; the monograph organized for the official presentation to the European Commission, Brussels, 7/5/2009. The final delivery of the BIRO project has been scheduled in the final event at Perugia (Italy), 25/5/2009.

integration, proposing a standard that can be fostered by a national coordinator. A country may collect sets of standard tables from regions, which can have sub-regional levels (e.g. local health authorities), running BIRO down to the level of the single clinical unit.

The SEDIS system would then grow bottom-up as a whole 'decision support system' composed of many local networks, facilitating an 'informed decision making' at different levels of the health system.

To agree on a definitive methodology, the Consortium had to classify the different categories of users of the BIRO system. The classification shown in Box 1.2.3 was useful to define dissemination activities, and design reports and publications that could have been used accordingly.

The characteristic needs of the dissemination strategy are reflected by the design of the data model.

The SEDIS *data model* is divided into two parts: a static part, related to data collection (hardly changes over time) and a more dynamic part, related to medical concepts (more susceptible to changes in the medical knowledge). Isolating the dynamic part will largely eliminate the risk of frequent updates in the software.

The complete SEDIS data cycle is based on the application of two consecutive data processing steps. The fundamental aspect of the system is to ensure its basic functionalities at the level of each single register ("*local SEDIS*"). The model is then generalized through its repeated application in all registers, followed by an overall step that compiles all "partial" results into a

Box 1.2.3 Who are the BIRO users

Governance • European Union • Commission and Parliament • National and Regional Governments • Local Health Care Authorities, Management Clinical Networks • Other local authorities • Payers • Social/Private Insurance • Non Governmental Organizations • WHO, OECD, IDF, National and Regional Diabetes Associations	• National and international scientific organizations • Research institutions • Universities, Foundations • Statistical Departments of Local Governements • Research areas • Epidemiology, health policy, clinical medicine **Health Care** • Primary Care Societies • Diabetes Care Units • Health Care Professional Associations • Quality Management Associations
Research • EU Directorates Research and Public Health • Scientific Organizations	**Citizens** • Consumer organizations • Patients organizations

global report. The basic technical characteristics of the model have been described in an unpublished seminal research study[23] paving the way for its realization.

A central element of BIRO is a precise organization of all relevant parameters that are required for statistical analysis, e.g. normal levels of glycated haemoglobin, categorization of continuous variables, coding, etc. All definitions are stored in a *data dictionary* using a common format. A "progressive diary" can also include algorithms, e.g. the list of tests recommended to patients with hypertension, over 65, with a high level of glycated haemoglobin (*guidelines*), or a particular "severity score" (*comorbidity index*). The result is a "concept and data dictionary".

The dictionary in the context of a "local SEDIS" can be represented as a chain of steps logically intertwined (Figure 1.2.1). It allows comparing different analysis, both geographically and longitudinally, representing the evidence-based component in the model chain. It follows the definition of a minimum dataset and needs to be regularly updated.

At the opposite end of the chain is the final health system report. The content of the report is based on the initial specification of a template that influences the selection of data procedures and statistical methods ("*database engine*" and "*statistical engine*").The engines operate on top of the local databases that are not directly accessible by other partners. The reports are composed through the amalgamation of statistical "*objects*" (*tables, parameters, graphs*) that need to be produced by the joint application of the engines.

The definition of an overall model (global SEDIS) directly follows the local implementation (Figure 1.2.2). Once the statistical objects are available for each register, they can be exchanged across the network using a secure format.

The level of aggregation chosen for each object is a combination of: a) formal agreement – the parties will need to support the level of detail; b) legislation – privacy at all levels, including individuals and institutions, need to be preserved; c) practical limits – not too much data can be rapidly transmitted. Such conditions had to be tested in practice through the BIRO collaboration, by acknowledging the independent values and judgement across partners: an associated centre may not agree to transmit some objects across the network, but minimal requirements can be met and be supported by other partners. For this reason a legislative review has been foreseen in close collaboration, so that each partner can reach an informed decision. Eventually, different procedures to summarize different types of objects will be applied.

Once objects reach a central location (*server*), these can be submitted to the global database and statistical engines (*central engine*) to finally return a global health system report that is valid for the whole EU collaboration.

The BIRO *workplan* has been directly derived from the design of SEDIS described above. The duration of the project was set to 36 months, to realize different work packages (WPs) described below.

Coordination and project management
Activities related to the coordination and project management included WPs running for the entire duration of the project: "*coordination*", "*project management*", and "*evaluation*". The evaluation step was performed with the help of independent experts.

Building the knowledge repository
Definitions have been organized through different WPs. Through the "*clinical review*", a systematic review of the evidence had to be carried out to define summary parameters and indicators for diabetes care in Europe. The WP "*common dataset*" defined the standard format to be created to extract data from local registers. The "*data dictionary*" defined the structure for information to be shared across centres, in the form of XML meta-data.

Developing the database and statistical engines
The "*database engine*" was in charge of developing SQL-compliant standards on top of the meta-data formats identified by the concept dictionary. Descriptive

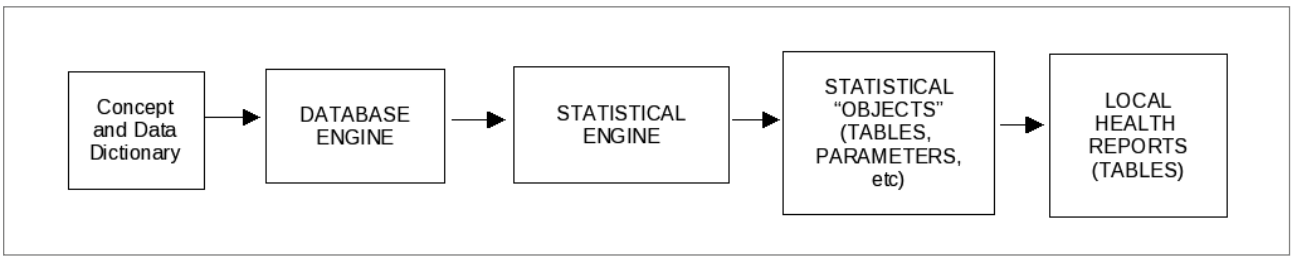


Figure 1.2.1. The BIRO data model for the "local SEDIS"

and multivariate models for case-mix adjustment have been designed for the "*statistical engine*", including two separate components: the local analyser and the meta-analyser. The latter must be implemented by the WP "*central engine*". Software had to be produced using various open source scripting and programming languages, available on both Windows and Linux.

Ensuring secure transmission and information exchange
The WP "*privacy impact assessment*"[35] had to identify the impact of privacy and data protection to reduce adverse effects of the introduction of new technologies. At the time of BIRO submission it was clear that "the diabetes community need(ed) guidance and a framework for safe practice, if the perceived benefits of these databases are to be realised"[36]. This specific WP was designed to ensure compliance with data protection law as a minimum acceptable practice, embedding strong rules by design in the realization of the system. Communication software for data transmission and exchange follows specifications advised by privacy assessment. These steps were to be implemented by the WP "*communication software*".

Dissemination of the results
The WP "*technology transfer*" aimed to develop the knowledge base for the use of end-users, particularly those from acceding and candidate countries (now New Member States).
The content of the health reports to be produced had to be set by the special WP "*reports template*", identifying the structure to deliver meaningful outputs for diabetes management.
Dissemination is insured to be the widest through the definition of an automated interface where BIRO reports can be published: the "*web portal*".
The WP "*dissemination*" had the task to translate the main achievements of the Consortium into various formats, including the website, newsletter, forum, web portal and the present monograph.

1.2.3 Results

The workplan of the project has been realized in 40 months, between 2nd December 2005 and 1st June 2009. A complete diary of activities can be found in Box 1.2.2. The initial duration required an extension: the workplan was completed in a timeframe of 40 months.

The first deliverable was the *Clinical review* resulting in a list of N=85 indicators, with a total of N=49 items selected as relevant for BIRO. For all indicators, the level of feasibility was taken into account to precisely specify data items that are required for computing such indicators. Further details on the WP, including the list of parameters/indicators, can be found in Chapter 2.1.

Several WPs allowed to specify the procedural flow of the BIRO Architecture through three consecutive steps, logically organized in two different parts: local and global (Figure 1.2.3).

The way WPs are associated to specific outputs is presented in Figure 1.2.4.

The *local* part of BIRO Architecture includes the set of software tools required by each collaborating centre to undertake two basic operations:

1) production of a standardized BIRO local report
2) data transmission to the BIRO server for the production of the global report.

Step 1 involves client data processing and statistical analysis.

A BIRO "*Adaptor"* has been realised to establish a connection to the local database and export data from any format used by the local diabetes register to the standardized format complying with specifications

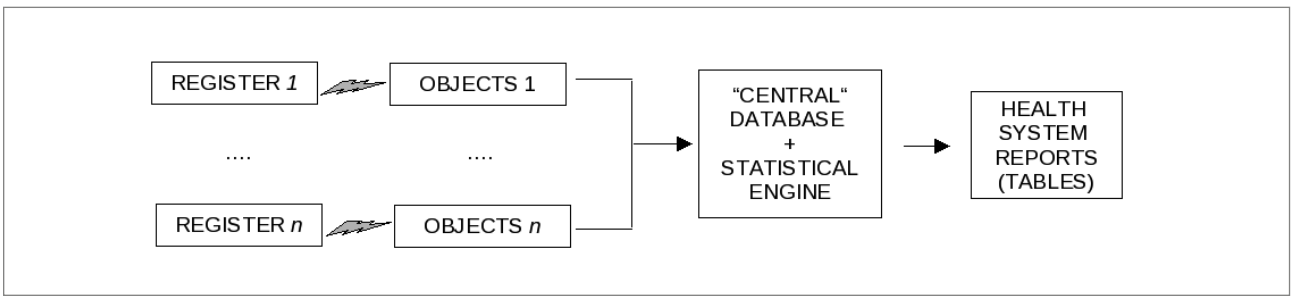


Figure 1.2.2. The data model for the "global SEDIS"

agreed for the BIRO common dataset.

Standardized instructions (XML Schema) have been specifically developed to implement common BIRO definitions into a uniformly defined database allowing the use and pooling of data collected from different centres. During the project, the need for more specifications than initially planned clearly emerged and more schemas were created for the scope of including more datasets from each local data source in the local database.

A "*Metadata Dictionary*" has been realized in XML to incorporate a broad and evolving set of diabetes-related concepts. New variables were derived from those incorporated into the common BIRO dataset. Details on all data definitions are provided in the specific Chapter 3.1.

A fair deal of work was required to ensure data management through various conversion tools and database operations.

According to data definitions, a flat text file (XML export) needs to be produced by each centre through the combined and repeated use of Java tools and the JDBC driver. The operation can be completed with some basic pre-processing of local data allowing to comply with basic requirements (e.g. storing one record for each individual subject in the production of the so-

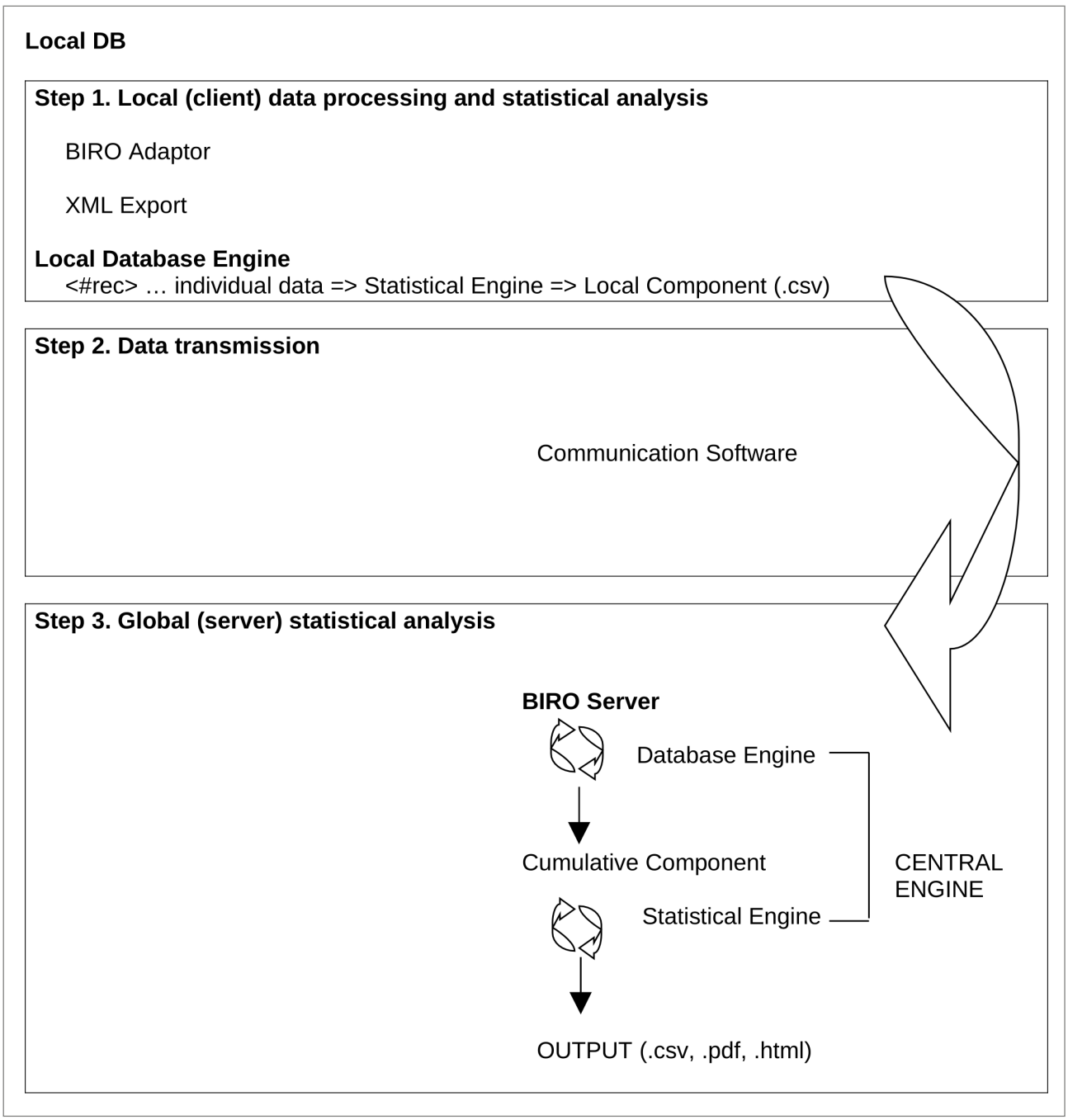


Figure 1.2.3. Procedural flow in the BIRO software

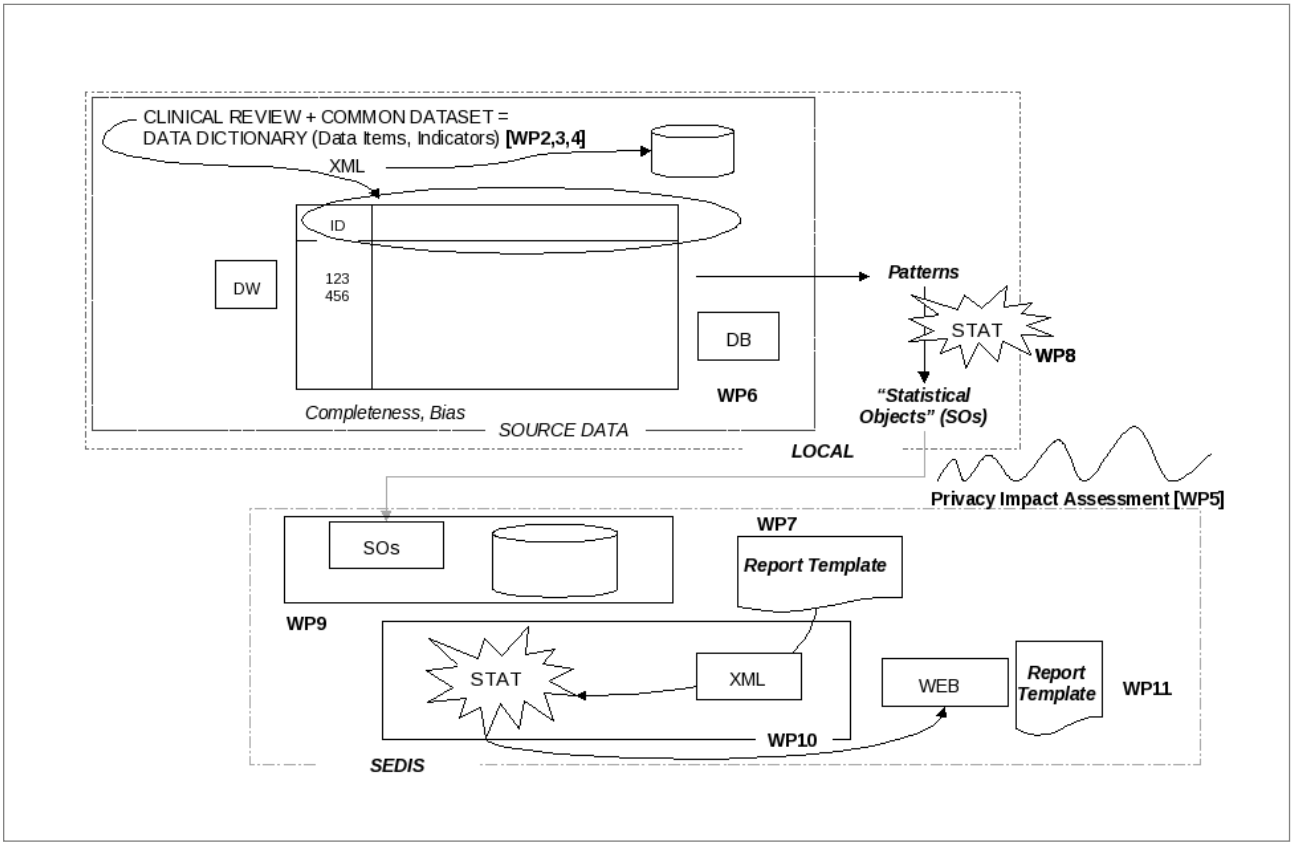


Figure 1.2.4. BIRO technology and work packages

called "Merge Table"). Details of these operations are provided in the Appendix.

A configuration file is needed by the BIRO Adaptor to apply specific options to the relevant driver, either directly or through the use of a visual application specifically realized to assist high level users.

The BIRO "*Database Manager*" has been developed to read XML files and store records into a local (Postgres) database that is used to organize local data in an optimal way, so that they could be automatically processed by the statistical engine. By the way, the same database can be used to run accurate and high performance SQL queries on top of standardized datasets, completely independently from the BIRO system.

The Database Manager has been developed using the Java language, Castor and Hibernate, through an ad hoc configuration file that must be defined by the local administrator.

Details on all BIRO database functions are provided in Chapter 3.3.

The *report template* served to specify precisely all outputs that must be produced by the BIRO system. This way, the same structure can be used to automate the production of reports for both the individual centres and the global collaboration. This feature is extremely convenient, as it allows using the same set of basic statistical functions for multiple, repeated applications. The template is described in great detail in Chapter 2.2.

The BIRO "*Statistical Engine*" connects to the local BIRO Postgres database to run statistical functions that create "*statistical objects",* i.e. essential data requested from local units that allows computing diabetes indicators (for more details refer to Chapter 3.3).

The statistical engine connects to the local database using the open source statistical R software, with proper Postgres drivers.

According to the specifications given by the report template, and the associated relevant definitions of the statistical objects, the statistical engine processes the database and delivers statistical objects in the form of small CSV datasets. These flat text files are further processed to output local reports in the form of pdf and html files, using the Latex software.

A compressed CSV folder is created to include all statistical objects produced by each run of the local reporting system, classified by date and centre id. This operation completes step 1) of the local engine.

Details on the local component of the distributed statistical analysis are provided in Chapter 3.4.

Step 2 involves data transmission.

This step was properly seen as the most relevant for privacy impact assessment: a clear definition of what was really necessary to be transmitted, if the target

could be actually sent from a legal perspective, whether alternative strategies were possible, and finally how to send the target, needed to be agreed from the very early stage of design.

The WP privacy impact assessment was successfully carried out, defining the best BIRO architecture herein described. For the scope, a literature review, a novel method including a Delphi panel guided by ad hoc materials, and a Consensus panel were duly organized. Specification for data transmission were finally agreed in the framework of the 2nd Investigator's Meeting in Cyprus.

Details regarding the methods and results of privacy impact assessment are reported in Chapter 3.2.

Specialized communication software has been developed to securely transmit the CSV folder including statistical objects from the local to the Central BIRO system. Web services have been used to comply with basic requirements, including availability of an open platform-independent standard, XML support, usability over Internet protocols, open source implementation and comprehensive security support. Confidentiality has been ensured by using encryption and data integrity, as well as non-repudiation provided by digital signatures.

Further details of communication software are provided in Chapter 3.3.

The *central* part of the BIRO Architecture includes the set of software tools required by the BIRO server to undertake Step 3: global statistical analysis.

Step 3 involves several operations including database processing and statistical analysis, covered by the WP Central Engine.

At the central level, individual data are no longer required as the BIRO system only requires aggregate data, so all database specifications include meta-data mainly referred to the concept of statistical objects.

A specialised application (BIRO CSV Importer) has been developed in Java to read CSV files embedding statistical objects and to store them as separate tables of the Central BIRO database.

As for the Adaptor and Database Manager, a configuration file is required to allocate proper options.

Related statistical objects, transmitted by separate centres, are appended to the same table to form a global collection of local aggregate data.

The BIRO Database component of the Central Engine has been specifically developed to load and to organize all central aggregate data, as well as to perform basic data processing.
Elementary Postgres functions have been used to compute a "*cumulative component*" for each statistical object as a pooled estimate of multiple "*local*" statistical objects.

Advanced statistical analysis in the Central Engine is performed by specific R functions. The cumulative components of statistical objects are processed to deliver all elements of the global report required to deliver the same template used for the local analysis. The template will be populated with results referring to the whole universe of BIRO collaborative centres.

Outputs of the Central Engine include a complete pdf report (as defined in the template), an html report (following specifications in the web portal), and CSV data, all produced using R and Latex software.

Details of the Central Engine are included in Chapter 3.3 and Chapter 3.4.

The final section of software development involves integration of the BIRO architecture into a unique, integrated software.

The BIRO process can be triggered by a simple "*local*" user friendly (GUI) application (see Chapter 3.3 and Appendix) allowing the user to:

- export local data stored into a local database to XML files running the BIRO Adaptor
- import XML files to the local database using the BIRO Database Manager
- produce the local statistical report
- send the local statistical objects to the Central BIRO System

The "central" application must be run by the BIRO Administrator to:

- import statistical objects stored as csv files
- run the global statistical analysis
- produce the global BIRO report

The BIRO architecture specified by the privacy impact assessment requires for the Central Engine to be managed by the BIRO coordinator, responsible of compliance with all national and international security rules for the maintenance of the server.

The global reports produced by the Central Administrator are automatically linked to the BIRO Web Portal, through a structure that is specified in detail in Chapter 3.5.

As far as installation of the BIRO system is concerned, a specific directory structure (Table A.8) has been identified to allocate all different components of the BIRO system and drive the construction of a comprehensive software setup for both the client and server side (see Appendix).

An analysis of the state of the art of diabetes information in New Member States, including a discussion of the limitations and barriers to the usage of BIRO, and various test runs of the software, were carried out by the WP Technology Transfer. The result showed that the system can be successfully used to load data from Cyprus, Malta and Romania, although further refinements are needed to optimize data translation to the

BIRO standard. In general, the BIRO system as it stands still requires a technical person to be operated. Technology transfer showed that BIRO should be more oriented to health professionals to influence more directly completeness of information, quality and outcomes, as originally planned. Details of this experience are reported in Chapter 4.1.

An evaluation of the BIRO project has been regularly undertaken as described in Chapter 4.2. It showed that apart from various aspects to be optimised, the plan is realistic and the project is relevant to fulfil the needs of diabetes information. However, the system must be easier, and for this it should be supported by appropriate training material, and a clear plan for dissemination and use of the software. It has been suggested that the whole initiative should be more "entertaining" for the general public, for which it is necessary to separate technical material from the actual application. A demo should be made widely available.

1.2.4 Discussion

The BIRO collaboration is the logical continuation of a long path spanning on a timeframe of more than 20 years of research and development in diabetes care, prevention and information technology.

Partners of the Consortium have been core activists in the implementation of the "St.Vincent Declaration"[34], a cooperative initiative of the European Region of the International Diabetes Federation and the WHO Regional Office for Europe for the improvement of Quality of Diabetes Care, as well as active members of the DIABCARE Q-NET project sponsored under the 4th Framework[1] and the DIABCARD Dataset repository.

After many years from the completion of DIABCARE, BIRO revamps its basic intuition, improving its design with new technology through which management of individual records is left to the single participating institution. At the same time, BIRO is substantially stronger as it may allow to deliver an in-depth analysis of systems of care, through the application of sophisticated epidemiological techniques and advanced statistical methods that were not adequately considered by previous projects.

The BIRO proposal was very timely presented in correspondence to the results of the EUDIP project[2]. The EUDIP report included a comprehensive list of quality and outcome indicators for both Type 1 and Type 2 diabetes, ranging from risk factors to process and outcome indicators. These measures provided an essential starting point for the construction of a knowledge repository that could drive the statistical analysis of registers' data.

Unfortunately, neither EUDIP nor DIABCARE were able to deliver a sustainable solution for routine reporting of diabetes indicators: while the former did not solve the problem of automatic updating of national indicators, the latter lacked a robust approach to privacy protection,

BIRO offers a solution that closes the loop between information and clinical practice using records that are normally available in health care organizations. Since regional registers are already linked to other data sources locally (hospital discharges, mortality, prescriptions), SEDIS can cover most indicators delivered by EUDIP. Furthermore, it makes possible to link population-based measures that can be obtained in a much more precise way at the regional level.

The ambitious goal of BIRO is to construct a European health information system helping different stakeholders to provide better clinical governance, better management of diabetes, and better analysis (Figure 1.2.5).

Clinical governance has recently influenced the organ-

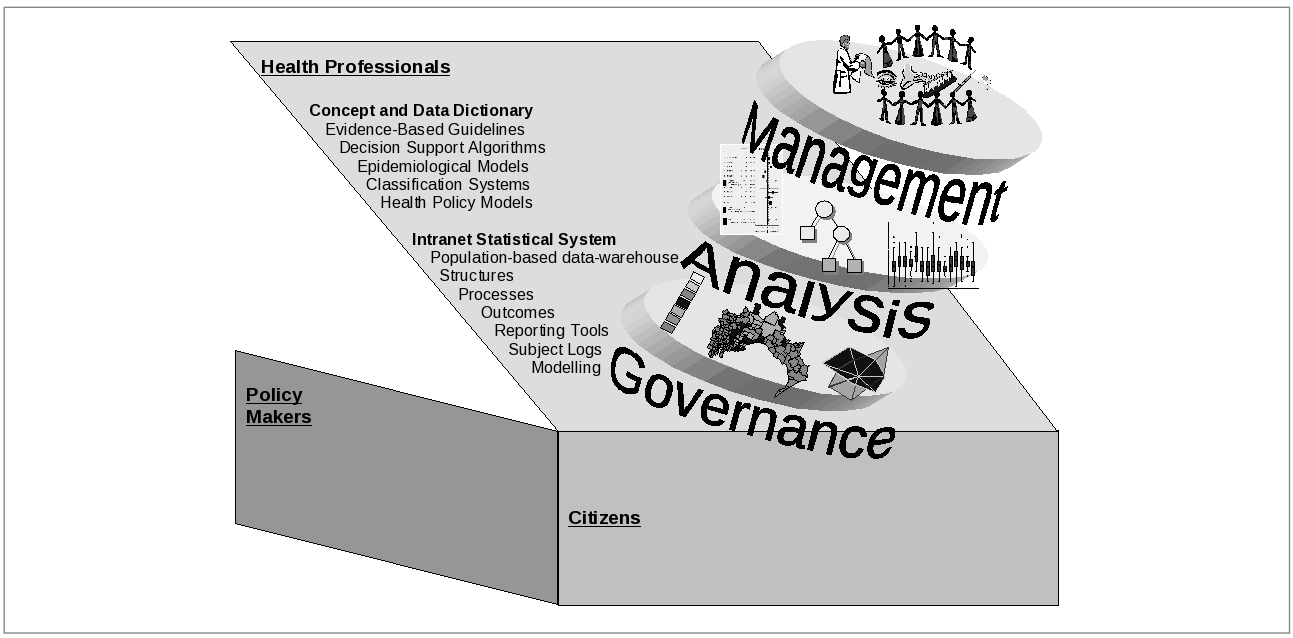


Figure 1.2.5. BIRO as a Multilevel Diabetes Initiative

ization of modern health systems and can play an important role in shaping quality improvement in diabetes[37]. The concept advocates an efficient use of resources on the basis of population needs, to support the adoption of adequate prevention strategies and health services based on the evidence[38-42].
The concept is intimately related to the principle of accountability, which can be realised only through the availability of up-to-date, well structured information[43].

Disease management is the key instrument that will support the continued availability of diabetes indicators. BIRO can activate a virtuous cycle and generate synergies to improve health outcomes. Managed care organisations that consistently monitor quality have shown significant improvements in outcomes, even short term, particularly because this information can support a more intensive follow up that can be effective in treating many different conditions[44-49].

Analysis can be also favoured by BIRO, as it allows for more data to be made available to more referents in anonymous, secure format. Researchers can also embed own procedures in the system without actually accessing data, by exploiting the open source nature of BIRO[50]. However, the distributed framework and associated specifications (local, global) must be duly taken into account. This way, new techniques e.g. multilevel models and data mining[51-59] can be added to update traditional approaches that may not be appropriate in the treatment of complex phenomena[28].

During the project, members of the Consortium have appreciated many other aspects that characterise the evolution implied by genuine participation.

Some aspect needed to be re-evaluated by those groups who have already established systems in place. The process allowed to rethink some basic functionalities of diabetes registers, including questions e.g. “are we truly population-based?”, “does this schema apply equally to all regions?”, “would it be worth to link other data sources to those that we currently use?”, “is information we collect analysed correctly from an epidemiological point of view?”.

Take mortality as an example. There is very little controversy on how mortality can be recorded and/or considered from the point of view of data definition. Yet, this may represent not a trivial point for local data sources: clinical units may not necessarily know if a subject has died, unless linkage with a client master index is routinely carried out. When discussing the “activity status” among BIRO partners, it seemed that only “active subjects” were relevant for analysis. By the way, that can be true for diabetes management, but it would be grossly incorrect from a population-based perspective. If deaths are not recorded, the longer the time reference, the more biased a diabetes indicator becomes if the cohort only includes “active alive patients”. In the end, a new schema was constructed to allocate a table that specified “activity” in broad terms, including residency and life status. That could be more difficult, but it can be crucial for better indicators.

We discovered that privacy impact assessment can be successfully used as a general method to guide the design of health information systems. Differently from other similar applications, the process is highly multidisciplinary and must facilitate self evaluation because for the system to be actually used, its strengths and limitations (even from a legal point of view) must be deeply understood, accepted and in the end also advocated by end users. The Consensus Panel held in Cyprus, with its marking exercise not so simple to perform for the different skills engaged, is going to be remembered by partners of the Consortium.

On the other hand, we learned that BIRO can help others to act on diabetes in a way that perhaps we could not foresee or believe possible. The Cyprus Ministry of Health used BIRO as a platform to create a regional network. A diabetic clinic was opened for the first time in Larnaca, directly managed by local referents of the BIRO project. Common specifications were used to create a fresh IT application that has been used in the clinic to collect patients data. After two years, records from a large sample of patients have been loaded on the BIRO system and can be analysed routinely.

A clinical disease management training program has been initiated in connection with the BIRO project, and the register is planned to spread its activities on the island to become the National Diabetes Register.

Finally, we learned how valuable can be to bring own data to a session, putting notebooks side by side, and perform a live analysis to refine the system progressively, as happened in Bergen. To realise it in a European framework, where languages and cultural differences are still (luckily and joyfully) different, may not be such an obvious matter.

These achievements may justify alone the effort for the BIRO system to be realised, regardless of the actual availability of an innovative product. That was perhaps unexpected and highly appreciated by all partners.

Further to this, there is the BIRO software.

It is a very practical, tangible product that can operate on top of databases already available. The BIRO system can provide users with an opportunity to get reliable information relative to: a) common patient management strategies, as expressed by the level of adherence to clinical guidelines, b) average results for specific subjects with diabetes in the particular local context; and c) measures of ‘global variability’ of diabetes care in Europe.

The same practicalities of the BIRO project also highlight some of its limitations.

The approach does not appear immediately simple and needs to be well explained to prospective partners. The BIRO budget was too limited to cover a targeted action of dissemination, including the preparation of material that would facilitate the uptake of the technology.

In the final phase of the BIRO project, a sequel project, EUBIROD, was successfully submitted by the BIRO Consortium and granted funds by the European Commission. This framework will allow to overcome these limitations, offering a challenging test-bed for a much broader group of partners from 20 EU countries, plus an extra European from the Middle East (Kuwait).

Open source software may facilitate the continuous development of BIRO, but it is also evolving very rapidly. Transforming a prototype into a product “out of the box” can be a real challenge for non industrial academic partners working in the field of public health.

A much wider audience may even require a radical change in many specifications - an aspect that has been taken into account into EUBIROD – which can become overwhelming.

To progress BIRO, continuity will represent the most important asset of the BIRO Consortium, as the project relies on the know how of few key referents, which in case of excessive turn over, may cause substantial difficulties for the continuation of the program.

1.2.5 Conclusions

The BIRO Consortium, over a timeframe of 40 months, has delivered a product that can be now rolled out to a network of clinical units, regions, and Member States.

The development of BIRO creates new chances to operate in the direction of a European Diabetes Register, towards which the project has contributed with an innovative and compelling prototype that deserves to be seriously taken into account by the European Union.

An expanding collaboration, under the banner of EUBIROD, will continue to operate to further develop the system through its direct application for the publication of reports on the web portal.

An alarmingly expanding population of people with diabetes requires to be followed up by clinicians and institutions more regularly and effectively.

The BIRO Consortium looks forward to accurately document any progress Europe will make in fighting and finally defeating such a burdening and subtle disease.

References

1. Piwernetz K. DIABCARE Quality Network in Europe--a model for quality management in chronic diseases, Int Clin Psychopharmacol 2001 Apr;16 Suppl 3:S5-13.
2. EUDIP Study Group, Establishment of indicators monitoring diabetes mellitus and its morbidity, EUDIP 2002 Report
3. Boyle D, Cunningham S, Sullivan F and Morris A, on behalf of the Tayside Regional Diabetes Network, Technology Integration for the provision of population-based equitable patient care: the Tayside Regional Diabetes Network: a brief description, Diabetes Nutrition and Metabolism 2001;14(2), 100-103.
4. Bergrem H, Leivestad T, Diabetic nephropathy and end-stage renal failure: the Norwegian story, Adv Ren Replace Ther. 2001 Jan;8(1):4-12
5. Beck, P, Battlogg, K, Gfrerer, R, Trajanoski, Z, Wach, P, and Pieber, T. Internet based client server application for diabetes management, Diabetes, Nutrition & Metabolism, Clinical and Experimental 13(4), 245. 2001.
6. Massi Benedetti M, Orsini Federici M, PROMODR: Progressive Model for Diabetes Register, Diabetes Nutrition and Metabolism 2001;14(2), 96-100.
7. Azzopardi J, Fenech FF, Junoussov Z, Mazovetsky A, Olchanski V., A computerized health screening and follow-up system in diabetes mellitus, Diabet Med. 1995 Mar;12(3):271-6.
8. Pruna S, Georgescu M, Stanciu E, Dixon RM, Harris N, The Black S ea Tele-Diab System: development-implementation-clinical evaluation, Stud Health Technol Inform. 2000;77:656-60.
9. The Diabetes Control and Complications Trial Research Group, The effect of insulin treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus, N Eng J Med, 1993; 329 (14):970-986.
10. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33).UK Prospective Diabetes Study (UKPDS) Group, Lancet, 1998 Sep 12;352(9131):837-53.
11. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group, Lancet, 1998 Sep 12;352(9131):854-65.
12. Tight blood pressure control and risk of macro-vascular and microvascular complications in type 2 diabetes: UKPDS UK Prospective Diabetes Study Group, BMJ, 1998 Sep 12;317(7160):703-13.
13. Massi-Benedetti M; CODE-2 Advisory Board, The cost of diabetes Type II in Europe: the CODE-2 Study, Diabetologia. 2002 Jul;45(7):S1-4. Epub 2002
14. Ellrodt G et al., Evidence-based Disease Management, JAMA, 1997; 278:1687-1692.
15. Roos NP, Black C, Roos LL, Frohlich N, DeCoster C, Mustard C, Brownell MD, Shanahan M, Fergusson P, Toll F, Carriere KC, Burchill C, Fransoo R, MacWilliam L, Bogdanovic B, Friesen D., Managing health services: how the Population Health Information System (POPULIS) works for policymakers., Med Care. 1999 Jun;37(6 Suppl):JS27-41
16. Vaughan NJA and Massi Benedetti M, A review of European experience with aggregated diabetes databases in the delivery of quality care to establish a future vision of their structure and their role, Diabetes, Nutrition and Metabolism 2001, 14 (2): 86-87.
17. Davis TM, Cull CA, Holman RR., Relationship between ethnicity and glycemic control, lipid profiles, and blood pressure during the first 9 years of type 2 diabetes: u.k. prospective diabetes study (ukpds 55), Diabetes Care. 2001 Jul;24(7):1167-74.
18. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, Hadden D, Turner RC, Holman RR., Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35):prospective observational study, BMJ, 2000 Aug 12;321(7258):405-12.
19. Quality of life in type 2 diabetic patients is affected by complications but not by intensive policies to improve blood glucose or blood pressure control (UKPDS 37). U.K. Prospective Diabetes Study Group. Diabetes Care 1999 22: 1125-1136
20. Meigs JB, Singer DE, Sullivan LM, Dukes KA, D'Agostino RB, Nathan DM, Wagner EH, Kaplan SH, Greenfield S., Metabolic control and prevalent cardiovascular disease in non-insulin-dependent diabetes mellitus (NIDDM): The NIDDM Patient Outcome Research Team, Am J Med, 1997;102(1):38-47.
21. Hayward RA, Manning WG, Kaplan SH, Wagner EH, Greenfield S., Starting insulin therapy in patients with type 2 diabetes: effectiveness, complications, and resource utilization, JAMA, 1997 Nov 26;278(20):1663-9.
22. Greenfield S, Rogers W, Mangotich M, Carney MF, Tarlov AR., Outcomes of patients with hypertension and non-insulin dependent diabetes mellitus treated by different systems and specialties. Results from the medical outcomes study, JAMA, 1995, 8;274(18):1436-44.
23. Carinci F, Rose W, Advantages of Distributed Statistical Processing in Health Information Systems: the H+MetaBase project, Submitted, Monash Institute of Health Services Research, Melbourne, Australia, 2003.

24. F.Carinci et al., RISS-H - Progetto Diabete - SDO di Pazienti Diabetici - Analisi stratificata per Ospedali - Regione Veneto 1999, available at: http://www.fabcarinci.net/projects/riss/riss_samples.html
25. Stier DM, Greenfield S, Lubeck DP, Dukes KA, Flanders SC, Henning JM, Weir J, Kaplan SH., Quantifying comorbidity in a disease-specific cohort: adaptation of the total illness burden index to prostate cancer, Urology. 1999 Sep;54(3):424-9.
26. Sullivan LM, Dukes KA, Harris L, Dittus RS, Greenfield S, Kaplan SH., A comparison of various methods of collecting self-reported health outcomes data among low-income and minority patients, Med Care, 1995 Apr;33(4 Suppl):AS183-94.
27. Greenfield S, Sullivan L, Dukes KA, Silliman R, D'Agostino R, Kaplan SH., Development and testing of a new measure of case mix for use in office practice., Med Care, 1995 Apr;33(4 Suppl):AS47-55.
28. Greenfield S, Kaplan SH, Kahn R, Ninomiya J, Griffith JL., Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results, Ann Intern Med 2002 Jan 15;136(2):111-21.
29. Charlson ME, Pompei P, Ales KL, MacKenzie CR, A new method of classifying prognostic comorbidity in longitudinal studies: development and validation, Journal of Chronic Diseases, 1987; 40:373.
30. Elixhauser et al., Comorbidity measures for use with administrative data, Medical Care,36,1,8-27.
31. Currie et al., Epidemiology and Costs of Acute Hospital Care for cerebrovascular disease in diabetic and non diabetic populations, Stroke,1997;28:1142-1146.
32. Krop et al, Predicting expenditures for Medicare beneficiaries with diabetes, Diabetes Care 1999; 22:1660-1666.
33. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG, The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease, JAMA, 1999 Jun 9;281(22):2098-105.
34. World Health Organisation (Europe) and International Diabetes Federation (Europe). Diabetes Care and research in Europe: the St.Vincent Declaration. Diabet Med 1990; 7:360.
35. Blair Stewart, Privacy Impact Assessment: some approaches, issues and examples, Assistant Commissioners Office of the Privacy Commissioner, New Zealand.
36. Vaughan N, Confidentiality and diabetes registers, Diabetes Nutrition and Metabolism 2001;14(2), 114-117.
37. Diabetes Atlas 2000, International Diabetes Federation, Brussels, Belgium 2000.
38. Stevens A, Gabbay J, Needs assessment needs assessment, Health trends 1991; 23:20-3.
39. Wright J, Williams R, Wilkinson JR, Development and importance of health needs assessment, BMJ, 1998; 316:1310-3.
40. Feldstein MS, Effects of differences in hospital beds scarcity on type of use, BMJ 1964; ii: 562-5.
41. Kirkup B, Forster D, How will health needs be measured in districts? Implications of variations in hospital use, J Public Health Med 1990; 12:45-50.
42. Wilkin D, Patterns of referral: explaining variation; In: Ronald M, Coulter A eds. Hospital referrals. Oxford:Radcliffe Medical Press, 1994.
43. Roos NP, Black C, Roos LL, Frohlich N, DeCoster C, Mustard C, Brownell MD, Shanahan M, Fergusson P, Toll F, Carriere KC, Burchill C, Fransoo R, MacWilliam L, Bogdanovic B, Friesen D., Managing health services: how the Population Health Information System (POPULIS) works for policymakers., Med Care. 1999 Jun;37(6 Suppl):JS27-41
44. National Committee on Quality Assurance, The state of managed care quality –1999, Washington DC: National Committee for Quality Assurance, 1999.
45. Mc Cullough DK, Price MJ, Hindmarsch M, Wagner EH, A population-based approach to disease management in a primary care setting: early results and lessons learned, Eff Clin Pract
46. Morris AD, Boyle DI, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, MacDonald TM, The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register, Br Med J, 1997 Aug 30;315(7107):524-8.
47. Kuhlthau K, Walker DK, Perrin JM, Bauman L, Gortmaker SL, Newacheck PW, Stein REK, Assessing managed care for children with chronic conditions, Health Affairs, 1998; 17 (4), 42-52.
48. Keever GW, Disease management on intranet saves $1.2 million, Health Management Technology, Nov 1998, p. 47.
49. New JP, Hollis S, Campbell F, McDowell D, Burns E, Dornan TL, Young RJ., Measuring clinical performance and outcomes from diabetes information systems: an observational study, Diabetologia. 2000 Jul;43(7):836-43.
50. Churches T, Carinci F, Open source at the interface between policy and academia: towards evidence-based information systems, 4th International Conference on the Scientific Basis of Health Services Research, Sydney, 22-25 September 2001.
51. Rice N, Leyland A, Multilevel models: applications to health data, J Health Serv Res Policy, 1996; 1(3), 154-164.
52. Daniels MJ, Gatsonis C, Hierarchical Polytomous Regression Models with Applications to Health

Services Research, Statistics in Medicine, 1997; 16, 2311-2325.
53. Carinci F, Pellegrini F, Nicolucci A, Exploring adherence to clinical guidelines: application of regression tree analysis in diabetes, 4th International Conference on the Scientific Basis of Health Services Research, Sydney, 22-25 September 2001.
54. Nicolucci A, Carinci F, Ciampi A., Stratifying patients at risk of diabetic complications: an integrated look at clinical, socioeconomic, and care-related factors. SID-AMD Italian Study Group for the Implementation of the St. Vincent Declaration, Diabetes Care, 1998, Sep;21(9):1439-44.
55. Carinci F, Nicolucci A, Pellegrini F, Regression trees in health services and outcomes research:an application of the RECPAM approach using quality of care as a criterion, Technical Report, Monash Institute of Health Services Research, Monash University, 2001.
56. Carinci F, Corrado D, Dettorre A, Pellegrini F, A multilevel approach to health systems analysis using RISS (Reporting-by-Intranet Stat System), 4th International Conference on the Scientific Basis of Health Services Research, Sydney, 22-25 September 2001.
57. Carinci F, RISS Samples, available at: http://www.fabcarinci.net/projects/riss/riss_samples.html
58. Nicolucci A., Cavaliere D., Scorpiglione N., Carinci F., Capani F., Tognoni G., Massi Benedetti M., the Italian Study Group for the Implementation of the St. Vincent Declaration, A comprehensive assessment of the avoidability of long term complications of diabetes mellitus: a case-control study, Diabetes Care, 19, 9:927-933, 1996.
59. F.Carinci, Southern Population Health Information System (SPHIS). An integrated Population-based Data Warehouse, available at: http://www.fabcarinci.net/projects/chsr/sphis.html

An evidence-base for EU diabetes information

Peter Beck

ABSTRACT

Introduction
The BIRO project proposes the creation of a Shared “Evidence-Based” Diabetes Information System to support European health policy. The essential items to be collected must be continuously revised, taking into account the actual relevance of new knowledge on diabetes for population health and everyday practice.

Objectives
Aim of the clinical review is to give an overview of the existing literature in order to propose a set of relevant measures collected routinely at the provider level that can also be used for benchmarking of diabetes prevention and care in the different European health care systems.

Materials and Methods
In a first step existing guidelines were viewed. Comments, literature references and potentially interesting data items were extracted and clustered per thematic area. A data item per definition is “one single piece of data” or “the smallest piece of information that can be obtained from a survey or census”. As several data items were not yet covered by existing indicators, several new or modified ones were suggested. In a second step indicators were defined, whereby indicators are seen as “a measure used to determine, over time, performance of functions, processes, and outcomes”. The selection of indicators was carried out along the recommendations for indicator evaluation developed by the US Institutes of Medicine and applied by the OECD Quality Indicators Project. According to this approach, indicators have to be relevant, secondly, they have to be scientifically sound, and thirdly, they have to be potentially feasible. In a third step the indicators were rated according to the above mentioned scheme in a consensus process within the consortium.

Results
A structured process was used to identify relevant and scientifically sound topics for data analysis in BIRO. A list of N=85 candidate indicators was suggested, with a total of N=49 items selected as relevant for BIRO. The level of feasibility was taken into account when defining the data items required for computing such indicators.

Conclusions
The set of indicators identified and precisely described by the clinical review allows to define the main characteristics of the “BIRO core data set” to be collected by participating regions.

2.1.1 Introduction

The BIRO project proposes the creation of a Shared "Evidence-Based" Diabetes Information System to support European health policy. The essential items to be collected must be continuously revised, taking into account the actual relevance of new knowledge on diabetes for population health and everyday practice. Aim of the clinical review is to give an overview of the existing literature in order to propose a set of relevant measures collected routinely at the provider level that can also be used for benchmarking of diabetes prevention and care in the different European health care systems.

2.1.2 Materials and methods

In a first step existing guidelines were viewed. Comments, literature references and potentially interesting data items were extracted and clustered into the following thematic areas:

- Risk profile for Diabetes
- Diagnosis and classification
- Risk profile for complications and interme diate outcomes
- Management and care of diabetes and its comorbidities
- Self management and lifestyle management
- Complications
- Individual characteristics
- Health status
- Demographic and socio-economic factors
- Health system & health care delivery

In a second step indicators were defined, whereby indicators are seen as "a measure used to determine, over time, performance of functions, processes, and outcomes". The selection of indicators was carried out along the recommendations for indicator evaluation developed by the US Institutes of Medicine as described in the OECD Health Technical Paper No.15[1] whereby firstly, indicators have to capture an important performance aspect, secondly, they have to be scientifically sound, and thirdly, they have to be potentially feasible.

In a third step the indicators were rated according to the above mentioned scheme, whereby the importance of an indicator can be further broken down into three dimensions:

Impact on health: What is the impact on health associated with this problem? Does the measure address areas in which there is a clear gap between the actual and potential levels of health?

Policy importance: Are policymakers and consumers concerned about this area?

Susceptibility to being influenced by the health care system: Can the health care system meaningfully address this aspect or problem? Does the health care system have an impact on the indicator independent of confounders like patient risk? Will changes in the indicator give information about the likely success or failure of policy changes?
The scientific soundness of each indicator can also be broken down into two dimensions:

Face validity: Does the measure make sense logically and clinically?

Content validity : Does the measure capture meaningful aspects of the quality of care?
A comprehensive discussion of the importance and scientific soundness of those indicators can be found in a paper by Fleming et al.[2], and in materials produced by the Alliance[3].

The feasibility of an indicator reflects the following two dimensions:

Data availability: Are comparable data to construct an indicator available on the international level?

Reporting Burden: Does the value of the information contained in an indicator outweigh the cost of data collection and reporting?

In a fourth step, the collected information and rating decisions were reviewed in a plenary sessionand discussed until a core list was selected on the basis of partners" considerations of the actual feasibility of their adoption within the context of the BIRO project.

Sources used in the literature review

- Health indicators: EUDIP, ECHI, OECD
- Guidelines: IDF, SIGN, Consensus on diabetic foot, New Zealand, ADA, Canada, German Diabetes Association
- Systematic literature search in: Cochrane database, Medline

2.1.3 Literature review

2.1.3.1 Risk profile for diabetes

Obesity and Overweight [4-6]

Overweight/obesity is a major risk factor for type 2 diabetes. (9). It causes insulin resistance, which will lead eventually to type 2 diabetes. WHO uses BMI to define obesity (overweight BMI > 25, obesity BMI >30). BMI = weight [kg] / (height [m])^2. IDF uses a definition by waist circumference with ethnicity specific values: Waist circumference of >94cm (men) and >80cm (women) for Europids, >90cm (men) and >80cm (women) for South Asians and Chinese, >85cm (men) and >90cm (women) for Japanese. BMI collection was considered more feasible than waist circumference.

Physical Inactivity [7-10]

Physical inactivity as an indicator of sedentary lifestyle, contributes to the development of type 2 diabetes, partly through increased risk for obesity. The *eupass* (European Physical Activity Surveillance System) project tested the International Physical Activity Questionnaires (IPAQ) a questionnaire which reflects dura-

tion, intensity and frequency of Health-enhancing physical activity (HEPA). This assessment instrument could well be used as basis for an indicator, but no data is available in the moment.

Nutritional habits[11-14]

Nutritional habits have an influence on obesity. Increased saturated fat intake, increased protein intake will influence insulin resistance and contribute to the development of type 2 diabetes. The effect of fast-acting carbohydrates on insulin resistance is questioned. *EUROHIS* contains questions to be potentially used for an indicator on nutritional habits, but there is no data collection currently going on in BIRO partner regions.

Gestational diabetes[15-17]

Gestational diabetes has been reported as a potential risk factor for the development of type 2 diabetes. This risk factor for type 2 diabetes in women should be re-evaluated and the prevalence of gestational diabetes possibly recommended as a core indicator.

2.1.3.2 Diagnosis and Classification

Diagnosis of Diabetes[18-23]

Diagnoses of diabetes may be either confirmed by a fasting plasma glucose (FPG) value of ≥7.0 mmol/l, >125 mg/dl) or by an oral glucose tolerance test (OGTT). The criteria for diagnosis of diabetes were adopted from the WHO and American Diabetes Association (ADA).

Classification of Diabetes[24-26]

The WHO uses the following classification: type 1 diabetes mellitus (with subtypes), type 2 diabetes mellitus, other specific types of diabetes (with subtypes) and gestational diabetes mellitus. NHS also includes pre-diabetes stages: impaired fasting glucose (IFG) and impaired glucose tolerance (IGT) - after discussion, these stages were not considered in scope for BIRO. MODY (Maturity onset diabetes of youth) and LADA (autoimmune diabetes mellitus in adults) are relatively small numbers, and data will not be reliable, because many are unrecognized. Therefore these types will also not be considered in BIRO.

Diabetes onset [27-29]

Age at onset plays an important role since duration of diabetes influences the risk for chronic complications. Diagnosis of type 2 diabetes is not always straightforward and the level of identified diagnosis might depend on national screening programmes. Age at onset also has to be considered for childhood onset of type 1 diabetes mellitus.

2.1.3.3 Risk profile for complications and intermediate outcomes

Glucose level [30-48]

Many prospective studies demonstrate an association between a good metabolic control and a reduction in micro vascular (retinopathy, nephropathy neuropathy) and macro vascular (cardiovascular) complications. As a measure for average blood glucose levels, the surrogate parameter HbA1c is assessed. For insulin treated patients, hypoglycaemic events should be assessed in combination.

Guidelines recommend low blood glucose levels and repeatedly were adjusted to even lower target HbA1c levels for patients. However, recent results from long term cohort studies showed that too low blood glucose levels/ too aggressive therapy may not always be beneficial and even may harm patients.

Continuous ambulatory blood gl ucose monitoring has become available in recent years. There is still no good evidence-base for its use, particularly in people with Type 2 diabetes.

Blood pressure[49-58]

Presence of hypertension is an independent risk factor for the development of complications. It is an established risk factor for the development of macular oedema and it is associated with proliferate retinopathy. In several randomized controlled trials the absolute benefit of a blood pre ssure reduction in regard to macrovascular mortality and morbidity was superior to the benefit of blood glucose reduction (e.g. UKPDS). It is recommended to collect numeric blood pressure values [mmHg] independent from target blood pressure values. However, currently a value of 140/90 mmHg was considered the target blood pressure for patients with diabetes, although some guidelines state lower targets. In the analyses also the blood pressure distribution should be considered. It was suggested to record the percentage of patients with SBP >140 separately from percentage patients with DBP >90. Correct blood pressure measurement according to international standards is a prerequisite and not further discussed here.

Lipids[59-65]

Abnormal lipid profiles in patients with type 2 diabetes contribute to higher rates of cardiovascular complications. Through dietary and therapeutic intervention, reduction of this risk can be obtained.

The Friedemann equation to calculate LDL cholesterol gives invalid results with non-fasting triglycerides. Its use in diabetic patients has been questioned. Fasting samples are not important for most lipids, with the exception of triglycerides. The values for fasting/non-fasting are very different. In data sets it is often difficult to know if a triglycerides recording is fasting or not. Information of Triglycerides is likely to be very inaccurate, as we will often not know the fasting status. Only few existing data sets contain information on fasting status. In some settings patients can not come to clinic fasted. Due to these limitations in practice, the Total Cholesterol/HDL ratio was chosen as the recommended lipid parameter. A value of 4.5 is the target value for diabetic patients, a value >8.0 means high risk.

Microalbuminuria [66-68]

Microalbuminuria is a marker of vascular risk in diabetes. However, therapeutic consequences of microalbuminuria tests are unclear.
As a suggestion to deal with regional differences in assessment methods and thresholds, BIRO recommends to use local thresholds for micro-albuminurea and qualify them as normal, micro-albuminurea and proteinurea.

Smoking [69-89]

Smoking increases the risk for developing type 2 diabetes. Smoking increases the risk for cardiovasc ular diseases, diabetic specific neuropathy, peripheral arterial occlusive disease, erectile dysfunction, apoplexy and hypertension. Smoking increases the risk for diabetic nephropathy. Abstinence can meliorate an existing proteinuria. With renal failure smoking is an important risk factor for mortality.

The findings concerning diabetic retinopathy are controversial. Smoking status is assessed as number of cigarettes per day. Smoking duration is hardly ever known in practice, so the calculation of “pack-years” is currently not feasible.

Alcohol [90-103]

Alcohol is a risk factor for hypertension, hyperlipidemia, polyneuropathy, diabetic foot syndrome, erectile dysfunction, hypoglycemia and fatal cetoazidosis. Practicable to assess are alcohol intake in grams per week or units (1 unit = 10g) or number of drinks per week. Recording of a numerical value is preferred since recommended consumption limits are subject to periodic revision and may differ for pregnant women. The calculation of grams or units per week can be supported by a conversion table.

The classification according to ICD 10 is an alternative assessment method for alcohol consumption. Alcohol abuse / problematic drinking (ICD10, F10.2): Widely adopted is an average alcohol consumption threshold of 60 (men) and 40 (women) grams per day for problematic drinking. The diagnose of alcohol dependence (ICD 10, F10.3) is not directly attributable to amounts of alcohol consumption.

Drug abuse/dependence[104]

Self-reported street drug usage in young adults with Type 1 diabetes is common and may contribute to poor glycaemic control and serious complications of diabetes.

Foot screening[105-109]

Foot screening has the aim to assess the risk status of the diabetic foot. Foot screening should be performed at least once a year to detect neuropathy. The examination consists of bilateral palpation of foot pulses as well as a neurological examination with reflex status, vibration (128 Hz tuning fork), pain and pressure sensation (10-g Semmes-Weinstein monofilament, pin prick sensation). Skin and nail status, muscle atrophy (distal muscle strength), deformations, hyperkeratosis and the assessment of previous ulcers / amputations complete the assessment.

Eye screening [110-115]

Eye screening annually is recommended by current guidelines. The importance of screening people with Type 2 diabetes at diagnosis relates to the finding that between 21 and 39% of them already have some retinopathy (which may already be sight-threatening) by this time. For people who have no retinopathy at diagnosis of Type 2 diabetes, the chance of developing sight-threatening retinopathy within 2 years is less than 1%. Cataract is another important cause of visual loss in people with diabetes, being twice as common as in people without diabetes.
Recent review of screening methods found that digital photography best met the needs of appropriate sensitivity/ selectivity, feasibility and opportunities for quality assurance [8]. SIGN found that direct ophthalmoscopy only rarely achieved 80 % sensitivity even when carried out by properly trained operators. Recording modality (ie. ophthalmoscope, retinal photograph, slit lamp etc.) was considered but regarded too complex.

2.1.3.4 Management and care of Diabetes and its comorbidities

Diet [116]

Reduction of weight leads to a reduction of HbA1c and reduced need of OAD.

Glucose Control [117-132]

Treatment options with oral antidiabetic agents (OAD): Biguanides (metformin), sulfonylurea, glucosidase inhibitors, glitazones, glinides and DPP4 inhibitors. In the documentation it is recommended to allow assessment of contraindications against treatment options. Many national standards recommend the following distinction: Treatment with metformin for reduction of blood glucose, if BMI > 26 kg/m^2 and no contraindications for Metformin do exist. Treatment with sulfonylurea for reduction of blood glucose, if BMI < 26 kg/m^2 and no contraindications for Sulfonylurea do exist.

The main injectable therapy option is insulin therapy. Insulin type (human insulin / insulin analogues / animal insulin) as well as long acting / short acting / combination insulin should be assessed together with the total insulin consumption (in units per day) and the average number of insulin injections per day. No indicator for “type of insulin therapy” was introduced because the terms (CIT, MDI, ODI, PIT) for the insulin therapy types do not cover various therapy mixes. However, pump therapy should be assessed separately.
A second, recently introduced injectable treatment option, are GLP-1 antagonists.

The intensity of glucose control has to be discussed after recent study results (see item “Glucose Level”)

Blood pressure control [133-147]

Treatment options for blood pressure control are Diuretics (Thiazide Diuretics, Spironolactone), Beta-Blockers, Ca-Antagonists, ACE inhibitors, Angiotensin Receptor Blockers (AT II Blocker), Alpha-Blockers. Spironolactone is recorded separately, since it may have treatment benefits in acute MI and congestive heart failure independent of thiazides.

Lipid lowering therapy[148-154]

Treatment options for lipid lowering treatment are: Statins (Simvastatin, Pravastatin, Atorvastatin), Gemfibrizol and other fibrates, Ezetimibe, Nicotinic acid derivates, Fish oil supplementation in hypertriglyceridia. Simvastatin and Pravastatin are recommended by guidelines for secondary prophylaxis in patients with coronary heart disease, cerebrovascular disease and PVD and for primary prevention in high-risk patients.

Treatment of Cardiovascular disease (CVD) & peripheral vascular disease (PVD)[155,156]

Treatment options for coronary revascularization: PTCA (Percutaneous transluminal coronary angioplasty) with/without stent; CABG (Coronary Artery Bypass Surgery) and anti-platelet therapy. Treatment options for peripheral revascularisation: PTA (Percutaneous transluminal angioplasty) with/without stent, Bypass, anti-platelet therapy. Anti-platelet therapy summarizes aspirin, platelet aggregation inhibitors, heparin as well as thrombolysis.

2.1.3.5 Self Management and Lifestyle Management

Self monitoring and lifestyle interventions[157-171]

Blood Pressure Self Monitoring: Cochrane Review found association with moderate net reduction in diastolic blood pressure (weighted mean difference (WMD) -2.0 mmHg (95% CI -2.7 to -1.4 mmHg).

Self Monitoring of Glycaemic Control: Self monitoring of blood glucose (SMBG) is recommended for type 1 diabetes. It is generally accepted that SMBG is useful in insulintreated Type 2 diabetes (evidence is better for intensive insulin therapy). Generally, the evidence for blood glucose self monitoring in type 2 diabetes without insulin therapy is not satisfactory. Self management strategies with independent insulin dose adjustments are more flexible.

Life Style Management: Evidence supports the effectiveness of nutrition therapy and physical activity in the prevention and management of Type 2 diabetes. This is reflected in the current ADA standards of medical care (which draw on a detailed evidence-based technical review on nutrition and a more recent review on physical activity) and in the Canadian guideline.

Physical activity[172-176]

Moderate to high levels of physical activity and cardio respiratory fitness are associated with substantial reductions in morbidity and mortality in both men and women and in both type 1 and type 2 diabetes. Large cohort studies have demonstrated that in people with type 2 diabetes, regular physical activity and/or moderate to high cardiorespiratory fitness were associated with reductions in cardiovascular and overall mortality of 45 to 70% over 12 to 14 years.
DPP recommended to engage in physical activity of moderate intensity, such as brisk walking, for at least 150 minutes per week (= 21min per day). EUPASS suggested Health-enhancing physical activity (HEPA) as half an hour a day of physical activity of moderate intensity (see item “Physical Inactivity”).

Education / Empowerment[177-185]

Participation in health promotion programmes with relation to physical activity & weight loss has shown to be effective in reducing the incidence of diabetes mellitus in patients with impaired fasting glucose/ impaired glucose tolerance.
Structured teaching and treatment programs for diabetes may lead to a reduction of weight and redu ction of medication use at unchanged HbA1c levels. A NICE technology appraisal found limited impact on clinical outcomes. Long term effectiveness of this intervention is not yet sufficiently investigated due to a lack of long term studies. A Cochrane meta analysis found that group-based training for self-management strategies in people with type 2 diabetes is effective by improving fasting blood glucose levels, glycated haemoglobin (1.4% after one year, -1% after two years) and diabetes knowledge and reducing systolic blood pressure levels (-5mmHg after 4-6months), body weight and the requirement for diabetes medication.
Hypertension education: Relative risk reduction for end stage events by up to 70%.
Diabetic patient organisation (membership, contact with) Target Agreements (HbA1c, blood pressure, physical activity, diet, smoking, alcohol).
Group-based training for self-management strategies in people with type 2 diabetes is effective by improving fasting blood glucose levels, glycated haemoglobin and diabetes knowledge and reducing systolic blood pressure levels, body weight and the requirement for diabetes medication.

Psychological care / Screening for depression[186-199]

All people with diabetes should be screened for depression and offered appropriate therapy (SIGN recommendation). Psychosocial aspects of diabetes care are included (to varying extents) in the guidelines from the CDA, SIGN, NICE (Type 1), ICSI and, for the first time in 2005, in the ADA standards of care. Depression has been found to be twice as prevalent in people with diabetes compared with the general population and is often under-detected.
Evidence-based guidelines for psychosocial care in adults with diabetes have been published under the auspices of the German Diabetes Association (DDG), indicating the level of evidence for psychological interventions in different problem areas. There is growing evidence that psychological counselling can contribute to improved adherence and psychological outcomes

in people with diabetes. A systematic review and meta-analysis has shown that, overall, psychological interventions are effective in improving glycaemic control in Type 2 diabetes. A n assessment instrument which is already used by some BIRO partners is WHO (Five) wellbeing index.

Health Related Quality of Life[200-212]

Two major types of Health related Quality-of-life (HR-QoL) measures are to consider, overall and disease-specific. Overall HRQoL refers to the patient's sense of his own health and well-being in the broad areas of physical, psychological, and social functioning. When evaluating overall HRQoL, objective health status is of secondary concern; it is the patient's personal perspective on his own well-being that is paramount.

Disease-specific HRQoL refers solely to patients' sense of how the disease in question is compromising their well-being in the three broad areas of physical, psychological, and social functioning. Researchers remark, that at this time, there is no well-accepted measure that comprehensively evaluates the many aspects of diabetes-specific HRQoL. Also, perceived HRQoL will not necessarily be closely tied to biomedical markers of diabetes, so it makes little sense to consider glycemic control or severity of complications as an appropriate gold standard.

The Euroquol has also been used in measuring the QoL of patients with complications, e.g. foot ulcers (Ragnarson et al, see references above). Modelling EuroQol health- related utility values for diabetic complications have also been undertaken by researchers (Bagust et al, see references).

2.1.3.6 Complications

Acute Complications

Acute complications which need to be considered are hypoglycaemia, hypoglycaemia requiring medical attention and hyperglycaemia / ketoacidosis / lactic acidosis.

Eye Complications [213,214]

Blindness due to diabetes is the core indic ator of microvascular pathology in the eyes. Definition of blindness in the different countries varies. Most reports use the legal definition of blindness for a certain country. Laser therapy within three months after the diagnosis of proliferate retinopathy was an indicator suggested by EUCID. After discussion this indicator was removed because of the difficulties collecting the follow-up data 3 months after diagnosis. Documentation can contain: Medical diagnosis (proliferative retinopathy (necessity for laser treatment), retinopathy (mild, severe) non proliferative, maculopathy, diabetic cataract, dry eye. Functional losses: Severe vision loss, partial sightedness (percentage), blindness as well as.

Procedures: Lasertherapy, photocoagulation, cataract operation, VEGF-Therapy (experimental in some countries).

Kidney Damage / Nephropathy [215-221]

Nephropathy represents the second major microvascular complication in persons with diabetes mellitus. Both micro- and macroalbuminuria are stronger predictors of cardiovascular mortality than of end-stage renal failure. Only a minority of patients with microalbuminuria will progress to end-stage renal failure, because death from a cardiovascular cause commonly occurs before renal failure has developed. Control of blood pressure in patients with type 2 diabetes significantly reduces the progression of diabetic kidney disease. If no action is taken micro vascular lesions in the kidneys will lead to renal insufficiency. First signals are the detection of microalbuminuria, followed by an increase in creatinine levels.Chronic kidney disease is defined as either kidney damage or decreased kidney function (decreased glomerular filtration rate (GFR)) for 3 or more months. of GFR varies according to age, sex, and body size. Normal GFR in young adults is approximately 120 to 130 mL/min per 1.73 m^2 and declines with age. The guidelines define kidney failure as either 1) GFR less than 15 mL/min per 1.73 m^2, which is in most cases by signs and symptoms of uremia, or 2) a need to start kidney replacement therapy (dialysis or transplantation).
The urinary albumin:creatinine ratio is a useful measure of renal function used in diabetic renal disease. The urinary albumin: creatinine ratio is measured using the first morning urine sample where practicable. End stage renal disease (ESRF) is defined as Creatinine over 400 mmol/l or previous renal transplant OR GFR < 15 mL/min per 1.73 m^2 OR on dialysis or transplant.

Foot complications [221-243]

Because of the potential for improvement of health and reduction of health-care costs, the evidence surrounding diabetes foot-care has been extensively and formally reviewed many times in recent years. The output from these documents is very consistent in suggesting that formal regular review to detect people at risk, more regular review of those found to be at risk, and intensive management of those developing foot ulceration and infection can produce major returns in avoiding the health and monetary costs of amputation. Providing foot-care education for all patients, with increased intensity for those at higher risk, and vascular interventions where critical ischaemia is identified (or is contributing to ulceration), are also common recommendations arising from the evidence-base.
Documentation may include: Acute ulcer, Healed ulcer, Amputation (above / below ankle), Wagner classification / San Antonio Wound classification, Foot deformities, Charcot, Non-surgical therapy received on foot disease, Regular visits at diabetic foot clinic, Number of patients admitted to hospital with foot related problems.

Neuropathy [244-251]

Discussion of diabetic foot neuropathy screening see item “Foot Screening”. Additional forms of nephropathy are painful sensory neuropathy, autonomic neuropathy and sexual dysfunction. There is general agreement

that stabilizing glycaemic control is important in the medium and longer term, and that tricyclic drugs should be used as first-line therapy for painful neuropathy, although side-effects are common. Exclusion of non-diabetic causes of neuropathy is important because these may account for 10 % of cases of neuropathy in people with diabetes. The range of tests available in clinical and research settings is detailed in two technical reviews. Erectile dysfunction is addressed by three of the guidelines, which draw on evidence from Type 1 as well as Type 2 diabetes. They conclude that the condition is rarely of simple causation, that it is important to consider the possible contribution of other medications and medical conditions, but that the expensive PDE5 inhibitors are worth a trial. The evidence-base on some of the rarer aspects of autonomic neuropathy is weak, including that for gastroparesis, and cardiovascular parasympathetic autonomic neuropathy. In general, other guidelines have relied on conventional wisdom in making recommendations over the management of gastroparesis, orthostatic hypotension, bladder dysfunction, and nocturnal diarrhoea.

Cardiovascular Disease [252-261]

Documentation on Cardiovascular Disease may include Myocardial Infarction / Former myocardial infarction, Coronary heart disease (CHD) as a risk factor, Stroke / Apoplexy, Transient ischaemic attacks (TIA).
Diagnosis of myocardial infarction is based on clear history, clinical findings and typical laboratory tests or ECG changes (EUROCISS definitions should be taken in to account). Stroke (by WHO) is a focal (or at time global) neurological impairment of sudden onset and lasting more than 24 hrs (or leading to death) and of presumed vascular origin (any permanent neurological brain damage, induced by vascular incidents). (EUROCISS definitions should be taken into account).

Peripheral Vascular Disease[262]

Stages for PVD according to Fontaine or Rutherford distinguish asymptomatic, mild/moderate/severe claudication, ischemic rest pain, ulceration or gangrene. Peripheral vascular disease, in addition to peripheral neuropathy and duration of diabetes over 10 years increases the risk for gangrene, foot ulcers and amputation. Myocardial infarction and stroke are increased in patients with diabetes mellitus as documented in many reports. There is a direct relationship between existence of polyneuropathy and/or PVD and the risk for foot lesions and even amputations.

Diagnosis of PVD and treatment is based on anamnesis, clinical symptoms and the respective stage (as outlined above), not on the ABI (Doppler sonography correlates in some cases well with symptoms but is misleading in very arteriosclerotic arteries.)

2.1.3.7 Individual characteristics, demographic and socio-economic factors

Individual characteristics and health status [263,264]

Documentation on individual characteristics may include: Age, gender, ethnicity, age at onset, socio-economic status (employment status, education, white/blue collar worker, income).

Gender perspective: Health services research has shown huge differences in access, process and outcomes between men and women. Socio-economic status should also be recorded with every indicator. The risk for chronic complications increases with diabetes duration. Duration of diabetes increases the risk of CHD death independent of coexisting risk factors.

Population and Socio-economic factors [265,266]

Population criteria for the comparison of regions are: Total population, Median age of population, distribution of age groups, Rate of urbanisation, Life expectancy in years at birth (at the age of 40 etc.) & related indicators, Sick days per year and person, Hospital days per year and person, Mortality (diabetes specific and because of diabetes specific comorbidities). Socio-economic factors to be considered for regional comparisons are: Literacy rate, Total labour force, Total employment, Total unemployment. Deprivation is closely related to health outcomes, but the problem is that it would be measured very differently across the countries. Ultimately it might be good to have a "pan-european" measure of deprivation, which is not yet available.

2.1.3.8 Health system & health care delivery

Health care resources, delivery of care[2 65,267]

Regional comparisons of health care resources: Facilities: hospital beds total (acute care / rehabilitation); Manpower: physicians (GPs, specialists), diabetologists, nurses, diabetes related nurses, pharmacies, ophthalmologists, dieticians, podiatrist, cast technicians; Education of personnel.

Health care delivery: Inpatient care utilisation (days per hospitalisation, hospitalisation rate per 1000 inhabitants); Outpatient care utilization (GP contacts per patient and year); Medicine use/medical aids; Gatekeepers; Availability and utilization of Disease Management Programmes (DMP).

Definition of "diabetologist" is unclear, individual specialist registers in different countries will be available as data sources, but will have different definitions. Disease Management Programmes in best-case scenarios are based on scientifically proven results.

Health care expenditures/financing [265]

The following parameters may be used for regional comparison of health care expenditures/financing: National expenditure on health (% of GDP); Public and private expenditure on health;

Expenditure on medical services (inpatient stays, outpatient care, medical aids, pharmaceuticals); Medical goods dispensed to outpatients; Total health expenditure by age group; Health expenditure by fund source

2.1.4 Data and Documentation

The following parameters may be used to characterize and compare regional data sources[268]:

Recording: electronic, paper, online;
Reliability: Bias, completeness (qualitative description, a numeric measure is not yet available);
Primary aim of data collection : Clinical Documentation; Registry; DiabCare System; Survey; Sentinel Practise Surveillance Network (SPSN); Accounting system; Insurance/Reimbursement; Patient association; Death certification.

2.1.5 Results

As a result of the clinical review, a list of BIRO indicators were selected, then divided for convenience into specific chapters: epidemiology, structural quality, process quality, intermediate outcomes and terminal outcomes.

The complete list of selected BIRO indicators is shown in Table 2.1.1 (a-c).

The review provided also an indication of data items to be collected, shown in Table 2.1.2, which was passed to the further step for the definition of the common dataset and data dictionary.

2.1.6 Conclusion

The BIRO clinical review tried to identify the most relevant concepts for epidemiology, diagnosis and classification of diabetes as well as diabetes treatment and the associated outcomes. In addition, parameters for regional comparisons, the institutional setting and the healthcare system in the region were identified. This clinical review mainly had the aim to identify the distinctive concepts and identify data items to be collected in the regions. Individual cut-offs and treatment recommendations recommended in the guidelines and will be adjusted in the corresponding indicators. The clinical review is the foundation for the definition of the BIRO data set, which will be collected by all participating countries and regions. The discussion in the BIRO Consortium gave very interesting insights on international variations and should provide best

Table 2.1.1 (a) Indicators selected for BIRO: Epidemiology and Structural Quality

Epidemiology

- **Annual Incidence of Type 1 Diabetes in children between 0-14 years of age at diagnosis (clinical) per 100,000 children**
  + Is described in the EUDIP final report and well evaluated.
  + Clear cut-off by choosing age group 0-14 to be more likely to count type 1 diabetes.
- **Prevalence of diabetes mellitus per 1,000**
  No distinction between type 1 and type 2 diabetes for cases when this information comes from prescription data => feasibility without given diabetes type
- **Age at diagnosis by 10 year age bands (incidence)**
  - Source: Patient records, DiabCare => One often has to rely on information given by the patient.

Structural Quality

- **Hospital beds per 100,000 population**
  Not all categories of beds have to be collected separately
- **Physicians employed per 100,000 population**
  + National statistics can provide this information feasibility
- **Number of diabetologists per 100,000 population**
  + Diabetologists play an important role in the process of treating patients with diabetes. (relevance)
  - Definition of a “diabetologist” is unclear. Comparison is difficult. (feasibility) Data should come from national specialist registers
- **Number of doctors who regularly take care of diabetic patients in diabetes clinics in primary or secondary care per 100,000 population**
  Number of doctors who regularly take care of diabetic patients was introduced to estimate whether patients with diabetes are treated by specialized entities or rather in the community
- **Number of diabetes nurses employed per 100,000 population**
  This indicator distinguishes between nurses and specialized diabetes nurses.
- **Number of physicians who offer structured Disease Management Programme (DMP) participations to patients per 1000 patients with diabetes mellitus**
  Availability of a DMP influences the level of structured and evidence based treatment.
- **Portion of diabetic patients enrolled in structured Disease Management Programmes (DMP)**
  Actual patient entrolment in DMP

Table 2.1.1 (b) Indicators selected for BIRO: Process Quality

Process Quality

- **Percentage of patients with one or more HbA1c tests during the last 12 months** + This indicator is one of the six suggested process indicators by the OECD. + For international comparability use the OECD definitions + High importance and scientific soundness.
- **Percentage of patients with one or more Total cholesterol/HDL tests during the last 12 months** EUDIP uses "lipid profile (total chol., LDL, HDL, trigl.) measured within the last 12 months", OECD uses LDL only, BIRO recommends Total chol./HDL chol.
- **Percentage of patients with at least one test for microalbuminuria during the measurement year or who had evidence of medical attention for existing nephropathy**
- **Percentage of diabetes patients who received a dilated eye examination or evaluation of retinal photography by a trained caregiver within the last 12 months** Modified version of the OECD indicator "Percentage of diabetes patients who received a dilated eye examination or evaluation of retinal photography by an ophthalmologist or optometrist during the current year or during the prior year if the patient is at low risk for retinopathy": The time measure "within the last 12 months" is used to be more consistent with the other indicators, the restriction to ophthalmologist or optometrist was removed because of differences in different countries, and the low risk for retinopathy was removed because it is hard to assess.
- **Percentage of diabetes patients receiving at least one foot examination within the last 12 months**
- **Percentage of diabetes patients whose smoking status was ascertained and documented within the last 12 months**
- **Percent with serum creatinine tested in last 12 months**
- **Percentage of patients with diabetes and one or more blood pressure measurements within the last 12 months** Blood pressure control is clinically important
- **Percentage of diabetes patients with clinically diagnosed hypertension who receive antihypertensive medication** Hypertension is defined by either hypertension treatment or blood pressure > 140/90
- **Percentage of patients with diabetes specific education at least once before** Diabetic specific education can lead to better outcome in patients with diabetes.
- **Type of oral therapy (distribution of agents) in patients with diabetes type 2** Which oral anti diabetic agents are used? Interesting for treatment processes, maybe even for research. If distribution is not feasible, change this indicator to "Portion of OAD treated patients"
- **Portion of patients with OAD therapy in patients with diabetes type 2** Relevant for type 2.
- **Portion of patients treated with insulin among patients with diabetes** Relevant for type 2.
- **Portion of patients treated with insulin in combination with OADs among patients with diabetes** Relevant for type 2.
- **Percentage of insulin treated patients with pump therapy**
- **Average number of insulin injections per day in insulin treated patients**It is recommended to display this indicator as a distribution (histogram)
- **Portion of diabetic patients treated with diet only** Item is present in DiabCare data set + this information can not be deduced from drug prescriptions and is therefore interesting to assess seperately
- **Portion of diabetes patients with anti hypertensive treatment**
- **Average number of antihypertensive agents used per diabetes patient with anti hypertensive treatment**
- **Portion of diabetes patients with lipid lowering medication** Important process in treatment practice
- **Percent of patients with diabetes performing self-monitoring of blood glucose/ urine testing** Important process for patient empowerment
- **Percent of patients with clinically diagnosed CVD and diabetes who are treated with anti-platelet therapy**

Table 2.1.1 (c) Indicators selected for BIRO: Outcomes

Intermediate Outcomes

- **Percentage of patients with most recent HbA1c level >9.0% (poor control)**OECD indicator definition used for international comparability.
- **Percentage of patients with most recent HbA1c level >7,5%** Modification for HbA1c threshold >7,5% (used in EUDIP) introduced in BIRO meeting in Malta -> display HbA1c distribution in addition
- **Percentage of patients with Total-Chol / HDL-Chol < 4.5** Modification of OECD indicator "Percentage of patient with most recent LDL<130 mg/dl"because LDL is problematic in practice. A value of 4.5 is the target value for diabetic patients, a value >8.0 means high risk, may be introduced later
- **Percentage of patients with most recent blood pressure <140/90 mmHg** OECD indicator definition used for international comparability. Some guidelines use lower threshold value – for outcome quality measurement 140/90 is appropriate to show distribution in addition.It was also recommended to analyze percentage of patients with SBP >140 separately from percentage patients with DBP >90
- **Percentage of patients with BMI ≥ 30 kg/m2** Overweight and obesity are considered as a major risk factor for developing micro and macro vascular complications. Overweight is defined as BMI ≥ 25 kg/m2. Obesity is defined as BMI ≥ 30 kg/m2
- **Percentage of persons with diabetes and proliferate retinopathy and/or maculopathy who had a fundus inspection in the last 12 months** EUDIP defines retinopathy as the presence of the growth of new blood vessels on the retina and the posterior surface of the vitreous. Reimbursement codes in some countries offer codes for laser treatment. Additionally to the ICD-Codes a validation of laser treatment is possible.
- **Percentage of patients with eye laser treatment ever** Intermediate outcome for retinopathy Interesting to compare how health care systems deal with retinopathy
- **Percentage with microalbuminuria in last 12 months (among those who have been tested)** that gives a rate of "newly found" patients with microalbuminuria- Difficult to compare because it is unclear who has been screened
- **Rate of current smokers among diabetes patients** Smoking is an important risk factor.
- **Rate of patients with current alcohol abuse/dependence**
- **Former or current foot ulceration** Classification of "Foot on Risk"

Terminal Outcomes

- **Annual incidence of blindness in patients with diabetes (among those visited during the last 12 months)** + Easier to assess than blindness incidence'due to diabetic retinopathy' is hard to assess The original EUDIP indicator is 'Annual incidence of blindness due to diabetic retinopathy/total annual incidence of blindness' because after discussion the BIRO group found it hard to record the reason for blindness
- **Annual incidence of dialysis and/or transplantation (renal replacement therapy in patients with diabetes**
- **ESRD in Persons with Diabetes** EUDIP uses two related indicators in connection with epidemiology of complications: 'Percent with ESRF in last 12 months in total population' and 'Prevalence (stock) of dialysis/ transplantation (renal replacement therapy) in patients with diabetes'
- **Annual incidence of amputations above the ankle** EUDIP definition "amputations above the ankle" was preferred. OECD suggestion is "Lower extremity amputation rates", major (above or below knee) amputations
- **Annual incidence of stroke in patients with diabetes** - stroke and diabetes have to be known in combination
- **Annual Incidence of myocardial infarction in patients with diabetes** - stroke and diabetes have to be known in combination
- **Annual death rate per 100,000 populations in patients, who have as primary or secondary cause of death, diabetes mellitus, adjusted for standard European population.** Major indicator for diabetes complications. EUDIP suggests the linkage of the death rate with gender and age. Data sources are national registries. - diabetes often is not well recorded as primary or secondary cause of death

Table 2.1.2 Data Items required for selected indicators

- Patient data
- Year of birth
- Sex
- Height
- Epidemiology
- Diabetes type
- Newly diagnosed diabetes
- Year of diagnosis
- Total number of children between 0-14 yrs
- Diabetes y/n
- Total number of general population in area
- Age at diagnosis
- Structural quality
- Hospital beds in area
- Physicians employed in area
- Number of diabetologists in area
- Number of doctors who regularly take care of diabetic patients in diabetes clinics in primary or secondary care in area
- Number of diabetes nurses employed in area
- Number of physicians offering structured Disease Management Programs (DMP) in area
- Patient enrolled in structured Disease Management Program (DMP)
- Process quality
- Number of clinically diagnosed diabetes patients in the area
- HbA1c tested within last 12 months
- Total Chol/HDL tested within last 12 months
- Microalbuminuria tested within last 12 months
- Medical attention for nephropathy within last 12 months
- Dilated eye examination or evaluation of retinal photography by a trained caregiver within the last 12 months
- At least one foot examination within last 12 months
- Smoking status ascertained within last 12 months
- Serum creatinine tested within last 12 months
- One or more blood pressure measurements within last 12 moths
- Hypertension prevalent within last 12 months
- Received antihypertensive medication within last 12 months
- Diabetes specific education at least once before
- Treatment with diet only
- Treatment with sulfonylurea y/n within last 12 months
- Treatment with biguanides within last 12 months
- Treatment with glucosidase inhibitors within last 12 months
- Treatment with glitzones y/n within last 12 months
- Treatment with glinides y/n within last 12 months
- Treatment with insulin within last 12 months
- Pump therapy within last 12 months
- Average number of insulin injections per day
- Self monitoring of blood/urine glucose within last 12 months
- Clinically diagnosed CVD
- Treatment with anti-platelet therapy within last 12 months
- Treatment with lipid lowering medication within last 12 months
- Outcome - intermediate outcomes
- Most recent HbA1c level (number)
- Most recent Total Cholesterol
- Most recent HDL Cholesterol
- Most recent systolic blood pressure
- Most recent diastrolic blood pressure
- Most recent Weight
- Most recent BMI (Calculated from Weight, Height)
- Retinopathy prevalent within last 12 months
- Maculopathy prevalent within last 12 months
- Fundus inspection within last 12 months
- Eye laser treatment ever
- Positive testing for urinary albumin within last 12 months
- Smoking currently
- Current alcohol abuse/dependence
- Former or current foot ulceration
- Outcome - terminal outcomes
- Blindness prevalent
- Blindness newly diagnosed within last 12 months
- Dialysis and/or transplantation new within last 12 months
- ESRD prevalent
- History of amputation above ankle
- Amputation above ankle new within last 12 months
- History of stroke
- Stroke new within last 12 months
- History of myocardial infarction
- Myocardial infarction new within last 12 months
- Death within last 12 months

possible feasibility for an implementation in European countries.

To reflect ongoing advancements in the clinical domain, the review should periodically be updated. Main changes in guidelines and recommendations may have an influence on BIRO indicators. However, only in exceptional cases individual studies will have to be considered in an update. A time frame of 3 years is suggested as appropriate for an update of the clinical review and as a consequence for the BIRO data set. It is suggested that for an update of the clinical review current and updated guidelines are screened for changes in core processes of diabetes care as well as epidemiological and outcome measures. These changes lead to additional or modified candidate indicators and additional or modified data items which have to be collected. The changes should be discussed and documented by clinical experts in a structured way and added to the documentation of the indicator in the BIRO system.

References

1. Greenfield S, Nicolucci A, Mattke S: Selecting Indicators for the Quality of Diabetes Care at the Health System Level in the OECD Countries. OECD health Technical Papers, No.15: 2004
2. Fleming BB, Greenfield S., Engelgau MM, et al., The Diabetes Quality Improvement Project: moving science into health policy to gain an edge on the diabetes epidemic Diabetes Care, Vol. 24(10), pp.1815-1820.
3. National diabetes quality improvement alliance (2003), Measures and supporting document, available at:http://www.nationaldiabetesalliance.org/measures.html
4. Alberti K, Zimmet P, Shaw J. Metabolic syndrome a new world-wide definition. A Consensus Statement from the International Diabetes Federation, Diabet. Med. 23, 469 480
5. Ferrannini E, Camastra S. Relationship between impaired glucose tolerance, non insulin dependent diabetes mellitus and obesity (1998) EJCI 28 S2: 3-7
6. Edelstein SL, Knowler WC, Bain RP, Andres R, Barrett-Connor EL, Dowse GK, Haffner SM, Pettitt DJ, Sorkin JD, Muller DC, Collins VR, Hamman RF (1997) Predictors of progression from impaired glucose tolerance to NIDDM: an analysis of six prospective studies. Diabetes 46: 701-710
7. Nosikov A and Gudex C (Eds.) Developing Common Instruments for Health Surveys, IOS Press, 2003 Chapter 6.Development of a common instrument for physical activity. (EUROHIS).
8. European Physical Activity Surveillance System (EUPASS) Final report to the European Commission (DG SANCO F/3, HEALTH MONITORING PROGRAMME), March 2001
9. Astrup A. Healthy lifestyles in Europe: prevention of obesity and type II diabetes by diet and physical activity. Public Health Nutr.;4(2B):499-515.
10. Liao D, Asberry PJ, Shofer JB, Callahan H, Matthys C, Boyko EJ, Leonetti D, Kahn SE, Austin M, Newell L, Schwartz RS, Fujimoto WY. Improvement of BMI, body composition and body fat distribution with lifestyle modifications in Japanese Americans with impaired glucose tolerance. Diabetes Care 25: 1504-1511
11. Nosikov A and Gudex C (Eds.) Developing Common Instruments for Health Surveys, IOS Press, 2003 Chapter 9.Development of a common instrument for use of preventive health care (EUROHIS).
12. Hu FB, van Dam RM, Liu S.Diet and the risk of type II diabetes: the role of types of fat and carbohydrate. Diabetologia 44: 805-817
13. Trichopoulou A. The DAFNE databank as a simple tool for nutrition policy. DAta Food NEtworking . Public Health Nutr. 4 : 1187-1198.
14. Brussaard JH, Lowik MR, Steingrimsdottir L, Moller A,, Kearney J, De Henauw S, Becker W; The EFCOSUM Group A European food consumption survey method--conclusions and recommendations. Eur J Clin Nutr.56 : S2 : S89-94
15. Ko GT, Chan JC, Tsang LW, Li CY, Cockram CS. Glucose intolerance and other cardiovascular risk factors in Chinese women with a history of gestational diabetes. Aust N Z J Obstet Gynaecol 39: 478-483.
16. Verma A, Boney CM, Tucker R, Vohr BR. Insulin resistance syndrome in women with prior history of gestational diabetes mellitus . J Endocrinol Metab 87: 3227-3235
17. Gestational diabetes mellitus. American Diabetes Association Diabetes
18. World Health Organization. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Report of a WHO Consultation. Part 1: Diagnosis and Classification of Diabetes Mellitus. Geneva: WHO Department of Noncommunicable Disease Surveillance, 1999: 1-59. http://www.who.int
19. Manley SM, Meyer LC, Neil HAW, Ross IS, Turner RC, Holman RR. Complications in newly diagnosed type 2 diabetic patients and their association with different clinical and biochemical risk factors. UKPDS 6. Diabetes Res 1990; 13: 1-11.
20. UKPDS Group. UK Prospective Diabetes Study 30: Diabetic retinopathy at diagnosis of type 2 diabetes and associated risk factors. Arch Ophthalmol 1998; 116: 297-303.
21. Colagiuri S, Cull CA, Holman RR. Are lower fasting plasma glucose levels at diagnosis of type 2 diabetes associated with improved outcomes? UKPDS 61. Diabetes Care 2002; 25: 1410-17.
22. The Expert Committee on the diagnosis and classification of diabetes mellitus. Follow-up report on the diagnosis of diabetes mellitus. Diabetes Care 2003; 26: 3160-67.
23. Alberti KGMM, Zimmet PZ for the WHO Consultation. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Part 1: Diagnosis and Classification of Diabetes Mellitus. Provisional Report of a WHO Consultation. Diabet Med 1998;15:539-53
24. Hans Reinauer, Philip D. Home, Ariyur S. Kanagasabapathy, Claus-Chr. Heuck, World Health Organization, Laboratory Diagnosis and Monitoring of Diabetes mellitus, 2002, available at: http://whqlibdoc.who.int/hq/2002/9241590483.pdf
25. Department of Noncommunicable Disease Surveillance. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Geneva: WHO; 1999.
26. Diagnostic criteria differ from non-pregnant state; venous plasma glucose>5.5mmol/l fasting OR

≥9.0mmol/l 2 hours after 75g OGTT; SIGN 55 guideline 2001
27. Green A, Gale EAM, Patterson CC, the EURODIAB Subarea A Study Group (1992) Incidence of childhood onset insulin dependent diabetes: The EURODIAB ACE study. Lancet 339: 905-909
28. Levy-Marchal C, Patterson CC, Green A, on behalf of the EURODIAB ACE Study Group (1995) Variation in age distribution and seasonality at diagnosis of childhood IDDM in Europe. Diabetologia 38: 823-830
29. De Lissovoy G, Ganoczy DA, Ray NF (2000), Relationship of hemoglobin A1c, age of diabetes diagnosis, and ethnicity to clinical outcomes and medical costs in a computer simulated cohort of persons with type 2 diabetes. Am J Manag Care 6, 573 584
30. de Boer IH, Kestenbaum B, Rue TC, Steffes MW, Cleary PA, Molitch ME, Lachin JM,Weiss NS, Brunzell JD, for the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) Study Research Group, Insulin Therapy, Hyperglycemia, and Hypertension in Type 1 Diabetes Mellitus, Arch Intern Med. 2008;168:1867-1873.
31. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA 10-Year Follow-up of Intensive Glucose Control in Type 2 Diabetes N Engl J Med. Published online 2008 Sep 10
32. The ACCORD Study Group. Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial: design and methods. Am J Cardiol 2007;99(Suppl):21i 33.
33. Gerstein HC, Riddle MC, Kendall DM, et al. Glycemia treatment strategies in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. Am J Cardiol 2007;99(Suppl):34i 43.
34. National Heart, Lung and Blood Institute, Action to Control Cardiovascular Risk in Diabetes (ACCORD) Trial, February 6, 2008. Available at www.nhlbi.nih.gov/ health/prof/heart/other/accord/index.htm [Accessed March 2008].
35. The Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, et al. Effects of intensive glucose lowering in type 2 diabetes. New Engl J Med 2008;358:2545 59.
36. ADVANCE Collaborative Group, Patel A, MacMahon S, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. New Engl J Med 2008;358:2560 72.
37. Duckworth W, Abraira C, Moritz T, et al., on behalf of the VADT Investigators. Glucose control and vascular complications in veterans with type 2 diabetes. N Engl J Med 2009;360:129-39.
38. Abraira C, Colwell J, Nuttall F, et al. Cardiovascular events and correlates in the Veterans Affairs Diabates Feasiility Trial: Veterans Affairs Cooperative Study Group on Glycemic Control and Complications in Type II Diabetes. Arch Intern Med 1997;157:181-8
39. UK prospective Diabetes Study Group 33 (1998). Intensive blood glucose control with sulfonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998:352: 837-853
40. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). Lancet 1998;352:854-65
41. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, Hadden D, Turner RC, Holman RR. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ. 2000; 321: 405-12
42. Diabetes Control and Complications Trial Research Group (1993). The effect of intensive treatment on diabetes on the development and progression of long term complications in type 1 diabetes mellitus N Engl J Med 329 : 977-986
43. Mahlhauser, I., Overmann, H., Bender R, et al. Predictors of mortality and end stage diabetic complications in patients with type 1 diabetes mellitus on intensified insulin therapy, Daibet Med 2000, 17: 727 34.
44. Sibley SD, Thomas W, de Boer I, Brunzell JD, Steffes MW. Gender and elevated albumin excretion in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) cohort: role of central obesity. Am J Kidney Dis. 2006 Feb;47(2):223-32.
45. Jenkins AJ, Lyons TJ, Zheng D, Otvos JD, Lackland DT, McGee D, Garvey WT, Klein RL; DCCT/EDIC Research Group. Lipoproteins in the DCCT/EDIC cohort: associations with diabetic nephropathy. Kidney Int. 2003 Sep;64(3):817-28.
46. Alicia J. Jenkins, Timothy J. Lyons, Deyi Zheng, James D. Otvos, Daniel T. Lackland, Daniel McGee, W. Timothy Garvey, MD, Richard L. Klein and The DCCT/EDIC Research Group: Serum Lipoproteins in the Diabetes Control and Complications Trial/Epidemiology of Diabetes Intervention and Complications Cohort. Associations with gender and glycemia. Diabetes Care 26:810-818, 2003
47. Catherine L. Martin, James Albers, William H. Herman, Patricia Cleary, Barbara Waberski, Douglas A. Greene, Martin J. Stevens, and Eva L. Feldman. Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Intervention and Complications (EDIC) Research Group. Neuropathy Among the Diabetes Control and Complications Trial Cohort 8 Years After Trial Completion. Diabetes Care 29:340-344, 2006
48. Michael W. Steffes, Shalamar Sibley, Melissa Jackson, and William Thomas. fl-Cell Function and the Development of Diabetes-Related Complications in the Diabetes Control and Complications Trial Diabetes Care 26:832-836, 2003

49. Haffner SM, Lehto S, Rˆnnemaa T, et al. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med 1998;339:229-34
50. Abraira C, Colwell J, Nuttall F, et al. Cardiovascular events and correlates in the Veterans Affairs Diabates Feasiility Trial: Veterans Affairs Cooperative Study Group on Glycemic Control and Complications in Type II Diabetes. Arch Intern Med 1997;157:181-8
51. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). Lancet 1998;352:854-65
52. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352:837-53
53. Adler AI, Stratton IM, Neil HAW, et al. Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study. BMJ 2000;321:412-9
54. UK Prospective Diabetes Study Group. Quality of Life in Type 2 Diabetic Patients is affected by Complications but not by intensive Policies to improve Blood Glucose or Blood Pressure Control (UKPDS 37). Diabetes Care 1999;22:1125-36
55. UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. BMJ 1998;317:703-13
56. Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure (2003). The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure. Hypertension, 2003, 42, 1206 1252.
57. Grossman E, Messerli FH, Goldbourt U. High Blood Pressure and Diabetes Mellitus. Arch Intern Med 2000;160:2447-58
58. Kannel WB, McGee DL. Diabetes and cardiovascular disease: the Framingham study. JAMA 1979; 241: 2035-38
59. New Zealand Guidelines Group (2003) Evidence-based best practice guidelines, Review date 2006, ISBN: 0-476-00092-0.
60. The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) (2001). Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection evaluation and treatment of high blood cholesterol in adults (adult treatment panel III) JAMA 285: 2486-2497
61. Pontrelli L, Parris W, Adeli K, Cheung RC (2002) Atorvastatin treatment beneficially alters the lipoprotein profile and increases low-density lipoprotein particle diameter in patients with combined dyslipidemia and impaired fasting glucose / type 2 diabetes. Metabolism 51(3):334-42
62. Sacco RL (2002) Reducing the risk of stroke in diabetes: what have we learned that is new? Diabetes Obes Metab 4 : S1;S27-34
63. Crespin SR (2001): What does the future hold for Diabetic dyslipidaemia? Acta Diabetol 38: S1;S21-S26
64. Vijan S, Hayward RA; American College of Physicians. Pharmacologic lipid-lowering therapy in type 2 diabetes mellitus: background paper for the American College of Physicians. Ann Intern Med. 2004; 40:650-8
65. W‰gner AM et al: Inaccuracy of Calculated LDL-Cholesterol in Type 2 Diabetes: Consequences for Patient Risk Classification and Therapeutic Decisions Clinical Chemistry 46: 1830-1832, 2000
66. Jarrett RJ, Viberti GC, Argyropoulos A, Hill RD, Mahmud U, Murrells TJ: Microalbuminuria predicts mortality in non-insulin dependent diabetics, Diaet Med 1984;1:17-19
67. Mogensen CE: Microalbuminuria predicts clinical proteinuria and early mortality in maturity onset diabetes (NEJM 1984;310:356-360)
68. NICE MeReC Briefing 2004;26:1-8 download from www.nice.org 05/07/06
69. Will JC, Galuska DA, Ford ES, Mokdad A, Calle EE: Cigarette smoking and diabetes mellitus: evidence of a positive association from a large prospective cohort study. Int J Epidemiol 30 (3) (2001) 540-546.
70. Ko GT, Chan JC, Tsang LW, Critchley JA, Cockram CS: Smoking and diabetes in Chinese men. Postgrad Med 77 (910) (2001) 551.
71. Manson JE, Ajani UA, Liu S, Nathan DM, Hennekens CH: A prospective study of cigarette smoking and the incidence of diabetes mellitus among US male physicians. Am J Med 109 (7) (2000) 538-542.
72. Rimm EB, Chan J, Stampfer MJ, Colditz GA, Willett WC: Prospective study of cigarette smoking, alcohol use, and the risk of diabetes in men. BMJ (England) 310 (1995) 555-559.
73. Kawakami N, Takatsuka N, Shimizu H, Ishibashi H: Depressive symptoms an occurrence of type 2 diabetes among japanese men. Diabetes Care 22 (1999) 1071-1076.
74. Dierkx RI, van de Hoek W, Hoekstra JB, Erkelens DW: Smoking and diabetes mellitus. Neth J Med 48 (4) (1996) 150-162.

75. Beach KW, Brunzell JD, Strandness DE Jr: Prevalence of severe arteriosclerosis obliterans in patients with diabetes mellitus. Relation to smoking and form of therapy. Arteriosclerosis 2 (4) (1982) 275-280. Evidenzklasse
76. Wei M, Gaskill SP, Haffner SM, Stern MP: Effects of diabetes and level of glycemia on all-cause and cardiovascular mortality. The San Antonio Heart Study. Diabetes Care 21 (1998) 1167-1172.Sawicki PT, Didjurgeit U, M¸hlhauser I, Berger M: Behaviour therapy versus doctor's anti-smoking advice in diabetic patients. J Int Med 234 (4) (1993) 407-409.
77. Chaturvedi N, Stephenson JM, Fuller JH: The relationship between smoking and microvascular complications in the EURODIAB IDDM complications study. Diabetes Care 18 (6) (1995) 785-792.
78. Holl RW, Grabert M, Heinze E, Debatin KM: Objective assessment of smoking habits by urinary nicotinine measurement in adolescents and young adults with type 1 diabetes. Reliability of reported cigarette consumption and relationship to urinary albumin excretion. Diabetes Care 21 (1998) 787-791.
79. Uchimoto S, Tsumura K, Hayashi T, Suematsu C, Endo G, Fujii S, Okada K: Impact of cigarette smoking on the incidence of type 2 diabetes mellitus in middle-aged Japanese men: the Osaka Health Survey. Diabet Med 16 (11) (1999) 951-955.
80. Mehler PS, Jeffers BW, Biggerstaff SL, Schrier RW: Smoking as a risk factor for nephropathy in non-insulin-dependent diabetes. J Gen Int Med 13 (12) (1998) 842-845.
81. Biesenbach G, Zazgornik J: Influence of smoking on the survival rate of diabetic patients requiring hemodialysis. Diabetes Care 19 (6) (1996) 625-628.
82. Ikeda Y, Suehiro T, Takamatsu K, Yamashita H, Tamura T, Hashimoto K: Effect on the prevalence of albuminuria in Japanese men with non-insulin-dependent diabetes mellitus. Diabetes Res Clin Pract 36 (1) (1997) 57-61.
83. Chase HP, Garg SK, Marshall G, Berg CL, Harris S, Jackson WE, Hamman RE: Cigarette smoking increases the risk of albuminuria among subjects with type 1 diabetes. JAMA 265 (5) (1991) 614-617.
84. Biesenbach G, Zazgornik J: Influence of smoking on the survival rate of diabetic patients requiring hemodialysis. Diabetes Care 19 (6) (1996) 625-628.
85. Stegmayr BG: A study of patients with diabetes mellitus (type 1) and end-stage renal failure: tobacco usage may increase risk of nephropathy and death. J Int Med 228 (2) (1990) 121-124.
86. Moss SE, Klein R, Klein BE: Cigarette smoking and ten-year progression of diabetic retinopathy. Ophthalmology 103 (9) (1996) 1438-1442.
87. Eadington DW, Patrick AW, Frier BM: Association between connective tissue change and smoking habit in type 2 diabetes and in non-diabetic humans. Diabetes Res Clin Pract 11 (2) (1991) 121-125.
88. Haire-Joshu D, Glasgow RE, Tibbs TL (1999) Smoking and Diabetes. Diabetes Care 22 1887-1898
89. Mikhailidis DP, Papadakis JA, Ganotakis ES (1998) Smoking, Diabetes and hyperlipidaemia. J R Soc Health 118 :91-93
90. Lorenzo C, Serrano-Rios M, Martinez-Larrad MT, Gabriel R, Williams K, Gonzalez-Villalpando C, Stern MP, Hazuda HP, Haffner S: Prevalence of hypertension in Hispanic and non-Hispanic white populations. Hypertension 39 (2) (2002) 203-208.
91. Laws A, Marcus EB, Grove JS, Curb JD: Lipids and lipoproteins as risk factors for coronary heart disease in men with abnormal glucose tolerance: the Honolulu Heart Program. J Int Med 234 (5) (1993) 471-478.
92. Adler AI, Boyko EJ, Ahroni JH, Stensel, Forsberg RC, Smith DG: Risk factors for diabetic peripheral sensory neuropathy. Results of the Seattle Prospective Diabetic Foot study. Diabetes Care 20 (7) (1997) 1162-1167
93. K‰stenbauer T, Sauseng S, Sokol G, Auinger M, Irsigler K: A prospective study of predictors for foot ulceration in type 2 diabetes. J AM Podiatr Med Assoc 91 (7) (2001) 343-350.
94. Martin-Morales A, Sanchez-Cruz JJ, Saenz de Tejada I, Rodriguez-Vela L, Jimenez-Cruz JF, Burgos-Rodriguez R: Prevalence and independent risk factors for erectile dysfunction in Spain: results of the Epidemiologia de la Disfuncion Erectil Masculina Study. J Urol 166 (2) (2001) 569-574.
95. Stepka M, Rogala H, Czyzyk A: Hypoglycemia: A major problem in the management of diabetes in the elderly. Aging 5 (2) (1993) 117-121.
96. Cusi K, Consoli A: Alcoholic ketoacidosis and lactic acidosis. Diabetes Reviews 2 (2) (1994) 195-208.
97. Balkau B, Eschwege E, Ducimetiere P, Richard JL, Warnet JM: The high risk of death by alcohol related diseases in subjects diagnosed as diabetic and impaired glucose tolerance: the Paris Prospective Study after 15 years of follow-up. J Clin Epidemiol 44 (6) (1991) 465-474.
98. Keilman PA: Alcohol consumption and diabetes mellitus mortality in different countries. Am J Public Health 73 (11) (1983) 1316-1317.
99. Lindegard B, Hillbom M: Associations between brain infarction, diabetes and alcoholism: Oberservations from the Gothenburg population cohort study. Acta Neurol Scand 75 (3) (1987) 195-200.
100. The ICD-10 Classification of Mental and Behavioural Disorders World Health Organization, Geneva, 1992 F10.2 Alcohol Dependence Syndrome
101. Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV), published by the

American Psychiatric Association, Washington D.C., 1994
102. Anderson, P. Management of Drinking Problems. WHO Regional Publications, European Series, No. 32, Copenhagen, 1990.
103. Saunders, J.B.; Aasland, O.G.; Amundsen, A.; Grant, M. Alcohol Consumption and Related Problems Among Primary Health Care Patients. WHO Collaborative Project on Early Detection of Persons With Harmful Alcohol Consumption - I. Addiction, 88, 349-362, 1993.
104. Ng RS, Darko DA, Hillson RM. Street drug use among young patients with Type 1 diabetes in the UK. Diabet Med. 2004 Mar;21(3):295-6.
105. Spraul M, Raunest J, Reike H. Der diabetische Fufl. In Berger M: Diabetes mellitus 2000
106. SIGN (Scottish Intercollegiate Guidelines Network). Management of diabetic foot disease 2001
107. Morbach S, M¸ller E, Reike H, Risse A, Spraul M. Diabetisches Fuflsyndrom Praxis-Leitlinie, DDG, Diabetes und Stoffwechsel 13/2004, 73-76
108. Morbach S, M¸ller E, Reike H, Risse A, Spraul M. Diagnostik, Therapie, Verlaufskontrolle und Pr‰vention des diabetischen Fuflsyndroms. Evidenzbasierte Leitlinien, Deutsche Diabetes-Gesellschaft 2004
109. International Consensus on the Diabetic Foot. International Working Group on the Diabetic Foot 1999
110. Scottish Intercollegiate Guidelines Network. SIGN 55. Management of Diabetes, 2001. http://www.sign.ac.uk
111. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Canadian Diabetes Association 2003 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Canadian Journal of Diabetes 2003; 27(Suppl 2): S76- S80. http://www.diabetes.ca
112. Klein R, Klein BEK, Moss SE, Davis MD, DeMets DL. The Wisconsin epidemiologic study of diabetic retinopathy III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. Arch Ophthalmol 1984; 182: 527-32.
113. The National Collaborating Centre for Chronic Conditions. Type 1 Diabetes in Adults. National clinical guideline for diagnosis and management in primary and secondary care. http://www.rcplondon.ac.uk/pubs/books/DIA/index.asp
114. Weber B, Burger W, Hartmann R, Hovener G, Malchus R, Oberdisse U. Riskfactors for the development of retinopathy in children and adolescents with type 1 (insulin-dependent) diabetes mellitus. Diabetologia: 29: 23-29.
115. American Diabetes Association, Diabetic retinopathy .Diabetes Care S1.: S90-93.
116. Kronsbein P, J^rgens V, M¸hlhauser I, et al. Evaluation of a structured treatment and teaching programme on non-insulin-dependent diabetes. Lancet 1988;ii:1407-10
117. The ACCORD Study Group. Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial: design and methods. Am J Cardiol 2007;99(Suppl):21i 33.
118. Gerstein HC, Riddle MC, Kendall DM, et al. Glycemia treatment strategies in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. Am J Cardiol 2007;99(Suppl):34i 43.
119. National Heart, Lung and Blood Institute, Action to Control Cardiovascular Risk in Diabetes (ACCORD) Trial, February 6, 2008. Available at www.nhlbi.nih.gov/ health/prof/heart/other/accord/index.htm.
120. The Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, et al. Effects of intensive glucose lowering in type 2 diabetes. New Engl J Med 2008;358:2545 59.
121. ADVANCE Collaborative Group, Patel A, MacMahon S, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. New Engl J Med 2008;358:2560 72.
122. Duckworth W, Abraira C, Moritz T, et al., on behalf of the VADT Investigators. Glucose control and vascular complications in veterans with type 2 diabetes. N Engl J Med 2009;360:129-39.
123. The University Group Diabetes Program. Effects of hypoglycaemic agents on vascular complications in patients with adult-onset diabetes. VIII: Evaluation of insulin therapy: final report. Diabetes 1982;31(Suppl. 5):1-26
124. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352:837-53
125. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). Lancet 1998;352:854-65
126. The University Group Diabetes Program. Effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. VIII: Evaluation of insulin therapy: final report. Diabetes 1982;31(Suppl. 5):1-26
127. Leibowitz G, Cerasi E. Sulfonylurea treatment of NIDDM patients with cardiovascular disease: a mixed blessing? Diabetologia 1996;39:503-14
128. Cleveland JC et al. Oral sulfonylurea hypoglycaemic agents prevent ischemic preconditioning in human myocardium. Circulation 1997;96:29-32
129. Ohkubo Y, Kishikawa H, Araki E, Miyata T, Isami S, Motoyoshi S et al. Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomised prospective 6-years study. Diabetes Research and Clin Pract 1995;

28:103-117
130. Staessen JA, Wang JG, Thijs L. Cardiovascular protection and blood pressure reduction: a meta-analysis. Lancet 2001;358:1305-15
131. Diabetes Control and Complications Trial Research Group (1993). The effect of intensive treatment on diabetes on the development and progression of long term complications in type 1 diabetes mellitus N Engl J Med 329 : 977-986
132. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352:837-53
133. Blood Pressure Lowering Treatment Trialists' Collaboration (BPLTTC). Effects of ACE inhibitors, calcium antagonists, and other blood-pressure-lowering drugs: results of prospectively designed overviews of randomised trials. Lancet 2000;356:1955-64
134. Staessen JA, Wang JG, Thijs L. Cardiovascular protection and blood pressure reduction: a meta-analysis. Lancet 2001;358:1305-15
135. Psaty BM, Smith NL, Siscovick DS, et al. Health outcomes associated with antihypertensive therapies used as first-line agents. A systematic review and meta-analysis. JAMA 1997;277:739-45
136. Curb JD, Pressel SL, Cutler JA, et al. Effect of diuretic-based antihypertensive treatment on cardiovascular disease risk in older patients with isolated systolic hypertension. JAMA 1996;276:1886-92
137. Borhani NO, Mercuri M, Borhani PA, et al. Final outcome results of the multicenter isradipine diuretic atherosclerosis study (MIDAS). A randomised controlled trial. JAMA 1996;276:785-91
138. The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major cardiovascular events in hypertensive patients randomized to Doxazosin vs Chlorthalidone. The Antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). JAMA 2000;283:1967-75
139. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs. diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). JAMA. 2002 Dec 18;288(23):2981-97. Erratum in: JAMA 2003 Jan 8;289(2):178. JAMA. 2004 May 12;291(18):2196
140. Freemantle N; Cleland J, Young P, Mason J, Harrison J. ≤ Blockade after myocardial infarction: systematic review and meta regression analysis. BMJ 1999;318:1730-7
141. Kjekshus J, Glipin E, Cali G, Blackey, Henning H, Ross J. Diabetic patients and betablockers after acute myocardial infarction. Eur Heart J 1990;11:43-50
142. Mangano DT, Layug EL, Wallace A, Tateo I,. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. N Engl J Med 1996;335:1713-20
143. Casas JP, Chua W, Loukogeorgakis S et al; Effect of inhibitors of the renin-angiotensin system and other antihypertensive drugs on renal outcomes: systematic review and meta-analysis, Lancet 2005; 366:2026-33.
144. Lindholm LH, Ibsen H, Dahlf B, Devereux RB, Beevers G, de Faire U, Fyhrquist F, Julius S, Kjeldsen SE, Kristiansson K, Lederballe-Pedersen O, Nieminen MS, Omvik P, Oparil S, Wedel H, Aurup P, Edelman J, Snapinn S, for the LIFE study group. Cardiovascular morbidity and mortality in patients with diabetes in the Losartan intervention for endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. Lancet 2002;359:1004-10
145. Dahlˆf B, Devereux RB, Kjeldsen SE, Julius S, Beevers G, de Faire U, Fyhrquist F, Ibsen H, Kristiansson K, Lederballe-Pedersen O, Lindholm LH, Nieminen MS, Omvik P, Oparil S, Wedel H, for the LIFE Study group. Cardiovascular morbidity and mortality in the losartan intervention for endpoint reduction in hypertension study. Lancet 2002;359:995-1003
146. Staesson JA, for the Systolic Hypertension in Europe Trial Investigators. Predicting cardiovascular risk using conventional vs. ambulatory blood pressure in older patients with systolic hypertension, JAMA 1999;282:539-46
147. European Society of Hypertension European Society of Cardiology Guidelines for the management of arterial hypertension. J Hypertens 2003; 21:1011-53
148. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Lancet 2002;360:7-22
149. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. N Engl J Med. 1998; 339:1349-57
150. Vijan S, Hayward RA; American College of Physicians. Pharmacologic lipid-lowering therapy in type 2 diabetes mellitus: background paper for the American College of Physicians. Ann Intern Med. 2004;40:650-8

151. Sacks FM, Pfeffer MA, Moye LA, Rouleau JL, Rutherford JD, Cole TG, Brown L, Warnica JW, Arnold JM, Wun CC, Davis BR, Braunwald E. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events (Care) Trial investigators. N Engl J Med. 1996;335:1001-9
152. Pyorala K, Pedersen TR, Kjekshus J, Faergeman O, Olsson AG, Thorgeirsson G. Cholesterol lowering with simvastatin improves prognosis of diabetic patients with coronary heart disease. A subgroup analysis of the Scandinavian Simvastatin Survival Study (4S). Diabetes Care 1997;20:614-20
153. Pyorala K, Ballantyne CM, Gumbiner B, Lee MW, Shah A, Davies MJ, Mitchel YB, Pedersen TR, Kjekshus J; Scandinavian Simvastatin Survival Study (4S). Reduction of cardiovascular events by simvastatin in nondiabetic coronary heart disease patients with and without the metabolic syndrome: subgroup analyses of the Scandinavian Simvastatin Survival Study (4S). Diabetes Care. 2004 Jul;27(7):1735-40.
154. Colhoun HM, Betteridge J, Durrington PN, Hitman GA, Neil, Livingstone SJ, Thomason MJ, Mackness MI, Fuller FJ. The collaborative atorvastatin diabetes study (CARDS). Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. Lancet 2004;364: 685-96
155. Scottish Intercollegiate Guidelines Network, SIGN 96: Management of stable angina, 2007.
156. Scottish Intercollegiate Guidelines Network, SIGN 55: Management of Diabetes, 2001.
157. Scottish Intercollegiate Guidelines Network, SIGN 55: Management of Diabetes, 2001
158. Fahey T, Schroeder K, Ebrahim S (Cochrane Hypertension Group). Interventions used to improve control of blood pressure in patients with hypertension. The Cochrane Database of Systematic Reviews. The Cochrane Collaboration Volume (2), 2006, Updated: 22-February-2006
159. McIntosh A, Hutchinson A, Home PD, Brown F, Bruce A, Damerell A, et al. Clinical guidelines and evidence review for Type 2 diabetes: management of blood glucose. Sheffi eld: ScHARR, University of Sheffi eld, 2001. http://www.nice.org.uk/pdf/NICE_full_blood_glucose.pdf
160. The National Collaborating Centre for Chronic Conditions. Type 1 Diabetes in Adults. National clinical guideline for diagnosis and management in primary and secondary care. http://www.rcplondon.ac.uk/pubs/books/DIA/index.asp
161. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Canadian Diabetes Association 2003 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Canadian Journal of Diabetes 2003; 27(Suppl 2): S18- S23. http://www.diabetes.ca
162. Coster S, Gulliford MC, Seed PT, Powrie JK, Swaminathan R. Self-monitoring in Type 2 diabetes: a meta-analysis. Diabet Med 2000; 17: 755-61.
163. Karter AJ, Ackerson LM, Darbinian JA, D Agostino Jr RB, Ferrara A, Liu J, et al. Self-monitoring of blood glucose levels and glycemic control: the Northern California Kaiser Permanente Diabetes Registry. Am J Med 2001; 111: 1-9.
164. Welschen LMC, Bloemendal E, Nijpels G, Dekker JM, Heine RJ, Stalman WAB, et al. Self-monitoring of blood glucose in patients with type 2 diabetes who are not using insulin: a systematic review. Diabetes Care 2005;28: 1510-17.
165. Pastors JG, Warshaw H, Daly A, Franz M, Kulkarni K. The evidence for the effectiveness of medical nutrition therapy in diabetes management. Diabetes Care 2002; 25: 608-13.
166. Sigal RJ, Kenny GP, Wasserman DH, Castaneda-Sceppa C. Physical activity/exercise and type 2 diabetes. Diabetes Care 2004; 27: 2518-39.
167. Herman WH, Hoerger TJ, Brandles M, Hicks K, Sorensen S, Zhang P, et al. Diabetes Prevention Program Research Group. The cost-effectiveness of lifestyle modifi cation or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. Ann Intern Med 2005; 142: 323-32.
168. The Diabetes Prevention Program Research Group. Impact of intensive lifestyle and metformin therapy on cardiovascular disease risk factors in the Diabetes Prevention Program. Diabetes Care 2005; 28: 888-94.
169. American Diabetes Association. Standards of Medical Care in Diabetes. Diabetes Care 2005; 28 (Suppl 1): S4-S36.
170. Franz MJ, Bantle JP, Beebe CA, Brunzell JD, Chiasson JL, Garg A, et al. Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. Diabetes Care 2002; 25: 148-98.
171. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Canadian Diabetes Association 2003 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Canadian Journal of Diabetes 2003; 27(Suppl 2). http://www.diabetes.ca
172. Wei M, Gibbons LW, Kampert JB, et al. Low cardiorespiratory fitness and physical inactivity as predictors of mortality in men with type 2 diabetes. Ann Intern Med. 2000;132:605-611.
173. Hu FB, Stampfer MJ, Solomon C, et al. Physical activity and risk for cardiovascular events in diabetic women. Ann Intern Med. 2001;134:96-105.
174. Moy CS, Songer TJ, LaPorte RE, et al. Insulin-dependent diabetes mellitus, physical activity, and

death. Am J Epidemiol. 1993;137:74-81.
175. European Physical Activity Surveillance System (EUPASS) Final report to the European Commission (DG SANCO F/3, HEALTH MONITORING PROGRAMME), March 2001
176. Diabetes Prevention Program (DPP) Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med, Vol. 346, No. 6 February 7, 2002
177. Kronsbein P, Jˆrgens V, M¸hlhauser I, et al. Evaluation of a structured treatment and teaching programme on non-insulin-dependent diabetes. Lancet 1988;ii:1407-10
178. Trocha AK, Schmidtke C, Didjurgeit U, Muhlhauser I, Bender R, Berger M, Sawicki PT. Effects of intensified antihypertensive treatment in diabetic nephropathy: mortality and morbidity results of a prospective controlled 10-year study.
179. Deakin T, McShane CE, Cade JE, Williams RD. Group based training for self-management strategies in people with type 2 diabetes mellitus. Cochrane Database Syst Rev, 2005, 2, CD003417.
180. International Consensus on the Diabetic Foot. International Working Group on the Diabetic Foot 1999
181. Jaakko Tuomilehto et al: Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 2001; 344: 1343-1350.
182. NICE. Technology Appraisal 60. Guidance on the use of patient-education models for diabetes. London, National Institute for Clinical Excellence, 2003. http://www.nice.org.uk
183. Piette JD, Glasgow RE. Education and home glucose monitoring. In: Gerstein HC, Haynes RB (eds) Evidencebased Diabetes Care. Hamilton, Ontario: BC Decker, 2001: pp 207-51.
184. Gary TL, Genkinger JM, Gualler E, Peyrot M, Brancati FL. Meta-analysis of randomized educational and behavioral interventions in type 2 diabetes. The Diabetes Educator 2003; 29: 488-501.
185. Warsi A, Wang PS, LaValley MP, Avorn J, Solomon DH. Self-management education programs in chronic disease. A systematic review and methodological critique of the literature. Arch Intern Med 2004; 164: 1641-49.
186. The association of comorbid depression with mortality in patients with type 2 diabetes, Diabetes Care. 2005 Nov;28(11):2668-72
187. Henkel V, Mergl R, Kohnen R, et al.Identifying depression in primary care: a comparison of different methods in a prospective cohort study. BMJ 2003;326:200
188. Sharp L, Lipsky M. Screening for Depression Across the Lifespan. A Review of Measures for Use in Primary Care Settings. American Family Physician, September 15,2002/Volume 66, Number 6
189. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Canadian Diabetes Association 2003 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Canadian Journal of Diabetes 2003; 27(Suppl 2): S50- S52. http://www.diabetes.ca
190. Scottish Intercollegiate Guidelines Network. SIGN 55. Management of Diabetes, 2001. http://www.sign.ac.uk
191. The National Collaborating Centre for Chronic Conditions. Type 1 Diabetes in Adults. National clinical guideline for diagnosis and management in primary and secondary care. http://www.rcplondon.ac.uk/pubs/books/DIA/index.asp
192. Institute for Clinical Systems Improvement (Bloomington, MN, USA). Management of Type 2 Diabetes Mellitus, 2004. http://www.icsi.org/knowledge
193. American Diabetes Association. Standards of Medical Care in Diabetes. Diabetes Care 2005; 28 (Suppl 1): S4-S36.
194. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes. A meta-analysis. Diabetes Care 2001; 24:1069-78.
195. Rubin RR, Ciechanowski P, Egede LE, Lin EHB, Lustman PJ. Recognizing and treating depression in patients with diabetes. Current Diabetes Reports 2004; 4: 119-25.
196. Herpertz S, Petrak F, Albus C, Hirsch A, Kruse J, Kulzer B. Psychosoziales und Diabetes mellitus. In: Deutsche Diabetes Gesellschaft (DDG) und Deutsches Kollegium Psychosomatische Medizin (DKPM) (eds) Evidenzbasierte Diabetes-Leitlinie DDG. Diabetes und Stoffwechsel 2003; 12 (Suppl 2). http://www.diabetes-psychologie.de/en/guidelines.htm
197. Snoek FJ, Skinner TC. Psychological counselling in problematic diabetes. Does it help? Diabet Med 2004; 19: 265-73.
198. Ismail K, Winkley K, Rabe-Hesketh S. Systematic review and meta-analysis of randomised controlled trials of psychological interventions to improve glycaemic control in patients with type 2 diabetes. Lancet 2004; 363: 1589-97.
199. Psychiatric Research Unit, WHO Collaborating Center for Mental Health. WHO-5 questionnaires, http://www.cure4you.dk/354/index.php?&a=e2&ei=4033&m=2
200. Redekop W, Koopmanschap M , Stolk R, Rutten G, Wolffenbuttel B, Niessen L. Health-Related Quality of Life and Treatment Satisfaction in Dutch Patients With Type 2 Diabetes. Diabetes Care 25:458-463, 2002
201. Rubin R. Diabetes and Quality of Life. Diabetes Spectrum Volume 13 Number , 2000, Page 21
202. Snoek F. Quality of Life: A Closer Look at Measuring Patients' Well-Being. Diabetes Spectrum Volume

13 Number, 2000, Page 24
203. Delamater A. Quality of Life in Youths With Diabetes. Diabetes Spectrum, Volume 13 Number, 2000, Page 42
204. Testa M. Quality-of-Life Assessment in Diabetes Research: Interpreting the Magnitude and Meaning of Treatment Effects. Diabetes Spectrum Volume 13 Number , 2000, Page 29
205. Polonsky W. Understanding and Assessing Diabetes-Specific Quality of Life. Diabetes Spectrum Volume 13 Number , 2000, Page 36
206. Glasgow RE, Ruggiero L, Eakin EG, Dryfoos J, Chobanian L. Quality of life and associated characteristics in a large national sample of adults with diabetes. Diabetes Care, Vol 20, Issue 4 562-567
207. AM Jacobson, M de Groot and JA Samson. The evaluation of two measures of quality of life in patients with type I and type II diabetes. Diabetes Care, Vol 17, Issue 4 267-274
208. Ragnarson Tennvall G, Apelqvist J. Health-related quality of life in patients with diabetes mellitus and foot ulcers. J Diabetes Complications. 2000 Sep-Oct;14(5):235-243
209. Hart HE, Redekop WK, Berg M, Bilo HJ, Meyboom-de Jong B Factors that predicted change in health-related quality of life were identified in a cohort of diabetes mellitus type 1 patients. J Clin Epidemiol. 2005 Nov; 58(11):1158-64. Epub 2005 Sep 12.
210. Gore M, Brandenburg NA, Dukes E, Hoffman DL, Tai KS, Stacey B. Pain severity in diabetic peripheral neuropathy is associated with patient functioning, symptom levels of anxiety and depression, and sleep. J Pain Symptom Manage. 2005 Oct;30(4):374-85.
211. Hart HE, Bilo HJ, Redekop WK, Stolk RP, Assink JH, Meyboom-de Jong B. Quality of life of patients with type I diabetes mellitus. Qual Life Res. 2003 Dec;12(8):1089-97.
212. Bagust A, Beale S. Modelling EuroQol health-related utility values for diabetic complications from CODE-2 data. Health Economics Volume 14, Issue 3 , Pages 217 230. Published online 7 Sep 2004
213. Weber B, Burger W, Hartmann R, Hovener G, Malchus R, Oberdisse U. Riskfactors for the development of retinopathy in children and adolescents with type 1 (insulin-dependent) diabetes mellitus. Diabetologia: 29: 23-29.
214. American Diabetes Association (2002) Diabetic retinopathy. Diabetes Care S1.: S90-93.
215. Mogensen CE, Keane WF, Bennett PH, Jerums G, Parving HH, Passa P, Steffes MW, Striker GE, Viberti GC (1995) Prevention of diabetic renal disease with special reference to microalbuminuria. Lancet 346: 1080-1084
216. Clinical Guideline F- Management of type 2 diabetes. Renal disease prevention and early management. Issue date: February 2002. Review date: March 2005 download from www.nice.org 05/07/06
217. Bruno G et al. Progression to overt nephropathy in type 2 diabetes: the Casale Monferrato Study. Diabetes Care. 2003 Jul;26(7):2150-5.
218. Levey A, Coresh J et al. National Kidney Foundation Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification. Annals of Internal Medicine, Volume 139, Number 2, 15 July 2003.
219. Taal M.W & Tomson C: Clinical Practice Guidelines for the Care of Patients with Chronic Kidney Disease. UK Renal Association Clinical Practice Guidelines 4th Edition 2007
220. Stages of chronic kidney disease according to the US K/DOQI-Classification
221. Kidney Disease Quality Outcomes Initiative. Clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis 2002;39(2 Suppl 1) S1-S266.
222. Spraul M, Raunest J, Reike H. Der diabetische Fufl. In Berger M: Diabetes mellitus 2000
223. SIGN (Scottish Intercollegiate Guidelines Network). Management of diabetic foot disease 2001
224. Morbach S, M¸ller E, Reike H, Risse A, Spraul M. Diabetisches Fuflsyndrom Praxis-Leitlinie, DDG, Diabetes und Stoffwechsel 13/2004, 73-76 (b)
225. International Consensus on the Diabetic Foot. International Working Group on the Diabetic Foot 1999
226. Steed DL, Ricotta JJ, Prendergast JJ, Kaplan RJ, Webster MW, McGill JB, et al. Promotion and acceleration of diabetic ulcer healing by arginine-glycine-aspartic acid (RGD) peptide matrix. RGD Study Group. Diabetes Care 1995; 18: 39-46.
227. Holloway G,Steed D,DeMarco M,Masmuto T,Moosa H,Webster M,Bunt T,Polansky M. A randomised controlled dose response trial of activated platelet supernatant,topical CT-102 in chronic,non-healing diabetic wounds. Wounds 1993;5:198-206.
228. Wieman TJ, Smiell JM, Su Y. Efficacy and safety of a topical gel formulation of recombinant human platelet-derived growth factor-BB (becaplermin) in patients with chronic neuropathic diabetic ulcers. A phase III randomized placebocontrolled double-blind study. Diabetes Care 1998; 21: 822-7.
229. Grayson ML, Gibbons GW, Habershaw GM, Freeman DV, Pomposelli FB, Rosenblum BI, et al. Use of ampicillin/sulbactam versus imipenem/cilastatin in the treatment of limb-threatening foot infections in diabetic patients. Clin Infect Dis 1994; 18: 683-93.
230. Lipsky BA, Baker PD, Landon GC, Fernau R. Antibiotic therapy for diabetic foot infections: comparison of two parenteral-to-oral regimens. Clin Infect Dis 1997; 24: 643-8.
231. Erstad BL Jr, McIntyre KE Jr, Mills JL. Prospective, randomized comparison of ampicillin/sulbactam

and cefoxitin for diabetic foot infections. Vascular Surgery 1997; 31: 419-26.
232. Chantelau E, Tanudjaja T, Altenhofer F, Ersanli Z, Lacigova S, Metzger C. Antibiotic treatment for uncomplicated neuropathic forefoot ulcers in diabetes: a controlled trial. Diabet Med 1996; 13: 156-9.
233. Scottish Intercollegiate Guidelines Network. SIGN 55. Management of Diabetes, 2001. http://www.sign.ac.uk
234. Morbach S, M¸ller E, Reike H, Risse A, Spraul M. Diagnostik, Therapie, Verlaufskontrolle und Pr‰vention des diabetischen Fuflsyndroms. In: Scherbaum WA, Kiess W, Landgraf R (eds) Evidenzbasierte Diabetes Leitlinien DDG. Diabetes und Stoffwechsel 2004; 13 (Suppl 2). http://www.deutsche-diabetes gesellschaft.de
235. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Canadian Diabetes Association 2003 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Canadian Journal of Diabetes 2003; 27(Suppl 2): S74- S75. http://www.diabetes.ca
236. National Institute for Clinical Excellence. Type 2 diabetes footcare. London: National Institute for Clinical Excellence, 2004. http://www.nice.org.uk/page.aspx?o=101518
237. Institute for Clinical Systems Improvement (Bloomington, MN, USA). Management of Type 2 Diabetes Mellitus, 2004. http://www.icsi.org/knowledge
238. Campbell L, Colagiuri S, O Rourke S, Chen M, Colagiuri R. Evidence Based Guidelines for Type 2 Diabetes. Detection and Prevention of Foot Problems. Canberra: Diabetes Australia & NHMRC, 2005. http://www.diabetesaustralia.com.au
239. International Working Group on the Diabetic Foot. Apelqvist J, Bakker K, Van Houtum WH, Nabuurs-Franssen MH, Shaper NC (eds) International Consensus on the Diabetic Foot. Maastricht, The Netherlands, 1999.
240. Lipsky BA. A report from the international consensus on diagnosing and treating the infected diabetic foot. Diabetes Metab Res Rev 2004; 20 (Suppl 1): S68-S77.
241. Eldor R, Raz I, Ben Yehuda A, Boulton AJM. New and experimental approaches to treatment of diabetic foot ulcers: a comprehensive review of emerging treatment strategies. Diabet Med 2004; 21: 1161-73.
242. Singh N, Armstrong DG, Lipsky BA. Preventing foot ulcers in patients with diabetes. JAMA 2005; 293: 217-28.
243. Valk GD, Kriegsman DMW, Assendelft WJJ. Patient education for preventing diabetic foot ulceration. A systematic review. Endocrinol Metab Clin North Am 2002; 31: 633-58.
244. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Canadian Diabetes Association 2003 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. Canadian Journal of Diabetes 2003; 27(Suppl 2): S72- S73, S81-S82. http://www.diabetes.ca
245. The National Collaborating Centre for Chronic Conditions. Type 1 Diabetes in Adults. National clinical guideline for diagnosis and management in primary and secondary care. http://www.rcplondon.ac.uk/pubs/books/DIA/index.asp
246. Boyko E, Ahroni J, Stensel V, Forsberg R, Davignon D & Smith D: A Prospective Study of Risk Factors for Diabetic Foot Ulcer The Seattle Diabetic Foot Study Diabetes Care 22:1036 1042, 1999
247. Vascular risk factors and diabetic neuropathy; N Engl J Med. 2005 Jan 27;352(4):341-50.
248. Haslbeck M, Luft D, Neundˆrfer B, Redaelli M, Stracke H, Ziegler D, et al. Diagnose, Therapie und Verlaufskontrolle der diabetischen Neuropathie. In: Scherbaum WA, Landgraf R (eds) Evidenzbasierte Diabetes-Leitlinien DDG, 2nd edn. Deutsche Diabetes-Gesellschaft 2004. http://www.deutsche-diabetes-gesellschaft.de
249. Dyck PJ, Kratz KM, Karnes JL, Litchy WJ, Klein R, Pach JM, et al. The prevalence by staged severity of various types of diabetic neuropathy, retinopathy, and nephropathy in a population-based cohort: The Rochester Diabetic Neuropathy Study. Neurology 1993: 43: 817-24.
250. Boulton AJM, Malik RA, Arezzo JC, Sosenko JM. Diabetic somatic neuropathies (Technical Review). Diabetes Care 2004; 27: 1458-86.
251. Vinik AI, Maser RE, Mitchell B, Freeman R. Diabetic autonomic neuropathy: a technical review. Diabetes Care 2003; 26: 1553-79.
252. The ACCORD Study Group. Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial: design and methods. Am J Cardiol 2007;99(Suppl):21i 33.
253. Gerstein HC, Riddle MC, Kendall DM, et al. Glycemia treatment strategies in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. Am J Cardiol 2007;99(Suppl):34i 43.
254. National Heart, Lung and Blood Institute, Action to Control Cardiovascular Risk in Diabetes (ACCORD) Trial, February 6, 2008, available at: www.nhlbi.nih.gov/ health/prof/heart/other/accord/index.htm.
255. The Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, et al. Effects of intensive glucose lowering in type 2 diabetes. New Engl J Med 2008;358:2545 59.
256. ADVANCE Collaborative Group, Patel A, MacMahon S, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. New Engl J Med 2008;358:2560 72.
257. Duckworth W, Abraira C, Moritz T, et al., on behalf of the VADT Investigators. Glucose control and

vascular complications in veterans with type 2 diabetes. N Engl J Med 2009;360:129-39.
258. Tuomilehto, J., Borch-Johnsen, K., Molarius, A., et al. Incidence of cardiovadcular disease in type 1 (insulin dependent) diabetic subjects with and without diabetic nephropathy in Finland, Diabetologie 1998, 41: 784 90.
259. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group. Intensive Diabetes Treatment and Cardiovascular Disease in Patients with Type 1 Diabetes. N Engl J Med 2005;353:2643-53.
260. EUROCISS. DIAGNOSTIC CRITERIA OF AMI/ACS, available at: http://www.cuore.iss.it/eurociss/reg_Ima/pdf/Ima_criteri-diagnostici.pdf
261. EUROCISS. DIAGNOSTIC CRITERIA OF STROKE, available at: http://www.cuore.iss.it/eurociss/reg_ictus/pdf/ictus_criteri-diagnostici.pdf
262. ACC/AHA 2005 practice Guidelines for the Management of Patients With Peripheral Arterial Disease (Lower Extremity, Renal, Mesenteric, and Abdominal Aortic) Circulation 2006;113;1474-1547
263. The Framingham Heart Study, Diabetes Care 27(3):704-708, 2004
264. De Lissovoy G, Ganoczy DA, Ray NF (2000). Relationship of hemoglobin A1c, age of diabetes diagnosis, and ethnicity to clinical outcomes and medical costs in a computer simulated cohort of persons with type 2 diabetes. Am J Manag Care 6: 573-584
265. ECHI working group. Design for a Set of European Community Health Indicators Final report of the ECHI Project. 1-93. 2001
266. United Nations, Department of Economic and Social Affairs Population Division (eds.). World Urbanization Prospects The 2005 Revision. Executive Summary. Fact Sheets. Data Tables. New York, 2006
267. Disease Management Association of America, Definition of Disease Management, http://www.dmaa.org/dm_definition.asp, 2002.
268. EUDIP Group. Establishing indicators monitoring diabetes mellitus and its morbidity, 2002.

Standardized European Diabetes Reports

Svein Skeie, Karianne Fjeld Loevaas

ABSTRACT

Introduction
A central element of the BIRO project is the ability to report diabetes outcome indicators for different regions in Europe. The creation of standardized reports requires the identification of a precise design for the output that can be usefully replicated. A template constitutes the basic reference which, through an agreed preset format, can be used repeatedly each time the report is applied or subsequently adapted.

Objectives
To define a reports template from the set of indicators defined by the clinical review. The template defines specifically how BIRO results are to be displayed and explained on the web portal, taking into account the needs of different types of audiences, which reflect the different areas covered by the project: governance, health care, research and people with/without diabetes.

Materials and methods
All data items and indicators from the clinical review have been listed in a table and assigned a specific level of priority and feasibility, where priority relates to the level of importance of its inclusion in the report, while feasibility indicates the possibility that the indicator could be practically estimated. Only data items and indicators with high priority and high feasibility were included in the final reports template. For each data item and indicator, any strata for output have been also indicated. A revision of graphical displays for diabetes reporting released by major agencies worldwide has been also undertaken. Best options have been selected as target statistical outputs for each category of audience, by data item and indicator, as well as the HTML code that would practically deliver results obtained by the application of the statistical engine.

Results
A total of data items/indicators were finally chosen, with an indication of the target audience and description of target statistical outputs. Results obtained have been directly submitted to the designers of the statistical engine and developers of the web portal to produce and organize all BIRO outputs accordingly.

Conclusions
For statistical reports to be best informative, they must strictly relate to the evidence base. The range of indicators must be restricted to those most reliable, computed properly using standardized methods. Through an objective procedure, the BIRO system has identified a minimal template that corresponds to the agreed set of indicators, referenced by a basic set of graphical outputs chosen among the plethora of those currently available in diabetes reports worldwide.

2.2.1 Introduction

The BIRO project aims to report diabetes outcome indicators for regions in Europe. To create standardized reports, a possible model in the form of a structured document must be developed.
A template is a document or file having a preset format, used as a starting point for a particular application, so that the format does not have to be recreated each time it is used.
In this context, a template can be viewed as the model for diabetes indicator reporting in the BIRO project, that is made available also to others who might choose to use the same format.

The aim of the reports template is to select an initial set of indicators from those defined in the clinical review, specifying how results will be reported, displayed and explained on the web portal.
The clinical review serves as the full backbone of diabetes indicators identified in the BIRO project, including fundamental details leading to the final formulation of all statistical indicators to be reported on.

The first goal of BIRO was to provide an initial, albeit limited, very useful report output system. The general value of the system lies in the fact that once a shared information system has been established, this can be extended to include other indicators and more sophisticated analyses, e.g.:

- analysis of longitudinal trends and average outcomes in a diabetes population
- identification of patterns of care and prevention consistently showing positive results
- identification of population strata and/or practices that do not show effective results
- verification of the application/applicability of best practice guidelines
- on-field testing of collaborative information systems in chronic diseases

The reports template should satisfy the need for different types of audience, reflecting the full scope of the project.

It has been agreed by the Consortium that BIRO will take account of three main categories of users: 1. Policy makers in charge of Governance, 2. Experts in Health care and research and 3. People with or without diabetes.

The three audiences have different reasons for being interested in diabetes and quality of diabetes care in Europe. The way they interpret diabetes data can be substantially different, based upon their level of interest in the field. Therefore any reporting system must fit the different needs and be aware of the different backgrounds involved.

2.2.2 Materials and Methods

All data items and indicators appearing in the clinical review have been listed in a basic table, with a scale of priority, and a score of feasibility assigned for each specific indicator. Priority indicates how important is to include the selected data item, while feasibility indicates how realistic is to collect the selected data item. Only data items and indicators with high priority and high feasibility were included in the further reports template. For each data item and indicator it is indicated if and how it should be stratified for output.
Priority scale is based on the question: “How important is it to include the selected data item?”. Therefore, the scale is subjective and has been consensus based in BIRO, with the following scores:
1=High, 2=Intermediate, 3=Low.

The feasibility scale is based on the question: “How feasible is to collect the selected data item?”. Again, the scale is subjective and consensus based, with scores A=High, B=Intermediate, C=Low. Finally, the column “Strata” indicates whether or not the data item should be stratified for output.

As a last step, a selection of statistical outputs is chosen for each category of audience and for each data item and indicator. An HTML structure for the statistical engine has been identified, based on this information.

As an example, Table 2.2.1 shows how data items and indicators from the clinical review were first listed in the table, with priority assigned for each data item and indicator.

The 2nd Annual BIRO Meeting has been used to finalize a first draft of the template agreed among members of the Consortium.

After completing the full table, the BIRO Consortium decided that only data items and indicators with high priority (1) and high feasibility (A) were to be included in the further reports template. Such fundamental choice addressed the relevance of ensuring good data quality as a fundamental platform for the implementation of a reporting system for the European Union.

Indicators chosen to be reported on and to be stratified have been listed with chosen stratification factors - decided by consensus in the BIRO-project consortium. Table 2.2.2 shows an extract of precise stratification factors and units for indicators with high priority and high feasibility.

Health reports from different countries in Europe have been used to find examples of ways to present the data that have been finally selected by the Consortium as candidate graphical outputs.

2.2.3 Results

A total of N=72 diabetes indicators have been finally selected for the standardized report of the BIRO Consortium, of which N=2 pertain to the domain of demographic characteristics, N=18 to clinical characteristics, N=21 to the health system, N=3 to the population level, N=28 to risk adjusted indicators. This last category is split into N=2 indicators for epidemiology, N=16 processes, and the residual N=10 for outcomes.

Table 2.2.1: Descriptive data items for diabetes population given priority and feasibility level (part of full table given as example).

	Priority Level	Feasibility	Strata
1) Demographic characteristics			
a) Age (Classes)	1	A	X
b) Gender	1	A	X
c) Ethnicity	2	C	
d) Education	2	C	
e) Employment status	2	C	
2) Clinical characteristics			
a) Diabetes status			
1. Type of diabetes	1	A	X
2. Duration of diabetes	1	A	X
b) Risk factors for diabetes complications			
1. Obesity			
a. Weight	1	A	X
b. Height	1	A	X
c. BMI	1	A	X
d. Waist circumference	2	B	
2. Lifestyle			
a. Smoking status	1	A	X
b. Cigarettes per day	2	B	
c. Alcohol intake	2	B	

Table 2.2.2: Stratification factors and units for selected data items and indicators

	Units	Strata
1. Demographic characteristics		
1.1. Age (Classes)		Gender
1.2. Gender		Age
2. Clinical characteristics		
2.1. Diabetes status		
2.1.1. Type of diabetes	1, 2, GDM, Other	Age
2.1.2. Duration of diabetes	years	Type of diabetes, Gender, Hba1c
2.2. Risk factors for diabetes complications		
2.2.1. Obesity		
2.2.1.1. Weight	kg	Gender x Age x Type of diabetes
2.2.1.2. BMI	Kg/m^2	Gender x Age x Type of diabetes
2.2.2. Lifestyle		
2.2.2.1. Smoking status		Gender x Age x Type of diabetes
2.2.3. Clinical measurements		
2.2.3.1. Systolic BP	mmHg	Gender x Age x Type of diabetes
2.2.3.2. Diastolic BP	mmHg	Gender x Age x Type of diabetes
2.2.3.3. Total cholesterol	mmol/L	Gender x Age x Type of diabetes
2.2.3.4. HDL-cholesterol	mmol/L	Gender x Age x Type of diabetes
2.2.3.5. Creatinine	µmol/L	Gender x Age x Type of diabetes
2.2.3.6. HbA1c	%	Gender x Age x Type of diabetes

The complete list of indicators with associated strata for the class of users: "Governance" are given in Tables 2.2.4 (1-4).

A selection of statistical outputs has been chosen for each category of audience and for each data item and indicator. The type of statistical output underlined in tables 2.2.4 is the one suggested for the report. Different examples of implementation of the statistical outputs are shown in Table 2.2.3 and Figures 2.2.1-2.2.5, based upon common approaches used in main diabetes health reports worldwide.

The situation shown in table 2.2.3 is typical of an output where the distribution of some relevant characteristic (age) may vary across strata of important variables e.g. type of diabetes and gender. In similar cases it would be highly beneficial to display a multidimensional

Table 2.2.3: Type of diabetes by age and gender (Source: The Norwegian diabetes register for adults)

Age (year)	Male		Female	
	Type 1 (%)	Type 2 (%)	Type 1 (%)	Type 2 (%)
0-9	0.1	0.1	0.1	0.1
10-19	0.2	0.2	0.2	0.2
20-29	1.3	0.3	1.3	0.3
30-39	1.6	1.0	1.6	1.0
40-49	2.0	3,1	2.0	3.1
50-59	2.3	4.0	2.3	4.0
60-69	3.4	5.0	3.4	5.0
70-89	3.6	7.0	3.6	7.0
>79	3.8	10.0	3.8	10.0

table where each cell is a particular combination of the target variable (age) with different levels of all the strata identified.

These kinds of complex partitions may be visualized using simple graphical outputs e.g. histograms (Figure 2.2.1[1]) and boxplots (Fig. 2.2.2[2]).

Lines connecting dots relative to single measurements across time are particularly effective to display longitudinal trends, as in Figure 2.2.3[3].

Multidimensional graphs may be highly informative when profiles of single units (e.g. hospitals or clinical centres) may be obtained from multiple variables, based upon positioning of each object component against the overall min/max values.
Figure 2.2.4[4] introduces the use of starplots, where the legend illustrates the content of each ray in terms of a key diabetes variable. Different profiles for different strata (in this example, patients with cardiac complications) indicate important deviations from an agree average.

Table 2.2.4 (1): The BIRO Reports Template (User Class: "Governance"), Demographic and Clinical Characteristic

Indicator	Strata	Statistical outputs
1. Demographic characteristics		
1.1 Age (Classes)	Gender	Table, histogram
1.2 Gender	Age	Table, histogram
2. Clinical characteristics		
2.1 Diabetes status		
2.1.1 Type of diabetes	Age	Table, histogram
2.1.2 Duration of diabetes	Type of diabetes, Gender, Hba1c	Table, histogram
2.2 Risk factors for diabetes complications		
2.2.1 Obesity		
2.2.1.1 Weight	Gender x Age x Type of diabetes	Table, lines
2.2.1.2 BMI	Gender x Age x Type of diabetes	Table, lines
2.2.2 Lifestyle		
2.2.2.1 Smoking status	Gender x Age x Type of diabetes	Histogram, table
2.2.3 Clinical measurements		
2.2.3.1 Systolic BP	Gender x Age x Type of diabetes	Table, lines
2.2.3.2 Diastolic BP	Gender x Age x Type of diabetes	Table, lines
2.2.3.3 Total cholesterol	Gender x Age x Type of diabetes	Table, lines
2.2.3.4 HDL-cholesterol	Gender x Age x Type of diabetes	Table, lines
2.2.3.5 Creatinine	Gender x Age x Type of diabetes	Table, lines
2.2.3.6 HbA1c	Gender x Age x Type of diabetes	Table, lines
2.3 Diabetes complications		
2.3.1 Retinopathy	Diabetes duration	Table, Histogram
2.3.2 End stage renal failure	Diabetes duration	Table, Histogram
2.3.3 Foot ulcer	Diabetes duration	Table, Histogram
2.3.4 Lower extremity amputation	Diabetes duration	Table, Histogram
2.3.5 Stroke	Diabetes duration	Table, Histogram
2.3.6 Myocardial infarction	Diabetes duration	Table, Histogram
2.3.7 Hypertension	Diabetes duration	Table, Histogram
3. Health system		
3.1 Structure (provider level)		
3.1.1 Type of provider		Table, Histogram
3.1.2 Average diabetes population per centre	Gender x Age x Type of diabetes	Table, Histogram
3.2 Structural quality		
3.2.1 Hospital beds per 100,000 population	Region	Maps, Table
3.2.2 Physicians employed per 100,000 population	Region	Maps, Table
3.3 Processes		
3.3.1 Foot examination		
3.3.1.1 Done	Age x Type of diabetes	Table, Histogram
3.3.2 Eye examination		
3.3.2.1 Done	Age x Type of diabetes	Table, Histogram
3.3.3 Measurement done		
3.3.3.1 BP done	Age x Type of diabetes	Table, Histogram
3.3.3.2 Lipids done	Age x Type of diabetes	Table, Histogram
3.3.3.3 Microalbumin done	Age x Type of diabetes	Table, Histogram
3.3.3.4 HbA1c done	Age x Type of diabetes	Table, Histogram
3.3.4 Treatment		
3.3.4.1 Antihypertensive medication	Age x Type of diabetes	Table, Histogram
3.3.4.2 Lipid lowering treatment	Age x Type of diabetes	Table, Histogram
3.3.4.3 ASA treatment (Aspirin)	Age x Type of diabetes	Table, Histogram
3.3.4.4 Glucose lowering treatment		
3.3.4.4.1 Diet only	Age x Type of diabetes	Table, Histogram
3.3.4.4.2 Tablets: 1. Sulphonyureas, 2. Biguanides, 3. Glucosidase Inhibitors, 4. Glitazones, 5. Glinides, 6. Other	Age x Type of diabetes	Table, Histogram
3.3.4.4.3 Insulin only	Age x Type of diabetes	Table, Histogram
3.3.4.4.4 Insulin and tablets	Age x Type of diabetes	Table, Histogram
3.3.4.4.5 Insulinpump	Age x Type of diabetes	Table, Histogram
3.3.5 Management		
3.3.5.1 Self-monitoring	Age x Type of diabetes	Table, histogram
3.3.5.2 Visit frequency	Age x Type of diabetes	Table, histogram
4. Population		
4.1 Area level		
4.1.1 Total population	Age x Gender	Table
4.1.2 Life expectancy	Age x Gender	Table
4.1.3 Mortality data	Age x Gender	Table

Table 2.2.4 (2): The BIRO Reports Template (User Class: "Governance"), Risk adjusted indicators

Indicator	Strata	Statistical outputs
5. Risk adjusted indicators		
5.1 Epidemiology		
5.1.1 Prevalence of diabetes mellitus per 1,000	Region, Type of diabetes	Maps, Table
5.1.2 Age at diagnosis by 10 year age bands (incidence)	Region, Type of diabetes	Maps, Histogram
5.2 Process quality		
5.2.1 % of persons with diabetes with one or more HbA1c tests during the last 12 months	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.2 % of persons with diabetes with at least one test for microalbuminuria during the last 12 months	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.3 % of persons with diabetes who received a dilated eye examination or evaluation of retinal photography by a trained caregiver within the last 12 months	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.4 % of persons with diabetes receiving at least one examination of the feet within the last 12 months	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.5 % of persons with diabetes whose smoking status was ascertained and documented within the last 12 months	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.6 % with serum creatinine tested in last 12 months	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.7 % of persons with diabetes and one or more blood pressure measurements within the last 12 months	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.8 % of persons with diabetes with hypertension who receive antihypertensive medication	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.9 Type of oral therapy (distribution of agents) in persons with diabetes type 2	Country, Gender x Age	Maps, tables, lines, histogram
5.2.10 % of persons treated with insulin among persons with diabetes	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.11 % of persons treated with insulin in combination with OADs among persons with diabetes	Country, Gender x Age	Maps, tables, lines, histogram
5.2.12 % of insulin treated persons with diabetes with insulinpump therapy	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.13 % of persons with diabetes with anti hypertensive treatment	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.14 % of persons with diabetes with lipid lowering treatment	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.15 % of persons with diabetes with ASA treatment	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.2.16 % of persons with diabetes performing self-monitoring of blood glucose	Region, Gender x Age x Type of diabetes	Maps, tables, lines, histogram
5.3 Outcome quality-intermediate outcomes		
5.3.1 % of persons with diabetes with most recent HbA1c level >9.0%	Region, Gender x Age x Type of diabetes	Tables, maps, histograms, lines
5.3.2 % of persons with diabetes with most recent HbA1c level >7,5%	Region, Gender x Age x Type of diabetes	Tables, maps, histograms, lines
5.3.3 % of persons with diabetes with most recent blood pressure <140/90 mmHg	Region, Gender x Age x Type of diabetes	Tables, maps, histograms, lines
5.3.4 % of persons with diabetes with BMI ≥ 30	Region, Gender x Age x Type of diabetes	Tables, maps, histograms, lines
5.3.5 % of persons with diabetes with microalbuminuria in last 12 months (among those who have been tested)	Region, Gender x Age x Type of diabetes	Tables, maps, histograms, lines
5.3.6 % of persons with diabetes who currently smoke among persons with diabetes	Region, Gender x Age x Type of diabetes	Tables, maps, histograms, lines
5.3.7 % of persons with diabetes with former or current foot ulceration	Region, Gender x Age x Type of diabetes	Tables, maps, histograms, lines
5.4 Outcome Quality – Terminal outcomes		
5.4.1 Annual incidence of dialysis and/or transplantation (renal replacement therapy) per 100,000 populations in persons with diabetes	Region, Gender x Age x Type of diabetes	Table, Map
5.4.2 % of persons with diabetes with ESRD	Region, Gender x Age x Type of diabetes	Table, Map
5.4.3 Annual death rate per 100,000 populations in the general population from all causes, adjusted for standard European population. Annual death rate per 100,000 populations in persons, who have as primary or secondary cause of death, diabetes mellitus, adjusted for standard European population	Region	Table, Map

Finally, figure 2.2.5[5] shows a figure extracted from the IDF Atlas exemplifying the use of epidemiological maps. In these cases, usually the frequency distribution of an individual data element or indicator is assigned a specific color and mapped against all the other values obtained from the whole area across a specific period of time.

Based upon the consideration that the template may present evidence-based indicators in different ways, a link has been maintained and properly referenced to connect the clinical review and the template. The web portal has been designed on the basis of a cross reference XML schema that links the clinical review to the reports template.

A base web report has been then designed using the template to precisely specify the target tables and all figures to be finally included in the report.

The base web report has been created using an HTML structure that is automatically populated with real numbers by the statistical engine, for each indicator (see Box 2.2.1).

The main tables in the HTML structure are visualized in a browser as follows (indicator 2.1.2):

In the example above Hba1c acts as an outcome, and the other strata as exposures. Partial column percentages are used to compare relative risks for different categories within the same sub-stratum. For instance, the relative risk (RR) of not doing an examination of glycated haemoglobin among females vs males, among persons with diabetes duration between 0-9 years is RR=(16/250)/(24/302)=.064/.079=.810.

BIRO REPORT - Site: centre1, 01 April 2009
Reference date: 31/12/02
Parameter: 2.1.2. Duration of diabetes (Classes).

HbA1c done	Type 1						
	0 - 9		10 - 19		20 +		
	Female	Male	Female	Male	Female	Male	
Yes	234 (93.6%)	278 (92.1%)	181 (94.3%)	229 (94.6%)	223 (93.7%)	170 (93.4%)	1315 (93.5%)
No	16 (6.4%)	24 (7.9%)	11 (5.7%)	13 (5.4%)	15 (6.3%)	12 (5.6%)	91 (6.5%)
	250 (45.3%)	302 (54.7%)	192 (44.2%)	242 (55.8%)	238 (56.7%)	182 (43.3%)	1406 (100.0%)

Chi-Square	p value
0.1117	0.7381

The result indicates that females in the lower duration level have close to 20% lower risk than males (250 (45.3%) vs. 302 (54.7%)) in not doing an Hba1c examination. However, the result in the upper classes are .057/.054=1.056 and .063/.066=.950.
A weighted average across strata would then lead to an estimate of .810 * (552/1406) + 1.056 * (434/1406) + .950 * (420/1406)= .318+.326+.284 =.928. In summary, the relative risk of female vs males is slightly lower than 7%, which is remarkably less than what was seen initially in the lower duration. Formulas for confidence intervals can simply lead to check for significance across all levels.

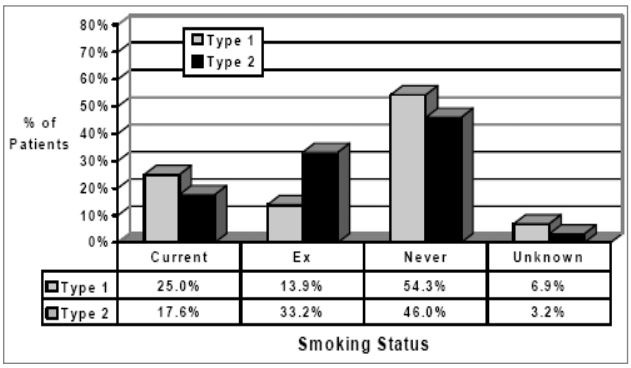


Figure 2.2.1: Histograms: Smoking Status

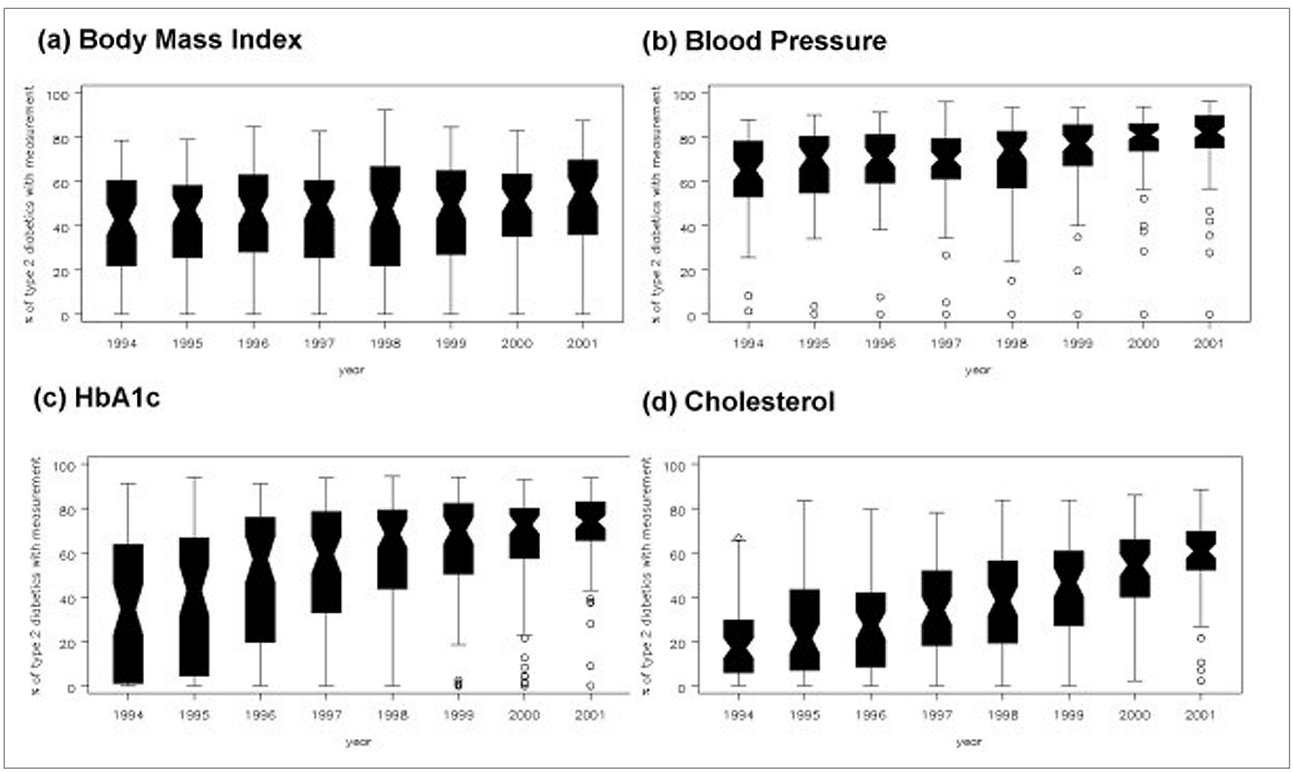


Figure 2.2.2: Boxplot: Distribution of diabetics with a risk factor measured by year

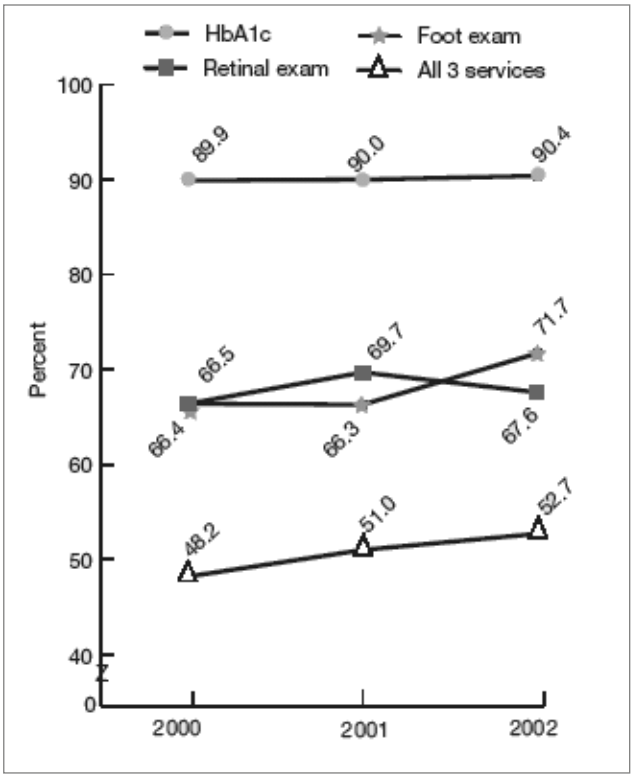


Figure 2.2.3: Lines: Adults age 18 and over with diagnosed diabetes who received Hba1c test, retinal exam, and all three tests, 2000-2002

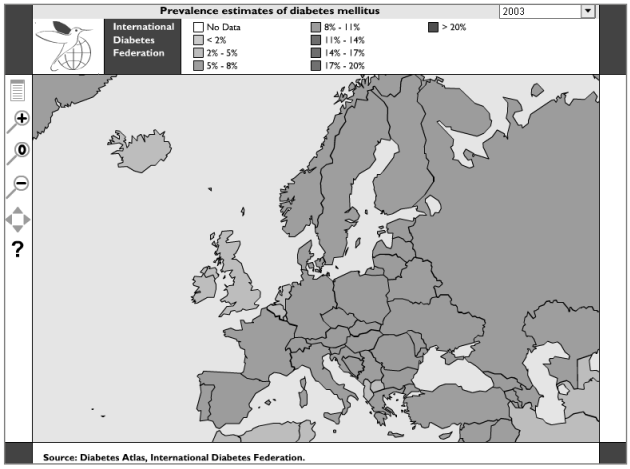


Figure 2.2.5: Maps: Prevalence Estimates

However, a Mantel Haensel[6] stratified chi square offers a single average measure of association between the exposure and the outcome, stratified by all exposure levels. For convenience, the chi-square statistic has been included in the BIRO report template as a possible candidate for the user class: governance.
R code performs a Cochran-Mantel-Haenszel chi-squared test of the null that two nominal variables are conditionally independent in each stratum, assuming that there is no three-way interaction. More sophisticated measures can be included in the future versions

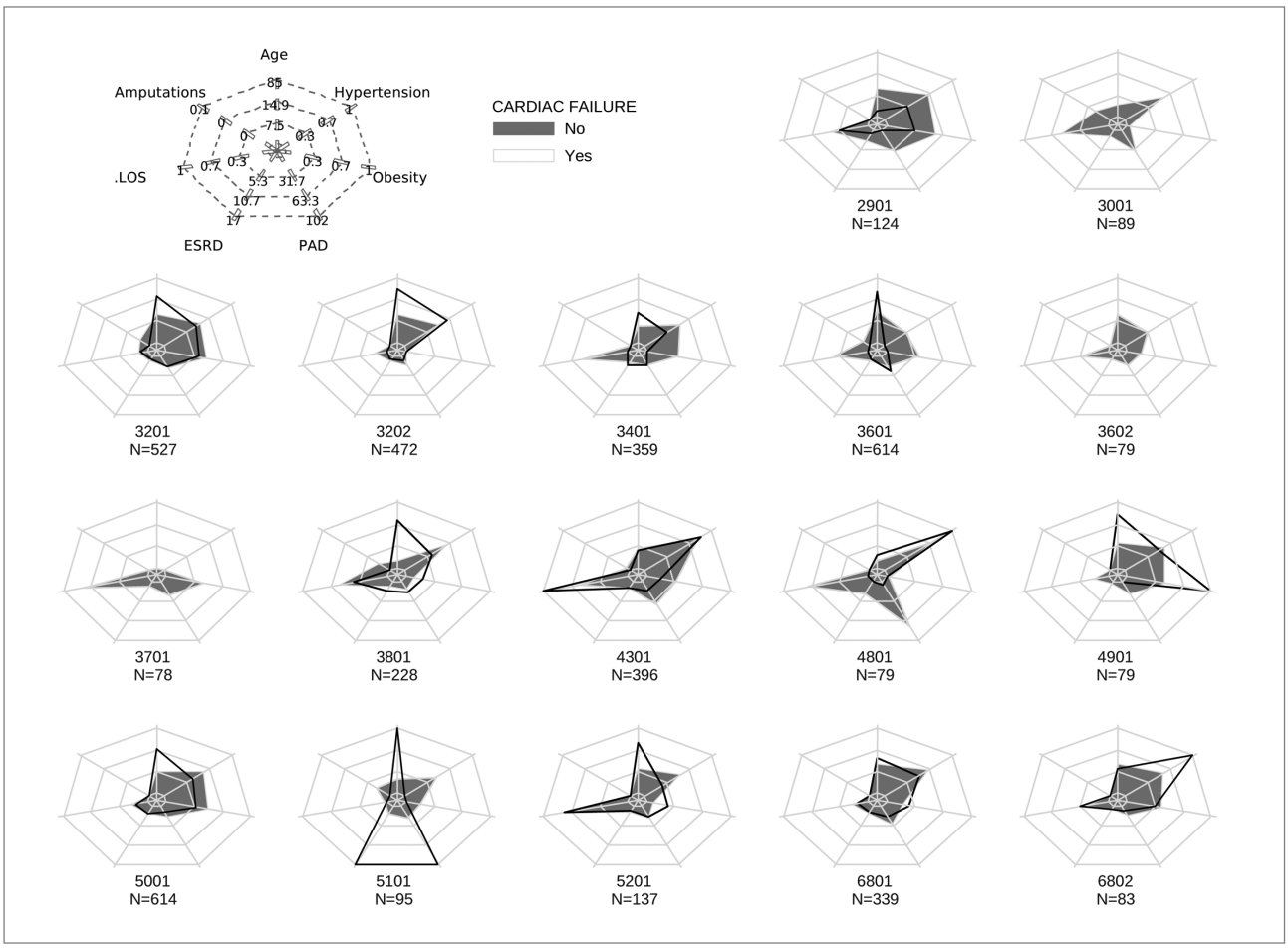


Figure 2.2.4: Starplots: Distribution by Discharge Ward stratified by Cardiac Complications

Box 2.2.1: HTML-codes of indicator "1.1 Age (Classes)"

```
<table border="1"><tbody>
  <tr><td>
     <table border="1"><tbody>
       <thead>
        <tr>
         <th rowspan=" 1 ">Age Classes</th><th>Female</th><th></th>
        </tr>
       </thead>
       <tr>
        <td class="catcol"> 0 - 34 </td>
        <td>n  ( p %)</td>
        <td>n  ( p %)<td class="rowsn">N</td>
       </tr>
       <tr>
        <td class="catcol"> 35 - 54 </td>
        <td>n  ( p %) </td>
        <td>n  ( p %) <td class="rowsn" >N</td>
       </tr>
       <tr>
        <td class="catcol"> 55 - 74 </td>
        <td>n  ( p %) </td>
        <td>n  ( p %) <td class="rowsn" >N</td>
       </tr>
       <tr>
        <td class="catcol"> 75 + </td>
        <td>n  ( p %) </td>
        <td>n  ( p %) <td class="rowsn">N</td>
       </tr>
       <tr>
        <td></td>
        <td class="colsn">N</td><td class="colsn">N</td><td class="grandn">N </td>
       </tr>
       </tbody></table>
      <br>
       <caption> Age Classes (by Gender) </caption>
       <br><br>
       <table border="1"><tbody>
         <thead>
          <tr>
           <th>statistic </th><th>p.value</th><th>df</th>
          </tr>
         </thead>
          <tr>
           <td>n</td><td>n</td><td>n</td>
          </tr>
        </tbody></table>
       <br><br>
   </td>
  </tr>
 </tbody></table>
<table border="1"><tbody>
  <tr><td>
  <img src="biro/software/_se_/output/reports/#<datetime>/graphs/bar1_1a.png
```

of the health care/research class of users.
The level of complexity in the interpretation of the results increases with the number of factors involved in the stratification. Graphical outputs may facilitate this step with a large (N>2) number of stratification factors.

The HTML structure also includes many figures selected among those existing in the diabetes reporting literature.

As mentioned above, histograms list among the most popular and easy to interpret in case of categorical variables (graphical representation of tables). In BIRO, these have been designed and developed as shown in Figure 2.2.6.

However, in case of continuous variables, boxplots constitute the most appropriate option (see Figure 2.2.7). In case of complex stratification, then it is clearly not easy to organize the presentation of a large number of graphical outputs.

In such situations, R offers a particular set of functions called "trellis", which are specifically built to allocate subgroups of graphs e.g. histograms and boxplots. Examples of this kind are shown in Figure 2.2.7, 2.2.8

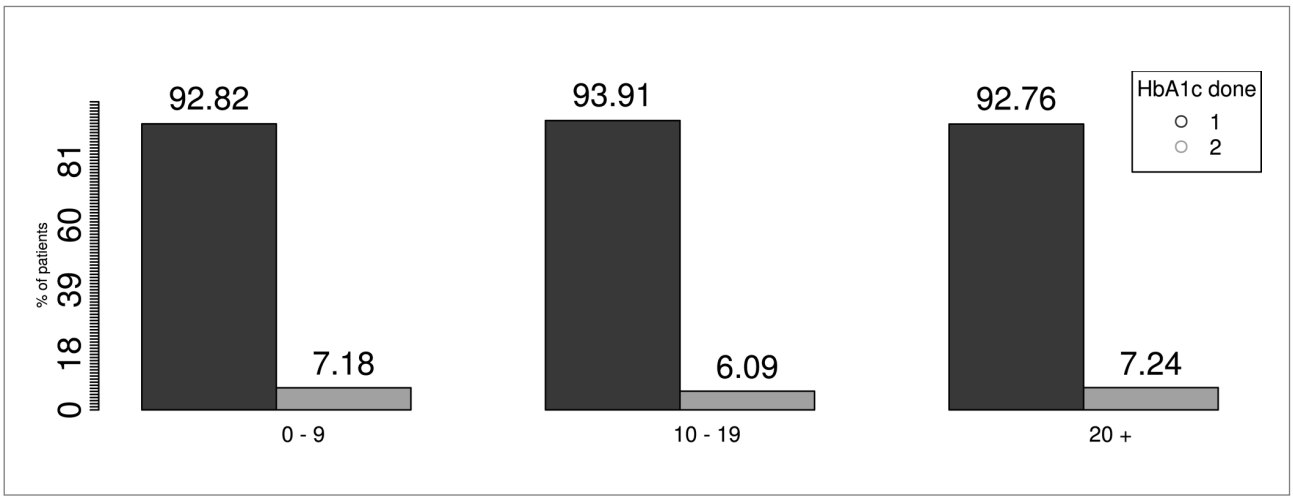


Figure 2.2.6: Indicator 2.1.2 Duration of Diabetes by Hba1c done

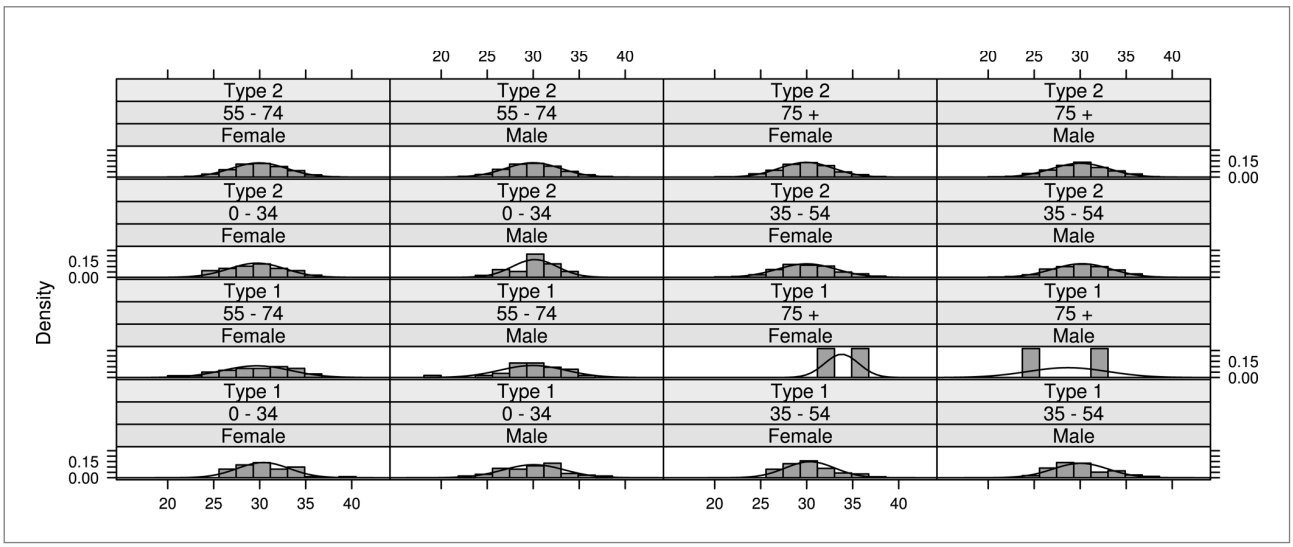


Figure 2.2.7: Indicator 2.2.1.2 Trellis density plot for BMI, by Gender, Age and Type of Diabetes

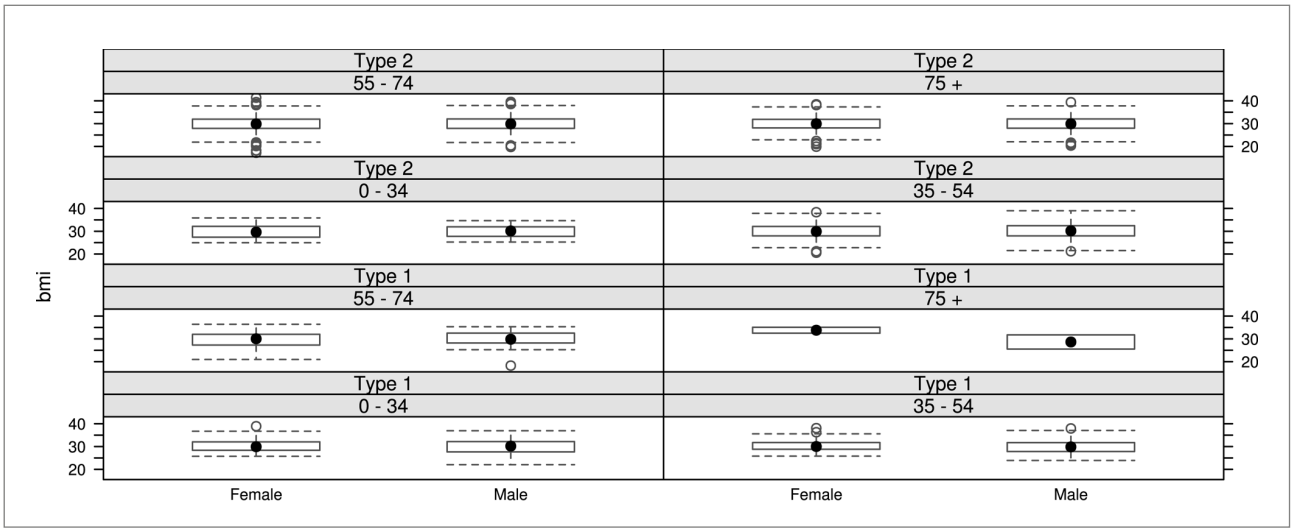


Figure 2.2.8: Indicator 2.2.1.2 Trellis Boxplot for BMI, by Gender, Age and Type of Diabetes

Conclusions

Through an objective process of review of the existing statistical reports for diabetes, the BIRO project has identified an ad hoc reports template including N=72 data items and indicators, with precise characteristics assigned to identify strata to be used for tabular and graphical outputs.

Different targeted audiences have different statistical outputs assigned to visualize results. The BIRO reports template provides the basis for the development of any regular update of the statistical engine. The web portal is to be designed to strictly comply with its structure and integrated across the whole system.

In the end, the reports template serves as a recipe for the publication of diabetes indicators by the BIRO Consortium, providing the flexibility required to undertake a potential update in the further development of BIRO and EUBIROD.

References

1. Annual Report 2004-2005, Tayside, Scotland
2. de Lusignan S, Sismanidis C, Carey I, DeWilde S, Richards N and Cook D. Trends in the prevalence and management of diagnosed type 2 diabetes 1994–2001 in England and Wales, BMC Fam Pract. 2005; 6: 13. available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1079812
3. Agency for Healthcare Research and Quality, 2002
4. F.Carinci, Valutazione dell'appropriatezza a livello di sistema, Mattoni Project: Measuring Appropriateness, Italian Ministry of Health, 2007
5. IDF (2006), Diabetes Atlas Third Edition
6. Alan Agresti (2002). *Categorical data analysis* (second edition). New York: Wiley

Standardised Diabetes Data Definitions

Scott Cunningham, Graham Leese, Massimo Brillante

ABSTRACT

Background
Core diabetes indicators for the BIRO project were identified in a clinical review of current evidence. In order to allow accurate comparisons of diabetes indicators, a standardised currency for data storage and aggregation is required.

Objectives
To define the BIRO Common Dataset. To explain 'data about the data' through a data dictionary and annexed series of XML documentation that would be widely available to authenticate the contributing information.

Methods
Clinical datasets used by all partner institutions were collated and common data items analyzed for consistency in terms of recording, data definition and units of measurement. Where necessary, data mappings and algorithms were specified in order to allow partners to meet the standard definitions. Data was then assigned a "validity" weight (high, medium or low) describing how well it could meet the agreed definitions. Based on such key requirements, a series of descriptive elements were created to document the 'metadata' for each of the records gathered. The first set of metadata describe the standardised elements, documenting units of measurement, data type, range values, definition, validity, mapping criteria etc. The second set of metadata must be provided by each partner in order to document how consistent their local systems are with the standard definition. Information gathered at this stage includes consistency with definition, completeness of local data and an overall quality score.

Results
The minimum dataset defined N=82 data items, of which N=53 describe clinical data such as clinical and process outcomes, N=21 were created to specify information regarding the local clinical site and N=8 to include geographical references. The classifications provide a method of capture for clinical process data and local geography These mappings and standardised definitions have been used to create an electronic directory for diabetes care, providing the foundation for the BIRO data repository. Clinical and non-clinical information has been gathered using a standardised XML data structure containing supplementary comments required to fully document the data analyses.

Conclusions
The development of data dictionaries and data standards can be used to improve the quality, relevance, consistency and comparability of national information about health. A European minimum common dataset and data dictionary have been developed to become applicable in every existing clinical dataset for diabetes. The process will continue as the project expands to 20 countries via the EUBIROD project.

3.1.1 Introduction

In order to create data analysis repositories for diabetes it is necessary to link to, and create extracts from all relevant sources of diabetes data. Successful systems requiring data extraction, consolidation and analysis[1,2,3] rely heavily on standardisation and definitive data criteria in order to analyze and present information consistently. A similar approach was specified for the BIRO project and it is detailed in this chapter.

European Core Diabetes indicators were identified in a Clinical Review[4] of current evidence. In order to allow accurate comparisons of Diabetes indicators for BIRO, a standardised currency for data storage and aggregation was required, defined as the "BIRO Common Dataset". This represents the first step towards the clear definition of multiple sets of data that could be collected on a routine basis to describe in detail a complex, population-based phenomenon e.g. a chronic disease.

At the same time, It is also essential to explain 'data about the data'. As a result, a Data Dictionary and series of XML Documentation were created to authenticate the contributing information.

This chapter describes the processes followed and the approach applied to reach these objectives.

3.1.2 Materials and Methods

A key objective of the project is to define a minimum common dataset that is applicable to all European partners in the context of their existing datasets.

To this end, an analysis of all datasets maintained in the participating regions was conducted, including an overview of the data formats used to collect clinical data. In addition to the common parameters, supplementary data items have been also defined for collection in order to accurately describe and perform comparative analyses across the sources of data.

The document defining the common dataset has been based on an analysis of the data items and clinical definitions from:

- DiabCare[6]
- Forum for Quality Systems in Diabetes Care (FQSD)[7]
- The Scottish Diabetes Core Dataset[8]
- Umbria Diabetes Register (PROMODR)[9]

In addition, the dataset was cross-referenced against the following datasets to ensure a common and complimentary approach:

(i) BIRO WP2 Clinical Review Indicator Development Results[4]
(ii) EUropean Core Indicators in Diabetes (EUCID)[10]
(iii) The Australian Diabetes, Obesity and Lifestyle (AusDiab) study[11]

Where possible, the International System of Units (SI)[12,13] definitions have been applied to parameters recorded with inconsistent units of measurement across each of the datasets. Mappings have been defined as appropriate to map local datasets to the common structure.

After the creation of the Common Dataset[14], clinical experts have been involved in a qualitative examination of data to ensure that appropriate data mappings and standards have been maintained. Where possible, data item definitions have been compared in order to create one universally acceptable definition. This is the first time that a dataset with corresponding data definitions has been created for a European population. After this step, the focus moved to expand this and create an electronic directory inclusive of concept and data dictionaries for diabetes care and prevention. The BIRO Data Dictionary[15,16] allowed to translate structures and elements of data storage and representation to an XML Schema. This process informs the structure of a relational database that will be available to all partners and which will provide mapping functionality to migrate original datasets to the common structure. The document clearly explains the calculations, translations, elements and mappings required for this purpose.

Each item in the BIRO dataset has been classified as a "*parameter*", with a unique reference code e.g. BIRO1, BIRO2, etc. Parameters may include patient data such as HbA1c, Clinic last attended, a Patient's GP Practice, or parameters used to describe clinical sites such as Clinic Population, Catchment Area Population and Specific Guidelines Used. Each parameter may have several relevant attributes. These are listed and defined in the sub-sections below:

Parameter Attributes

Parameter Attribute: Definition Definition:	The clinical or agreed definition of the parameter
Parameter Attribute: Definition Source Definition:	The published source of the definition
Parameter Attribute: Data Type Definition:	The database storage type. e.g. String, Integer, Decimal, Datetime, etc
Parameter Attribute: Data Type Length Definition:	Numeric value to define the length a String value
Parameter Attribute: Units Definition:	The units of measurement for the data item. e.g. m, mmol/l, kg/m^2, %
Parameter Attribute: Enumerated Types Definition:	If the data items contain a finite number constant values, these should be listed

Parameter Attribute:	Upper Range
Definition:	A defined upper value that cannot be exceeded
Parameter Attribute:	Lower Range
Definition:	A defined lower value that cannot be exceeded
Parameter Attribute:	Mandatory
Definition:	Whether or not the recording of the parameter is compulsory
Parameter Attribute:	Guideline Value
Definition:	The recommended clinical guideline value
Parameter Attribute:	Guideline Source
Definition:	The published source of the guideline value

Clinical parameters may include patient data such as HbA1c, Weight and Retinopathy Screening Status. For data items where there are similarities but mismatches in the units of measurement or definition, a "Data Mapping" is specified in order to translate the local data to the standard BIRO Common Dataset format. A summary of all published Clinical Parameters can be found in the Appendix, Table A.1.

Data will be gathered from a wide range of clinical and administrative data sources, national registers and screening programmes. In order to identify and describe the sources contributing to a data feed, definitions of the various clinic demographics are required. These are necessary in order to perform comparative analyses of the partner sites. Effectively, this includes data regarding the organization of healthcare and the various aspects of service delivery within the contributing data source. Much of this data may not be routinely recorded, but will be required for submission by each site for use in the statistical analysis.

A summary of all published Clinical Site Parameters can be found in the appendix, Table A.6-7.

The original BIRO Data Dictionary made the assumption that data would only be exported for patients currently active within a geographical area, or a specific clinic or data source. This was designed in order to simplify the data extraction process so that data could be compared based on a time-stamped extract.

At a later stage it was agreed that it would be useful for the purpose of estimating indicators that are epidemiologically correct (number of subjects at risk across a timeframe) to be able to track a patients' periods of 'activity' if the contributing data source is able to supply this information.

The changes agreed were to allow a series of activity start and end dates, alongside a corresponding reason for status change. Agreed Activity Start Reason's are:

- Birth
- Diabetes Diagnosis
- Transferred In

Agreed Activity End Reason's are:

- Death
- Transferred Out
- Lost to Follow-up

It is possible for a patient to have one continuous or several disjointed periods of activity based on their diagnosis dates, location of residence or follow up status. A summary of all published Patient Activity Status can be found in the appendix, Table A.2.

As part of discussions towards the end of 2008 it was agreed that it would be of great benefit to capture *Geographical References* for partner territories. The objectives of requirement are to allow the creation of graphical outputs incorporating maps of contributing regions and countries.

Two international standards were considered for this purpose: the "ESRI Shapefiles"[17] and the "of Territorial Units for Statistics (NUTS)"[18] classifications. After careful consideration of the likelihood of each BIRO partner to be able to obtain the required information for these schemes, a modified 'NUTS' classification was agreed as the first approach for capturing geographical information.

The NUTS levels of classification and additional levels of granularity for health systems recording within BIRO are detailed below:

- Level 0 Classification: Continent (BIRO Custom Level)
- Level 1 Classification: Country (NUTS Level 0)
- Level 2 Classification: Sub-National Area (NUTS Level 1)
- Level 3 Classification: Region (NUTS Level 2)
- Level 4 Classification: Local Health Authority (BIRO Custom Level)
- Level 5 Classification: Province (NUTS-3)
- Level 6 Classification: District Health Unit (BIRO Custom Level)
- Level 7 Classification: Post Code (BIRO Custom Level)

It is likely that larger countries will be able to supply NUTS data for their whole country and if data is reported at a 'sub-region' the presentation software must be made aware of the specific area of reference. This consideration has been made when designing the schema. A summary of all published Geographical References can be found in the appendix, Table A.8.

During the regional dataset comparison and analysis, it was clear with some data items that, although the basic concept was the same, the clinical data item or definition contained slight variations.

The concept of the "validity" of a data item has been defined to clearly identify those items for which comparisons may not have a complete correlation.

• *Validity Classification: High*
• Definition: High Validity items are those which are consistent across all analysed datasets.

• *Validity Classification: Medium*
• Definition: Inconsistencies in 1 dataset during comparison
• Example: Blindness - Defined as “Registered Blind” in Scotland but “Receiving Money for Blindness” in DiabCare.

• *Validity Classification: Low*
• Definition: Inconsistencies across >1 datasets
• Example: Ethnic Group is not recorded in the Umbria dataset or DiabCare Basic Information Sheet. Definition is also controversial.□

All Medium validity items will be clearly marked to indicate any discrepancies. Low validity items are only included in the dataset where the data items are required in the core indicators detailed in the Clinical Review.

Data source definitions complete one of the fundamental building blocks of the BIRO applications.
Data sources can define any clinical data source, geographical region or clinical domain providing data for the systems. A high-level architecture of data sources may include a top-level domain covering each partner involved in the process.

For example, at the minimum level, BIRO may consist of 7 top-level Data Sources (Figure 3.1.1):

• DARTS Dataset, Tayside, Scotland
• Umbria Dataset, Italy
• Healthgate Dataset, Austria
• Paulescu Datasets, Romania
• Norway Diabetes Register
• Cyprus Diabetes Register
• Malta Diabetes Register

These top-level domains can be further split to provide a more detailed level of granularity. For example, if BIRO wished to analyse data from within the DARTS dataset, contributing data could be split by Tayside sub-regions and marked appropriately:

• DARTS Dataset - Dundee
• DARTS Dataset - Perth & Kinross
• DARTS Dataset - Angus

This example would allow for the aggregation of data at a sub-regional level (Figure 3.1.2).

Similarly, depending on the types of data sources available within a partner site, different types of contributing data can be specified. For example, by data source type:
• Healthgate Clinical System
• Healthgate Insurance System

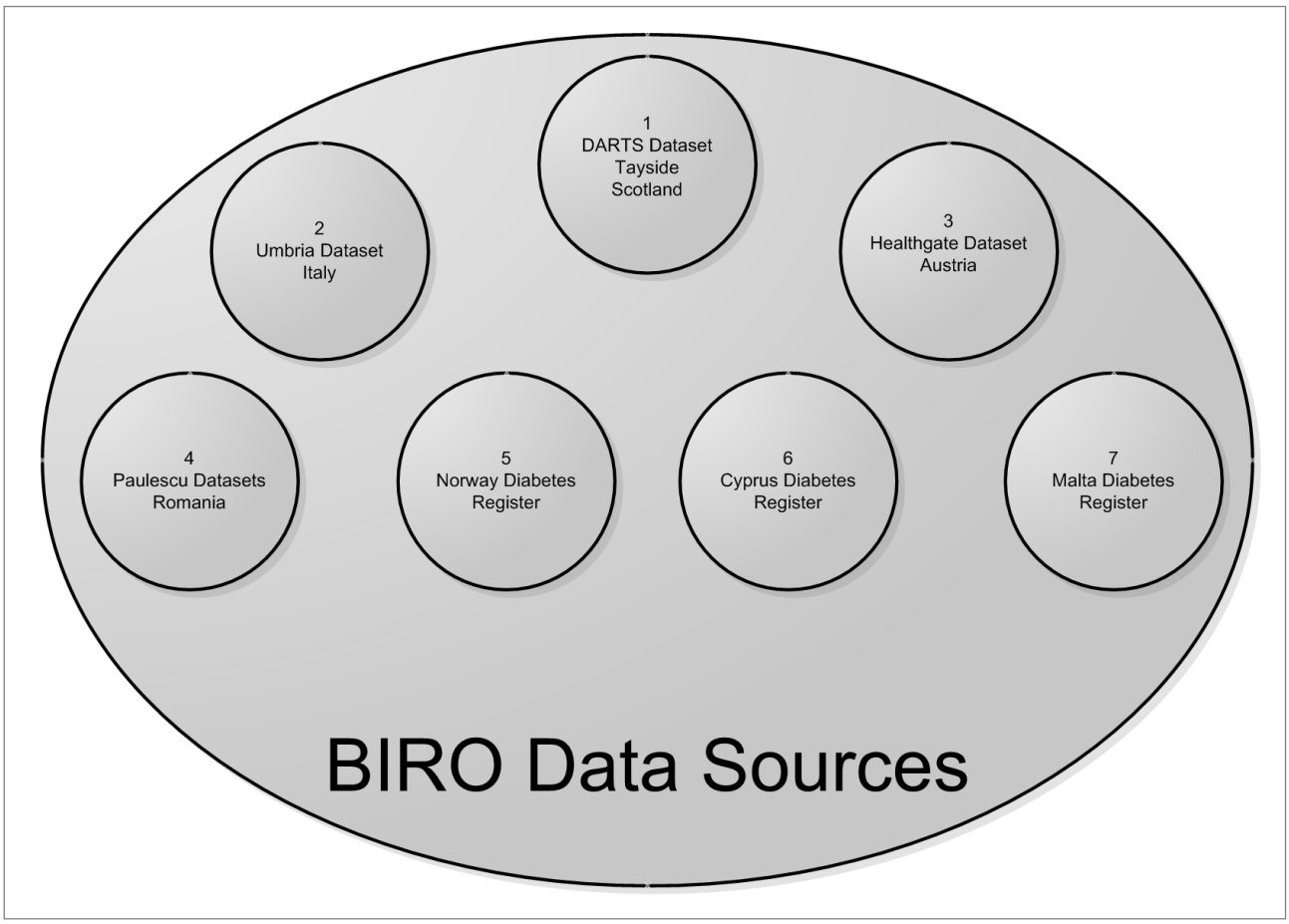


Figure 3.1.1: BIRO Data Sources

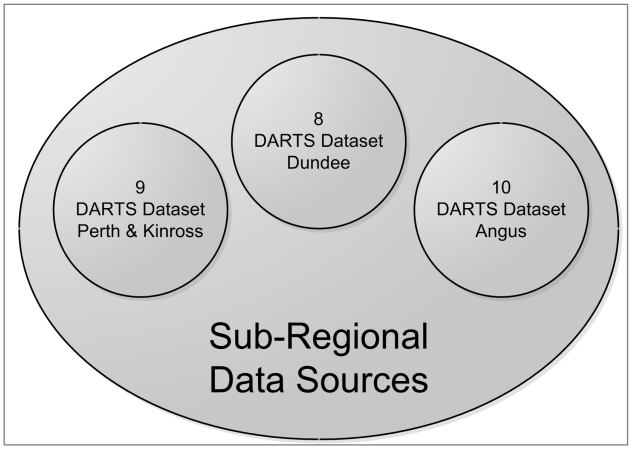


Figure 3.1.2: National Clinical Datasets Development Programme

The Data Dictionary shows how individual data items can be described within each data source. Some data items, for example, may only be available within insurance systems, whereas others may be routinely collected within clinical processes. Each of these scenarios can then be appropriately documented and documented within the BIRO data dictionary.
It is important for every BIRO partner to detail and describe each data source that they wish to separate from their main submissions so that they can be summarized and linked back to the contributing region or country.

In Italy, for example, we can have a straightforward input of Umbria datasets due to their role in the development of BIRO, but this may vary and not directly apply to other regions within the country, regardless of their ability to provide a variety of complete and high quality clinical/administrative datasets. This logical separation of data and subsequent linkage to National contributions will allow consistent and fully documented analyses of results while maintaining the high level of granularity required.

Data will vary depending on the source from which it is gathered. Each clinical site will be given a description to identify the type of data source.

Data Source Type:
GP
Definition:
A site operated by a Primary Care practitioner

Data Source Type:
Hospital Clinic (Internal Medicine)
Definition:
A general secondary care
hospital clinic

Data Source Type:
Hospital Clinic (Diabetes)
Definition:
A specialist Diabetes secondary care clinic

Data Source Type:
Regional Shared-data Register
Definition:
A database containing data gathered from multiple data sources

---

Data Source Type:
Regional Primary Care Project
Definition:
Data gathered from Primary Care which may cover multiple sites

---

Data Source Type:
Disease Management Programme
Definition:
A data source covering patients with a specific complication or disease

---

Data Source Type:
Hospital Discharge Information
Definition:
Administrative hospital discharge system

Data Source Type:
Insurance Programme
Definition:
A source dealing with data relevant for insurance

Data Source Type:
Retinal Screening Programme
Definition:
A source containing data for the purposes of retinal screening

---

Data Source Type:
Diabetes Specialist Nurse Clinic
Definition:
A clinic run by DSN's

Such list is not fixed and is likely to be expanded as new data sources are added.

The development of *data dictionaries and data standards* can be used to improve the quality, relevance, consistency, availability and comparability of national information about the health[8,19]. The rationale for the BIRO Data Dictionary arises from the need for better information and automated collection of data on Diabetes Outcomes throughout the European Community.

Data dictionaries can be used to describe the expected meaning and acceptable representation of data for use within a defined context. The need for consistency of meaning and clinical definition is vital to facilitate information sharing among all end-users of the data. Much of the work involved in establishing a data collection methodology is in the development of data standards to ensure comparability and consistency of the data collected and produced from the collection.

This is imperative for the BIRO model as data is collected across seven European partner sites, all using varying data standards and collection techniques. As the BIRO model expands to further partner sites, this data dictionary will form the cornerstone for data

collection and analysis. Data is shared we need to ensure that all of those who need to use data can clearly understand the meaning, regardless of how the data is collected or stored at source.

A central concept in the construction of the Data Dictionary is that of *metadata*, often referred to as "data about data"[20]. Specifically, it is the definition or structured description of the content, quality, condition or other characteristics of the data. It is well accepted in the world of statistics and database warehousing that metadata leads to better data and subsequent analysis. This is because they enable all people collecting, using and exchanging data to share the same understanding of its meaning and representation.

In BIRO, the Data Dictionary documentation[14] described each of the relevant data items for the project, but also some of the metadata associated with it. For example, information on the test name, units of measurement, clinical definition, data type and a flag indicating the comparability of the data. These representation definitions include how data elements are stored in a computer structure. Well-defined, agreed and precise clinical data definitions are essential to ensure that the data is collected in an internationally consistent way. In order to explain the comparability of data across all BIRO partners, *data quality and consistency* must be fully documented and visible to the end-users of the BIRO system. This information is a key requirement for the statistical analysis and to explain variations between sites for each of the defined indicators.

Data mapping information will be stored in the BIRO Data Dictionary. The various attributes assigned to each mapping will translate to elements in the BIRO Data Dictionary XML Schema.

Some systems contributing to the BIRO data analysis store clinical data in a format that is inconsistent to the BIRO dataset. In most cases, it is possible to map from the source data format to the BIRO format. It was agreed during the Common Dataset development that The International System of Units (SI) would be used as the standard storage format.

Data item assessment will be carried out for each item in the BIRO dataset. Every partner will be required to assess its quality in comparison with the agreed definitions. Although not an insignificant piece of work, this information will provide substantial metadata describing the feeding systems and will contribute considerably to the interpretation of the indicators. This will make the BIRO outcomes unrivalled in terms international data analysis.

A number of criteria must be applied to the data items by each partner. This assessment exercise must detail:

consistency with the BIRO definition
- High: Exact match
- Medium: Minor discrepancy - e.g. Source units require mapping
- Low: Major discrepancy - e.g. mapping unavailable
- Data Item Unavailable

Box 3.1.1: Data quality information in BIRO

Data Completeness

All data for this item should be complete and accurate for all patient encounters
This describes well-recorded, accurate data that is suitable for any analysis

Data is not routinely collected
This is data that may be recorded in specific situations, but not for all patients

Data is only collected for a subset of the population
This is data that has been collected for a portion of the population, but is not routinely collected for all

Data Collection

Data is entered immediately by the medical staff
This is data that is entered into the source system by clinical staff during the consultation

Data is entered retrospectively by medical staff
This is data that is entered into the source system by clinical staff after the consultation

Data is entered retrospectively by administrative staff
This is data that is entered into the source system by non-clinical staff after the consultation

Data Values

The laboratory analyser is not standardised for this test
Data analysis technology is inconsistent across laboratory sites

The results are estimates only
Data are not actual results, but approximate values

completeness
- Complete: Full data available for all currently diagnosed patients
- Incomplete: Partial data available for all currentlydiagnosed patients
- Data Item Unavailable

overall quality score: a value judgement on the ability for the data source to provide complete and consistent data in line with the definition
- High: Can provide complete and consistent data
- Medium: Minor completeness and consistency issues
- Low: Major completeness and consistency issues
- Data Item Unavailable

comments: This field should be used to describe any further information known about the data item at source that may affect longitudinal analysis or data presentation

The example below explains the submission for the 'HBA1C' field within the DARTS dataset in Tayside, Scotland.

Data Item:	HBA1C
Consistency:	High
Completeness:	Complete
Quality Score:	High
Comments:	In Tayside, HbA1c was DCCT aligned in August 2002.

Further examples of data quality information of interest are given in Box 3.1.1.
This type of information provides key insights and local knowledge of data and may be displayed alongside BIRO outputs and statistics.

It should also be possible to 'drill-down' and view the data quality information supplied by each data source for the contributing data items for the indicator.

3.1.3 Results

The practical implementation of the Common Dataset and associated Data Dictionary (see Tables in Appendix) has been heavily based on the realization of the correspondent *XML definitions*.

EXtensible Markup Language XML is commonly defined as a standardised specification for the creation of structured data for the purposes of information sharing[5].

The BIRO infrastructure has built upon the Dataset and Data Dictionary defined to create a series of XML specifications to aid data capture.
The <BIRODataSet.xsd> contains the clinical parameters used in BIRO. This schema simply details the BIRO data fields identified in the core dataset alongside their unique identifiers.
The purpose of this file is to act as a parameter reference file for the following schemas:

1.ECClinicalDefinitions.xsd
2.ECDataExtractDefinition.xsd
3.ECDataSourceExtractDefinition.xsd

An example of how data is stored in the schema is shown below:

```
<xsd:simpleType name="BIRODataSet">
      <xsd:restriction base="xsd:string">
            <xsd:enumeration value="PAT_ID" id="BIRO001"/>
            <xsd:enumeration value="DS_ID" id="BIRO002"/>
            <xsd:enumeration value="TYPE_DM" id="BIRO003"/>
            .
            .
            .
            .
            <xsd:enumeration value="SELF_MON" id="BIRO044"/>
            <xsd:enumeration value="EDUCATION" id="BIRO045"/>
      </xsd:restriction>
</xsd:simpleType>
```

As new parameters are added to this schema over time, these will simply proliferate and immediately become available for use in the remaining schemas listed above and defined below.
The <ECClinicalDefinitions.xsd> is the schema that maps the comm on dataset items to their clinical definitions. It allows the input of a full version history, including associated comments and reason for update. The definitions may also be held in an unlimited number of languages, therefore allowing local translations to be recorded. The XML file associated with this schema will be maintained centrally and will tie in with the Clinical Indicators, and Common Dataset documentation.

A version history will be available within the document to detail changes:

```
<VersionHistory>
      <Version>0.2</Version>
      <VersionDate>2007-05-25</VersionDate>
      <VersionComments>Amended during Cyprus meeting</VersionComments>
</VersionHistory>
```

Version: The value assigned to the version number
VersionDate: The date of the numbered document version
VersionComments: Additional freetext used to describe changes

Supported languages can be defined using the following XML:

```
<SupportedLanguage>
      <Language>EN</Language>
      <TranslatorName>Scott Cunningham</TranslatorName>
      <TranslatorEmail>scott.cunningham@nhs.net</TranslatorEmail>
</SupportedLanguage>
```

Language: This refers to the language of reference using a two-digit language code
TranslatorName: This is the person responsible for the definition of the BIRO data items in the language specified
TranslatorEmail: The email address of the person responsible for the translation in the language specified.

This schema refers to the <BIRODataSet.xsd> in order to allow only entries for those items currently valid in BIRO. These DataItem's can then be defined individually:

```
<DataItem>
 <FieldName>HEIGHT</FieldName>
 <DateDataItemReviewed>2007-05-25</DateDataItemReviewed>
 <DateDataItemUpdated>2007-05-25</DateDataItemUpdated>
 <DataItemDefinition>
  <Language>EN</Language>
  <ClinicalDefinition>Height in metres - measured without shoes.
   It is particularly important to measure regularly the height
   of children. In adults a single recording will usually
   be sufficient
  </ClinicalDefinition>
 </DataItemDefinition>
 <Units>m</Units>
 <LowerRange></LowerRange>
 <UpperRange></UpperRange>
 <DataMapping>
  <SourceUnits>cm</SourceUnits>
  <MultiplicationFactor>100</MultiplicationFactor>
  <MappingComments>Height in centimetres can be mapped
   to metres by multiplying by 100
  </MappingComments>
 </DataMapping>
 <Mandatory>false</Mandatory>
 <Validity>High</Validity>
</DataItem>
```

FieldName: The standard BIRO field name as defined in BIRODataSet.xsd
DateDataItemReviewed: Most recent date of review
DateDataItemUpdated: Most recent date of update
DataType: The data type to be applied within the database structure
DataItemDefinition
DataItemDefinition Language: The language of the definition
DataItemDefinition ClinicalDefinition: The clinical definition of the data item in the language specified
DataItemDefinition LanguageComments: Any additional comments on the definition
Units: Standard BIRO-compatible units of measurement
LowerRange: The acceptable lower range of the data item
UpperRange: The acceptable upper range of the data item
DataMapping: New complex element used to record details of mappings from non-BIRO units to standard units of measurement. This allows any number of mappings for an individual data field
DataMapping SourceUnits: Original units of measurement in source dataset
DataMapping MultiplicationFactor: Value to multiply source units by to calculate BIRO-acceptable units
DataMapping DivisionFactor: Value to divide source units by to calculate BIRO-acceptable units
DataMapping MappingDetail: Details of any mapping that cannot be quantified mathematically
DataMapping MappingComments: Freetext comments associated with mapping
Mandatory: A flag indicating whether the field must always be recorded

Validity: The ranking given in WP3 for the data item (High/Medium/Low)
DataItemReferences: Details of published peer reviewed literature relevant to the data item
DataItemReferences Reference: The full reference of the article
DataItemReferences ReferenceComments: Any freetext comments associated with the reference
DataItemComments: Any additional information associated with the data item

---

This schema also ties up each of the defined data items with the BIRO candidate indicators under the "OutcomeIndicator" tag. For each indicator, a reference is provided, along with details of the contributing data items and the possible algorithms (e.g. R, SAS, SQL, etc) required to calculate the indicator.

```
<OutcomeIndicator>
 <IndicatorReference>1</IndicatorReference>
  <DateIndicatorReviewed>2007-03-25</DateIndicatorReviewed>
  <DateIndicatorUpdated>2007-03-25</DateIndicatorUpdated>
  <IndicatorDefinition>
   <Language>EN</Language>
   <IndicatorText>Annual Incidence of Type 1
     Diabetes in children between 0 - 14 years of age
     at diagnosis (clinical) per 100,000 children
   </IndicatorText>
   <Numerator>Number of children between 0-14 yrs,
     diagnosed (clinical) within the last 12 months
     with type 1 diabetes mellitus
   </Numerator>
   <Denominator>Total number of children between 0-14 yrs
     in the study region/country/100,000
   </Denominator>
   <Source>EUDIP</Source>
   <DatasetIssues>None</DatasetIssues>
  </IndicatorDefinition>
  <ContributingData>PAT_ID</ContributingData>
  <ContributingData>DS_ID</ContributingData>
  <ContributingData>TYPE_DM</ContributingData>
  <ContributingData>DOB</ContributingData>
  <Algorithm>
   <ComputerLanguage>Pseudocode</ComputerLanguage>
   <AlgorithmCalculation>
     Total Patients (PAT_ID) /
     (Data Source Denominator (DS_DENOM) / 100000)
     With Type 1 Diabetes (TYPE_DM = 1)
      Grouped By Year of Birth (in DOB)
     and Data Source ID (DS_ID)
   </AlgorithmCalculation>
   <AlgorithmOutput>
     Number of Type 1 patients/100000
     grouped by year and by data source.
     Reference to age bandings defined
     in section 8 of this document.
   </AlgorithmOutput>
  </Algorithm>
</OutcomeIndicator>
```

IndicatorReference: The ID of the indicator that is referred to
DateIndicatorReview: Date of last review of the indicator
DateIndicatorUpdated: Date of last update of indicator
IndicatorDefinition: The full definition of the indicator
IndicatorDefinition Language: The language that the indicator is defined in
IndicatorDefinition IndicatorText: The full text title of the indicator
IndicatorDefinition Numerator: The numerator used to calculate the indicator
IndicatorDefinition Denominator: The denominator used to calculate the indicator
IndicatorDefinition Source: The source that originally defined the indicator (e.g. EUDIP, OECD)
IndicatorDefinition DatasetIssues: Details of known dataset compatibility issues
IndicatorDefinition IndicatorDescriptionComments: Any other comments associated with the indicator
ContributingData: list of data items contributing to the indicator are listed here. Each clinical outcome needs to derive its data from fields.

Algorithm: Details of algorithms used to calculate an indicator in any computing language are stored here
Algorithm ComputerLanguage: The language used in the following algorithm (SQL, R, SAS, etc)
Algorithm AlgorithmCalculation: Full syntax of the calculation
Algorithm AlgorithmOutput: The specific output produced from the calculation
Algorithm AlgorithmComment: Any addition comments about the algorithm

---

In addition to the data items defined above, it is possible for some data items to be calculated based on other existing parameters:

```
<CalculatedData>
 <CalculatedField>BMI</CalculatedField>
 <CalculationDataItem>HEIGHT</CalculationDataItem>
 <CalculationDataItem>WEIGHT</CalculationDataItem>
 <Calculation>WEIGHT / (HEIGHT * HEIGHT)</Calculation>
</CalculatedData>
```

CalculatedData: Data items that can be calculated based on other data items
CalculatedDataCalculatedField: The data item that can be calculated using other data items
CalculatedDataCalculationDataItem: An unbounded list of data items that contribute to the calculated data item.
CalculatedDataCalculation: The definition of how the data item can be calculated
CalculatedData : Any other information about the calculated value

Within BIRO several indicators, data will be split based on age criteria. The following section explains how these groupings will based:

```
<AgeBandings>
 <BandingID>BAND1</BandingID>
 <LowerAge>0</LowerAge>
 <UpperAge>14</UpperAge>
</AgeBandings>
```

AgeBandings: The structure defining the age bands
AgeBandings BandingID: A unique identified for the age band.
AgeBandings LowerAge: The lowest age contributing to the defined age band
AgeBandings UpperAge: The upper age contributing to the defined age band
AgeBandings AgeBandingComment: Any additional freetext available regarding the age banding

---

The <ECDataExport.xsd> is used to define the structure for the patient data extract required from each BIRO partner. This data extract is defined with profile data that is recorded only once (e.g. data of diagnosis, type of diabetes, etc) separated from clinical results that may be recorded several times (e.g. HbA1c, blood pressure, etc). This format is flexible enough to allow any number of clinical results to be recorded although each XML file must only contain data for one patient.

Each clinical result must be recorded with an associated date of episode, alongside the following fields:
BIRO Data Item
Value of result

The full details of the export are defined within the XML Schema. The following section describes the XML elements required for each patient:

```
<Profile>
 <ProfileFieldName>PAT_ID</ProfileFieldName>
 <ProfileFieldValue>2222222222</ProfileFieldValue>
</Profile>
<Profile>
 <ProfileFieldName>TYPE_DM</ProfileFieldName>
 <ProfileFieldValue>2</ProfileFieldValue>
</Profile>
<Profile>
 <ProfileFieldName>SEX</ProfileFieldName>
 <ProfileFieldValue>2</ProfileFieldValue>
</Profile>
<Profile>
 <ProfileFieldName>DOB</ProfileFieldName>
 <ProfileFieldValue>1927-11-10</ProfileFieldValue>
```

```
</Profile>
<Profile>
 <ProfileFieldName>DT_DIAG</ProfileFieldName>
 <ProfileFieldValue>1981-06-06</ProfileFieldValue>
</Profile>
<EpisodeData>
 <EpisodeDate>2005-07-04</EpisodeDate>
 <Data>
  <EpisodeFieldName>HBA1C</EpisodeFieldName>
  <EpisodeFieldValue>6.5</EpisodeFieldValue>
 </Data>
 <Data>
  <EpisodeFieldName>HEIGHT</EpisodeFieldName>
  <EpisodeFieldValue>150</EpisodeFieldValue>
 </Data>
 <Data>
  <EpisodeFieldName>SBP</EpisodeFieldName>
  <EpisodeFieldValue>140</EpisodeFieldValue>
 </Data>
  <Data>
   <EpisodeFieldName>DBP</EpisodeFieldName>
   <EpisodeFieldValue>80</EpisodeFieldValue>
  </Data>
</EpisodeData>
```

Profile: The patient profile is non-event-based data such as surname, date of diagnosis and date of birth

ProfileProfileFieldName: Standard BIRO field name
ProfileProfileFieldValue: Result of the data item above
EpisodeData: Patients have events that happen chrolonogocally (patient episodes)
EpisodeDataEpisodeDate: Date of patient episode
EpisodeData : Data corresponding to the patient episode
EpisodeDataDate EpisodeFieldName: Standard BIRO field name
EpisodeDataDate EpisodeFieldValue: Result of field specified

The <ECDataSourceExport.xsd> is used to define the profile of the clinical site providing the patient data extract. This may be an individual clinic, full regional database, national sample, etc and must be described using the category that is most appropriate. This allows a picture to be created of the data source and the associated meta-data can be used to highlight inconsistencies with local mappings to data definitions, ultimately leading to presentation on BIRO reports. The first section of this schema allows for the description of general clinic contact information.

```
<SiteHeader>
 <DateHeaderInformationChecked>2007-03-25</DateHeaderInformationChecked>
 <DS_ID>1</DS_ID>
 <DS_WEBSITE>http://www.diabetes-healthnet.ac.uk</DS_WEBSITE>
 <DS_ADDRESS_1>Diabetes Centre</DS_ADDRESS_1>
 <DS_ADDRESS_2>Level 8</DS_ADDRESS_2>
 <DS_ADDRESS_3>Ninewells Hospital</DS_ADDRESS_3>
 <DS_POST_CODE>DD1 9SY</DS_POST_CODE>
 <DS_COUNTRY>Scotland</DS_COUNTRY>
 <DS_C_CONTACT>Dr Graham Leese</DS_C_CONTACT>
 <DS_C_EMAIL>graham.leese@tuht.scot.nhs.uk</DS_C_EMAIL>
 <DS_T_CONTACT>Scott Cunningham</DS_T_CONTACT>
 <DS_T_EMAIL>scott.cunningham@nhs.net</DS_T_EMAIL>
 <HeaderComments>DARTS Dataset - Shared Patient Record
  for Tayside, Scotland
 </HeaderComments>
</SiteHeader>
```

SiteHeader: General information about the data source
SiteHeaderDateHeaderInformationChecked: Date of last review
SiteHeaderDS_ID: Unique centre identification number (Defined as a BIRO Clinical Site)
SiteHeaderDS_WEBSITE: Internet address for Data Source
SiteHeaderDS_ADDRESS_1: First line of Data Source address

SiteHeaderDS_ADDRESS_2: Second line of Data Source address
SiteHeaderDS_ADDRESS_3: Third line of Data Source address
SiteHeaderDS_ADDRESS_4: Fourth line of Data Source address
SiteHeaderDS_POST_CODE: Post Code of data source
SiteHeaderDS_COUNTRY: Country in which data source exists
SiteHeaderDS_C_CONTACT: Clinical contact for data source
SiteHeaderDS_C_EMAIL: Email address for clinical contact
SiteHeaderDS_T_CONTACT: Technical contacts for data source
SiteHeaderDS_T_EMAIL: Email address for technical contact
SiteHeaderHeaderComments: Comments regarding the header information

---

The following section allows the storage of data regarding clinic demography.

```
<SiteProfile>
 <DateProfileInformationChecked>2007-03-25</DateProfileInformationChecked>
 <DS_TYPE>4</DS_TYPE>
 <DS_DENOM>385000</DS_DENOM>
 <DS_AREA>1</DS_AREA>
 <DS_BEDS>1</DS_BEDS>
 <DS_PHYSICIANS>1</DS_PHYSICIANS>
 <DS_DIABETOLOGISTS>1</DS_DIABETOLOGISTS>
 <DS_DOCTORS>1</DS_DOCTORS>
 <DS_DSN>1</DS_DSN>
 <DS_PROGS>1</DS_PROGS>
</SiteProfile>
```

SiteProfile: Details regarding the profile of the data source
SiteProfileDateProfileInformationChecked:
SiteProfileDS_TYPE: The type of source from which data has been extracted
SiteProfileDS_DENOM: Current data source population (with or without diabetes)
SiteProfileDS_AREA: The total population of patients with known diabetes in the catchment area of the clinic
SiteProfileDS_BEDS: Total hospital beds within data source geographical area - not separated by category
SiteProfileDS_PHYSICIANS: Physicians within data source geographical area. National statistics can provide information on this indicator
SiteProfileDS_DIABETOLOGISTS: Diabetologists within data source geographical area. Data should come from national Specialist Registers
SiteProfileDS_DOCTORS: Number of doctors who regularly take care of diabetic patients in diabetes clinics in primary or secondary care within data source geographical area
SiteProfileDS_DSN: Specialist diabetes nurses within data source geographical area
SiteProfileDS_PROGS: Number of disease management programmes in data source geographical area. Availability of a DMP influences the level of structured and evidence based treatment
SiteProfileDS_ProfileComments: Additional comments regarding the site profile

---

Most importantly, this schema also allows for data to be held regarding each data items in the BIRO dataset in terms of its quality and consistency with the BIRO definition.

```
<FieldExportProfiles>
 <FieldName>DT_DIAG</FieldName>
 <DateStatusLastReviewed>2007-03-25</DateStatusLastReviewed>
 <Recorded>true</Recorded>
 <Consistency>High</Consistency>
 <Completeness>95%</Completeness>
 <Mandatory>false</Mandatory>
 <Routine>true</Routine>
 <QualityScore>High</QualityScore>
</FieldExportProfiles>
```

FieldExportProfiles: Metadata regarding the BIROdata items from the specified data source
FieldExportProfilesFieldName: The BIRO data items that is being described
FieldExportProfilesDateStatusLastReviewed: Date of last review
FieldExportProfilesRecorded: Whether or not the data source records this data item
FieldExportProfilesConsistency: An indicator detailing how well the source data complies with the BIRO definition
FieldExportProfilesCompleteness: A percentage indicating how much of the data item can be identified
FieldExportProfilesMandatory: Whether or not the recording of the data item is mandatory at source
FieldExportProfilesRoutine: Whether or not the recording of the data item is routine at source

FieldExportProfilesQualityScore: An objective overall score describing how well the source can meet the data requirements
FieldExportProfilesFieldExportComments: Any additional free text comments

Patient Activity status can be represented using the <ECDataExport.xsd> schema.
This data can be repeated to allow several periods of patient activity.

```
<ActivityData>
<StartDate>2005-01-01</StartDate>
<StartReason>Diagnosis</StartReason>
<EndDate>2009-01-01</EndDate>
<EndReason>Death</EndReason>
</ActivityData>
```

StartDate: The date of the period of care
StartReason: Reason for inclusion in diabetes care systems
Enddate: The date of the end of this period of care
EndReason: Reason for exclusion for diabetes care systems

A method of recording geographical areas was also defined to import potentially large NUTS data files.

```
<GeoClassification>
<Continent>EU</Continent>
<Country>IT</Country>
<MacroRegion>ITE</MacroRegion>
<Region>ITE2</Region>
<HealthAuthority>101</HealthAuthority>
<Province>ITE21</Province>
<DistrictUnit>10101</DistrictUnit>
<PostCode>54011</PostCode>
</GeoClassification>
```

Continent: At present, this will always be EU, until the project expands beyond Europe
Country: The country of the data source being described (NUTS-0)
MacroRegion: A NUTS-1 sub-national area
Region: A NUTS-2 region
HealthAuthority: A BIRO modification to allow the capture of health board information
Province: A NUTS-3 compliant province
DistrictUnit: A BIRO modification to allow sub-health-board analyses
PostCode: The lowest level of geographical data

For patient data it is desirable to identify their physical location for monitoring purposes. Depending on the level of detail allowed, the <ECDataExport.xml> files can now include the following for patient geo-referencing.

```
<GeoRef>Province</GeoRef>
<GeoValue>ITE21</Geovalue>
```

GeoRef: The GeoClassification point referred to
GeoValue: The precise location of the GeoRef indicated.

3.1.4 Discussion

The key objective of BIRO is to allow the automated comparison of national Diabetes indicators across Europe. In addition to this it is also desirable to browse the BIRO dataset and data dictionary defined to gain an understanding of the data collected. Existing data dictionary and data standards resources allow the end user to browse the dataset and associated metadata, as further detailed in the BIRO Reports Template[22].

In the UK, there has been a longstanding approach to the development of *datasets and data dictionaries*. The Scottish Intercollegiate Guidelines Network[15] published their first minimum dataset for people with Diabetes in 1998 in SIGN 25. Since then, NHS Scotland has developed the National Clinical Datasets Development Programme[16] which now allows now holds the current datasets across all specialties online (Figure 3.1.3).

In addition to allowing the user to search for, and to browse the details of every single data item within these datasets, all related metadata is also made available to the user. The BIRO website will provide similar functionality containing the content defined in

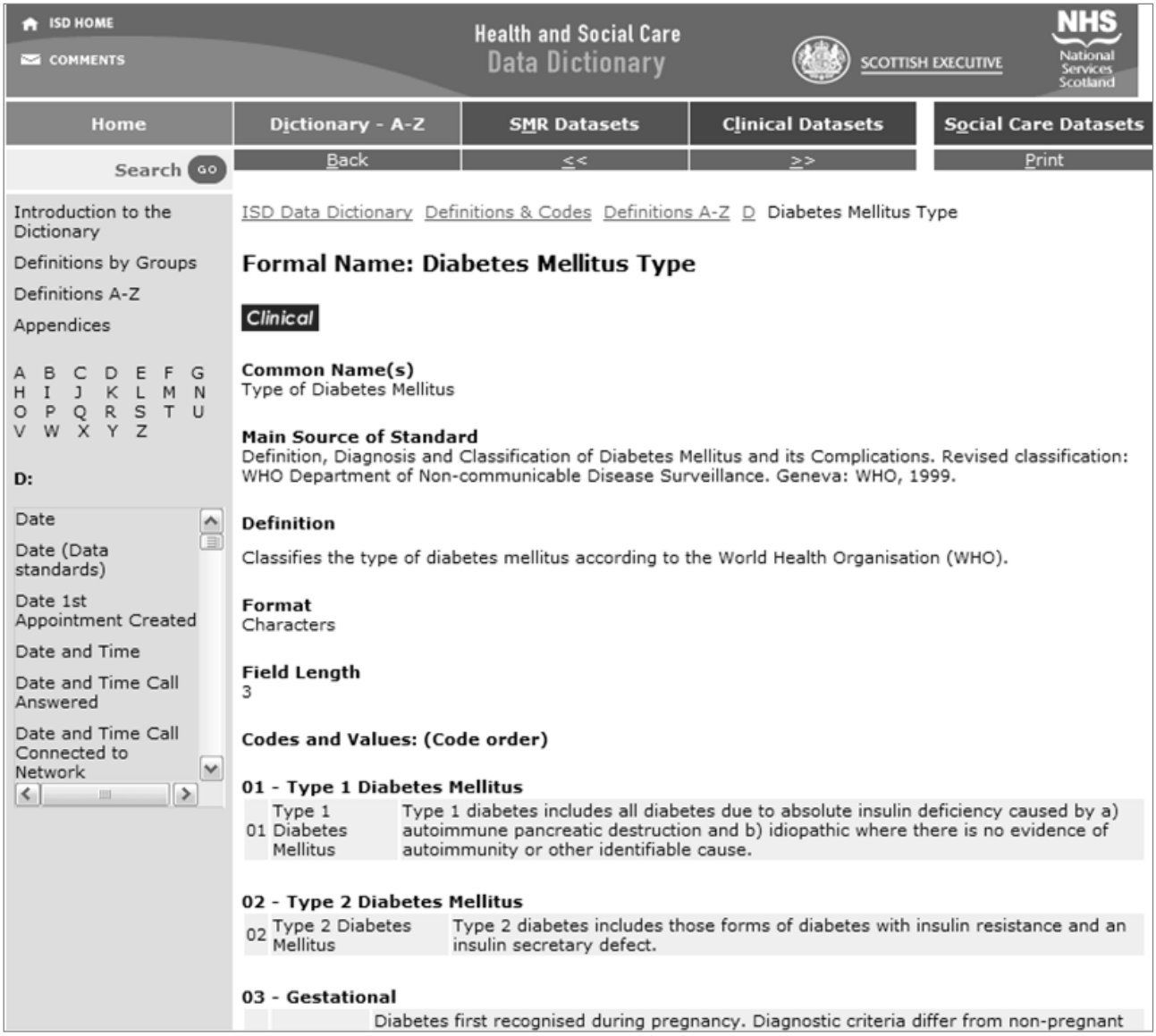


Figure 3.1.3: Aggregation of data at a sub-regional level

this document, and the contents supplied by each of the partner institutions detailing the intricacies of their data.

The NHS in England and Wales[17] and Australian Institute of Health and Welfare Metadata Online Registry[18] provide similar resources, and many more examples are available online.

'Geocode' information can now be collected and subsequently displayed online concurrently with clinical results. The main benefits being that specific geographical areas can be highlighted and marked depending on partner performance.

Information of this type can be used to highlight a country, e.g. Germany (Figure 3.1.4) on a map of Europe. The various levels of granularity can then allow the user to 'drill-down' to the sub-regional level (Figure 3.1.5).

It is believed that this level of data presentation and display will lead to a greater level of user satisfaction and interest when browsing the final BIRO website.

In order to provide a low-level description of confounding factors and to outline discrepancies in data collection, manipulation and data storage between partners, it is necessary to capture metadata. A series of metadata collection principles were outlined in order to facilitate the speedy collection, storage and update of this information, an online data entry tool has been created.

This tool will allow representatives from each partner to submit local knowledge regarding every BIRO data item, but will remain online to allow easy maintenance and update. The information collected will provide extremely powerful and unique commentary on the web outputs created by the Central BIRO Engine.e.

The online questionnaire consists of 5 main sections:

Login and Data Source Selection
Within each partner country, one or more local data

sources may be described and documented. Once the user has logged on to their account, they will be shown the existing data sources that they can edit, but also have the opportunity to add a new source for their country (Figure 3.1.6).

Site Header
The second section ((Figure 3.1.7) allows for the entry of the administrative contact details associated with the specific data source. This allows partners to distribute responsibility for sub-regional data sources within their country.

Site Profile
This section (Figure 3.1.8) allows the entry of some aggregated data related to the data source being described. These may not be relevant for all data sources, but for each clinical data source (e.g. DARTS, Umbria) this is necessary.

Field Export Profiles
For each data item within every source, BIRO needs to obtain information about the data quality, completeness and consistency with the BIRO definitions. It is important that the designated partner representative entering the data into this section has considerable in-depth knowledge regarding data quality and completeness within their designated area.

Particularly useful will be the free-text comments section which will allow the user to provide a commentary on any issues or features they are aware of contained within their local data.

This data entered in this section (Figure 3.1.9) will be used for further presentation alongside the final outputs. The following screen within the wizard shows how the application loops through every BIRO data item in turn until data is completed for each.

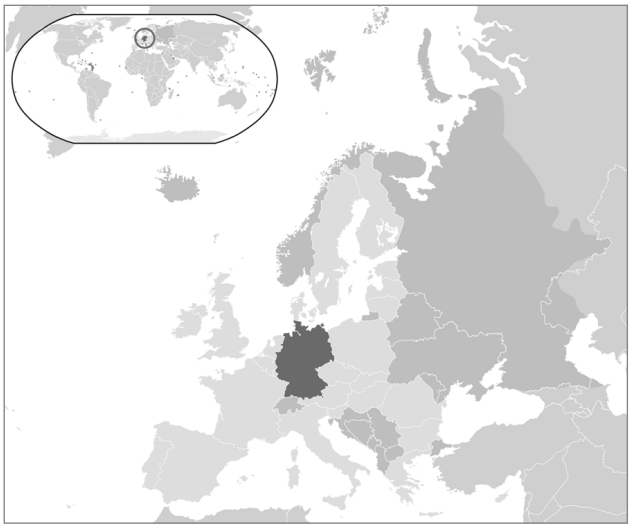


Figure 3.1.4: Highlighting a country on a map: Germany

Figure 3.1.5: Drilling down to the sub-regional level

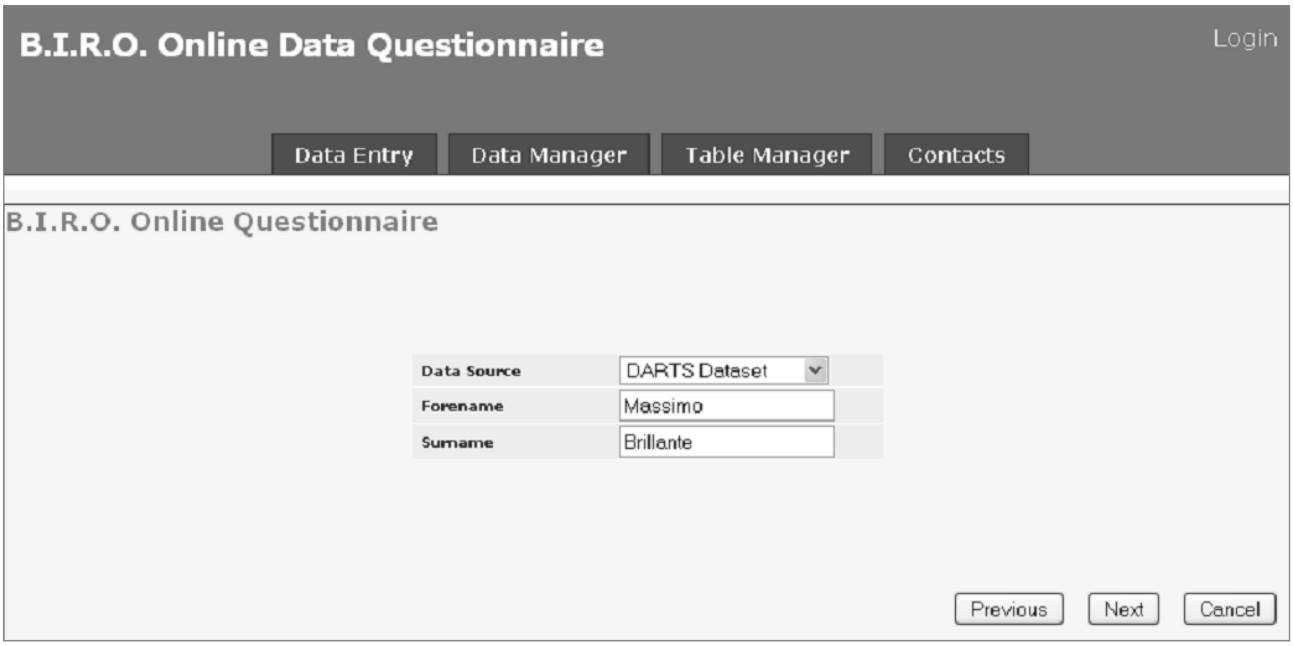


Figure 3.1.6: and Data Source Selection

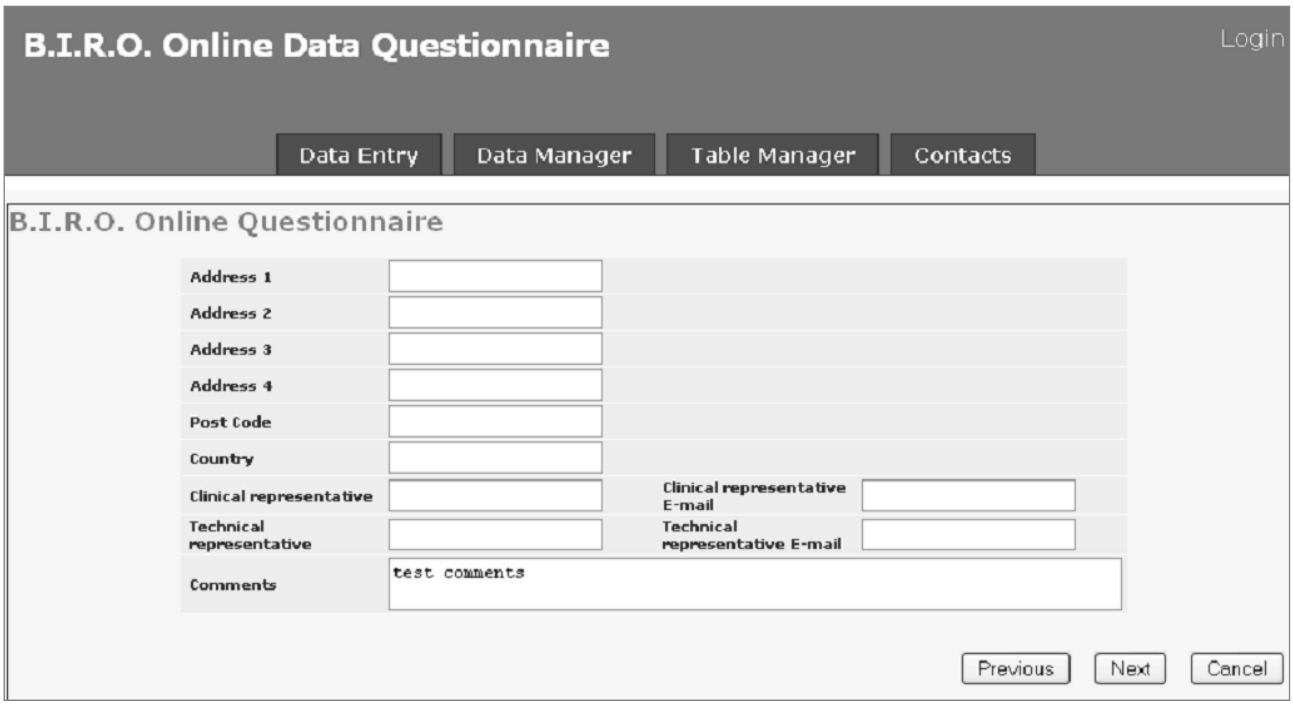


Figure 3.1.7: Site Header

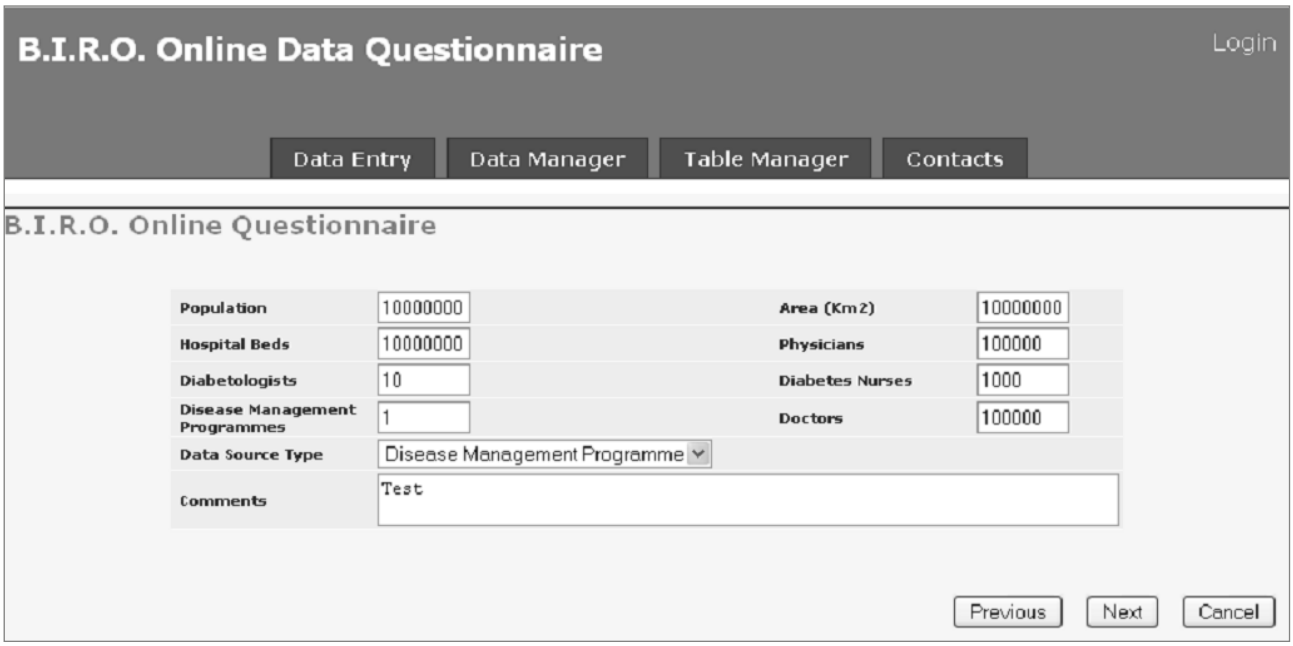


Figure 3.1.8: Site Profile

Summary of Data Entry
On completion of data entry, a summary screen (Figure 3.1.10) is produced for review and validation. At this point, all data entered will be stored to database, meaning that any future updates simply build on the data entered during the first submission.

The full data capture process should take no longer than 15 minutes the first time it is completed. As the system saves any previous submissions, subsequent action consists of only updating data items that have either changed or been added. It is recommended that all data source metadata is reviewed by a local representative at least annually.

As a result of the online data capture, it is now possible for all descriptive data source to generate *dynamic XML documentation*.
This documentation is essential for transmission to the central engine for subsequent display online, and this process improves reliability, consistency and validity. The questionnaire results are captured in a database and from there they can be
translated into the appropriate XML format.

By aligning this with the agreed schema, validity checks can be performed during generation and in addition to reducing the manual overhead of creating these files, a consistent approach can be applied.

The example shown in Figure 3.1.11 shows a sample extract generated by the software for the DARTS, Tayside data source area.

The naming convention of these files is as follows: Datasource_X_Extract.xml, where 'X' is the unique data source ID assigned to the clinical domain being described.

3.1.5 Conclusions

The development of data dictionaries and data standards can be used to improve the quality, relevance, consistency and comparability of national information about health. A European minimum common dataset for diabetes has been created based on an in-depth analysis of all contributing data sources. Within this

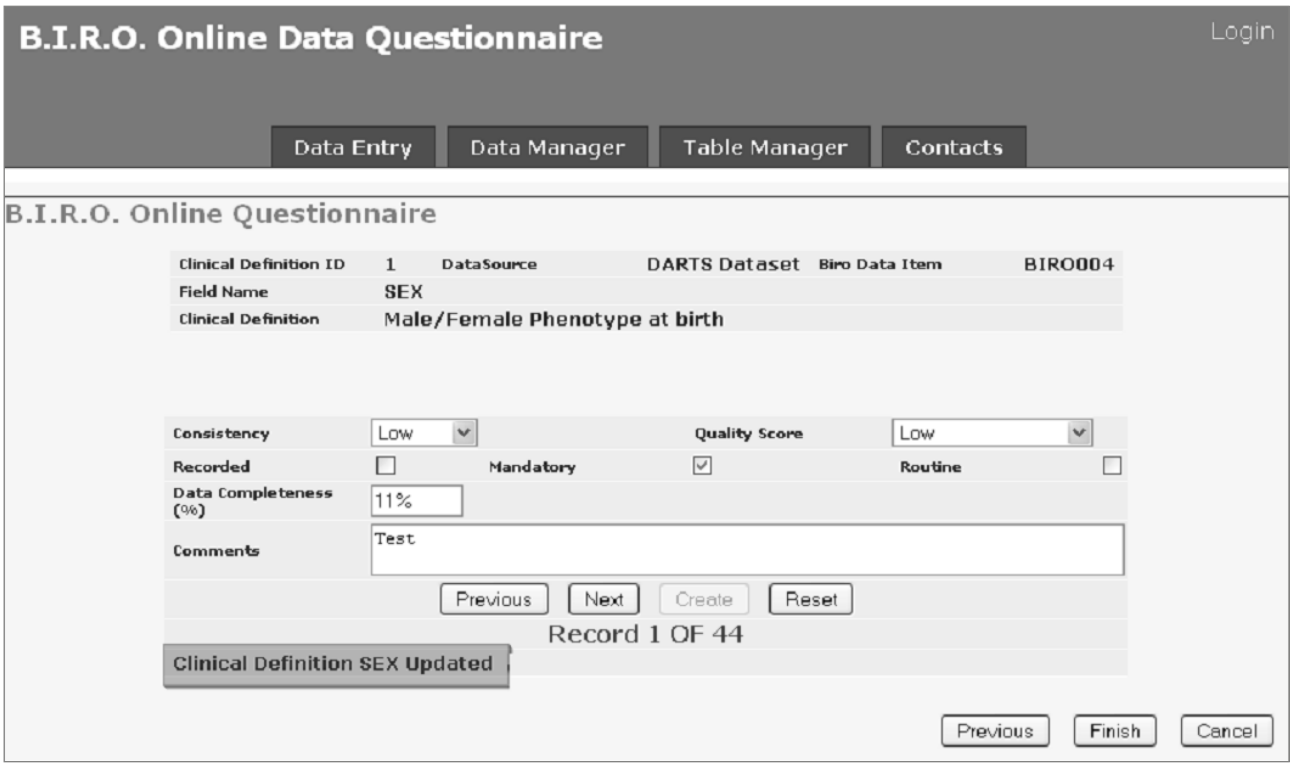


Figure 3.1.9: Field Export Profiles

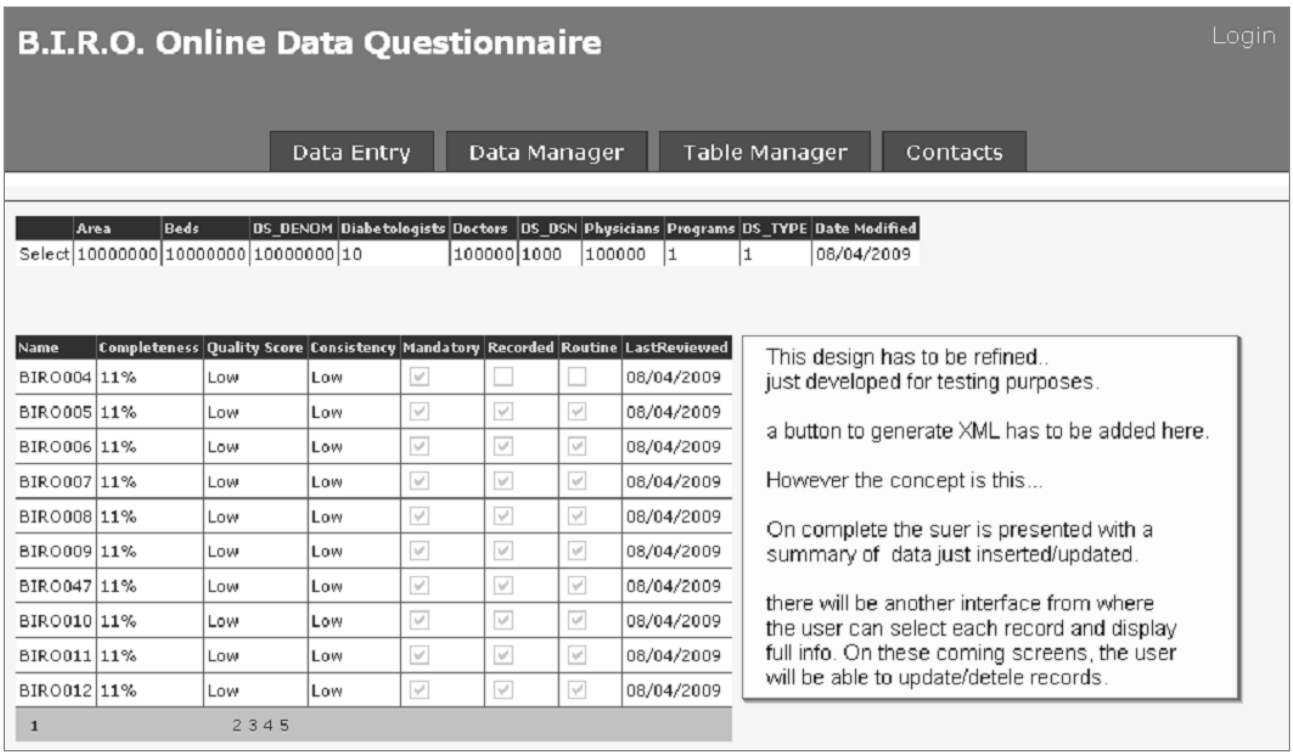
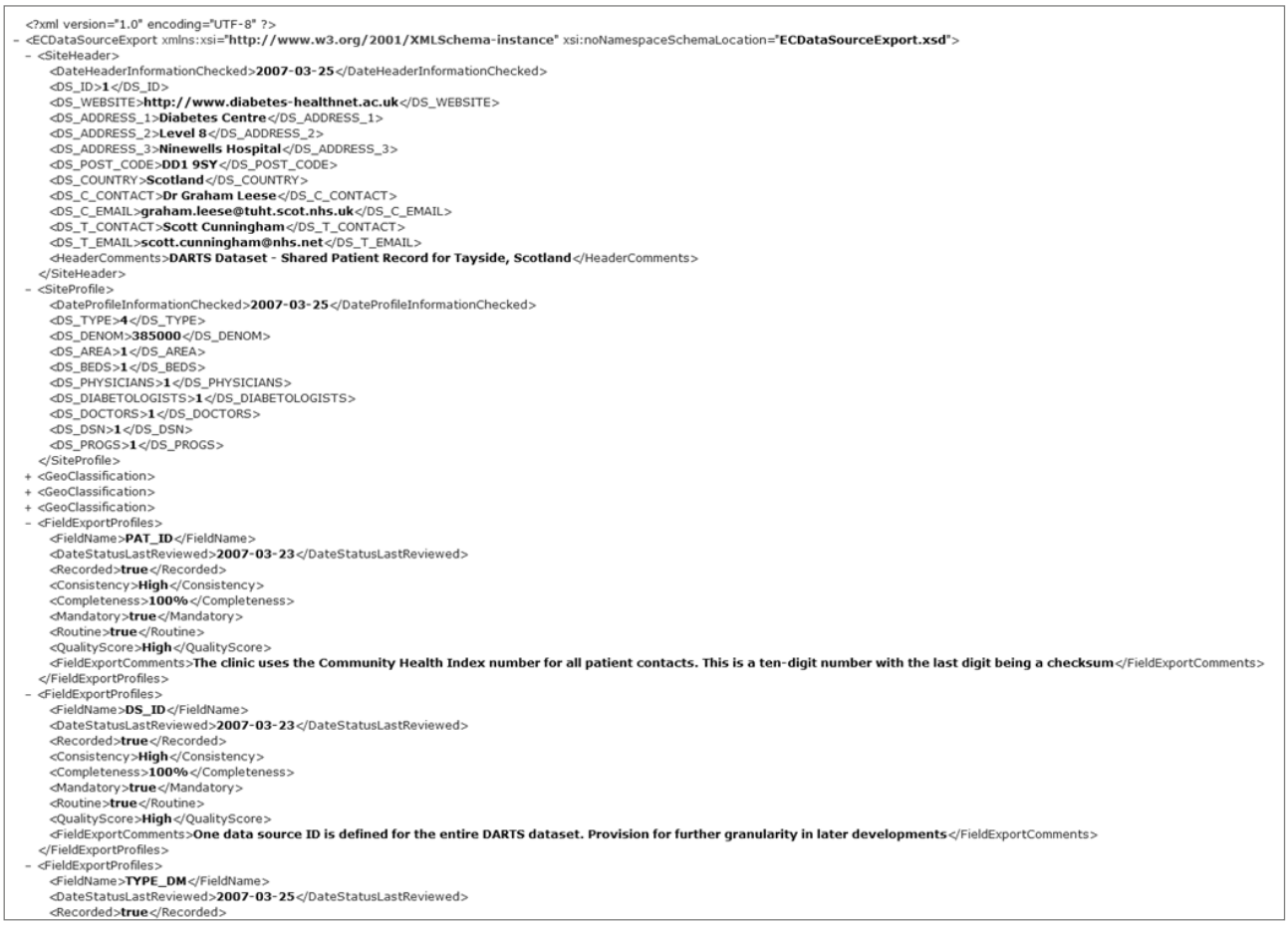


Figure 3.1.10: Summary of Data Entry

dataset, European Diabetes Data Definitions applicable using existing clinical datasets, have been documented and agreed.

To complement the dataset, a data dictionary and technical infrastructure have been developed in order to capture metadata from all contributing data sources. This allows local knowledge and expertise to be summarised and documented within a standardised XML format. The electronic storage of this data allows for future reference and presentation alongside final outputs.

A wider dataset and data dictionary review will commence as the project expands to 20 countries via the EUBIROD project. In addition to the review of new partners' datasets, this will also allow the existing BIRO datasets to be re-validated to maintain consistency and to take any recent changes into account.



```
<?xml version="1.0" encoding="UTF-8" ?>
- <ECDataSourceExport xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="ECDataSourceExport.xsd">
  - <SiteHeader>
      <DateHeaderInformationChecked>2007-03-25</DateHeaderInformationChecked>
      <DS_ID>1</DS_ID>
      <DS_WEBSITE>http://www.diabetes-healthnet.ac.uk</DS_WEBSITE>
      <DS_ADDRESS_1>Diabetes Centre</DS_ADDRESS_1>
      <DS_ADDRESS_2>Level 8</DS_ADDRESS_2>
      <DS_ADDRESS_3>Ninewells Hospital</DS_ADDRESS_3>
      <DS_POST_CODE>DD1 9SY</DS_POST_CODE>
      <DS_COUNTRY>Scotland</DS_COUNTRY>
      <DS_C_CONTACT>Dr Graham Leese</DS_C_CONTACT>
      <DS_C_EMAIL>graham.leese@tuht.scot.nhs.uk</DS_C_EMAIL>
      <DS_T_CONTACT>Scott Cunningham</DS_T_CONTACT>
      <DS_T_EMAIL>scott.cunningham@nhs.net</DS_T_EMAIL>
      <HeaderComments>DARTS Dataset - Shared Patient Record for Tayside, Scotland</HeaderComments>
    </SiteHeader>
  - <SiteProfile>
      <DateProfileInformationChecked>2007-03-25</DateProfileInformationChecked>
      <DS_TYPE>4</DS_TYPE>
      <DS_DENOM>385000</DS_DENOM>
      <DS_AREA>1</DS_AREA>
      <DS_BEDS>1</DS_BEDS>
      <DS_PHYSICIANS>1</DS_PHYSICIANS>
      <DS_DIABETOLOGISTS>1</DS_DIABETOLOGISTS>
      <DS_DOCTORS>1</DS_DOCTORS>
      <DS_DSN>1</DS_DSN>
      <DS_PROGS>1</DS_PROGS>
    </SiteProfile>
  + <GeoClassification>
  + <GeoClassification>
  + <GeoClassification>
  - <FieldExportProfiles>
      <FieldName>PAT_ID</FieldName>
      <DateStatusLastReviewed>2007-03-23</DateStatusLastReviewed>
      <Recorded>true</Recorded>
      <Consistency>High</Consistency>
      <Completeness>100%</Completeness>
      <Mandatory>true</Mandatory>
      <Routine>true</Routine>
      <QualityScore>High</QualityScore>
      <FieldExportComments>The clinic uses the Community Health Index number for all patient contacts. This is a ten-digit number with the last digit being a checksum</FieldExportComments>
    </FieldExportProfiles>
  - <FieldExportProfiles>
      <FieldName>DS_ID</FieldName>
      <DateStatusLastReviewed>2007-03-23</DateStatusLastReviewed>
      <Recorded>true</Recorded>
      <Consistency>High</Consistency>
      <Completeness>100%</Completeness>
      <Mandatory>true</Mandatory>
      <Routine>true</Routine>
      <QualityScore>High</QualityScore>
      <FieldExportComments>One data source ID is defined for the entire DARTS dataset. Provision for further granularity in later developments</FieldExportComments>
    </FieldExportProfiles>
  - <FieldExportProfiles>
      <FieldName>TYPE_DM</FieldName>
      <DateStatusLastReviewed>2007-03-25</DateStatusLastReviewed>
      <Recorded>true</Recorded>
```

Figure 3.1.11: Data Source Sample Extract

References

1. Morris AD, Boyle DI, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, MacDonald TM. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. BMJ, 1997, 315(7107):524–528.
2. Boyle DI, Cunningham S, Sullivan FM, Morris A (2001). Technology integration for the provision of population-based equitable patient care: The Tayside Regional Diabetes Network - a brief description Diabetes Nutrition & Metabolism 14():100 – 103
3. Boyle DIR, Cunningham SG. Resolving fundamental quality issues in linked datasets for clinical care. Health Informatics Journal 8():73 – 77, 2002
4. BIRO WP2 Clinical Review Indicator Development Results, available at: http://www.biro-project.eu/restricted/deliverables/D2_1%20Clinical%20Review.pdf
5. eXtensible Markup Language, available at: http://en.wikipedia.org/wiki/XML
6. DiabCare Q-NET: Diabetes Quality Network, available at: http://www.diabcare.de
7. Forum for Quality Systems in Diabetes Care, available at: https://apps.healthgate.at/bars/help/input/en/itemsdef.jsp
8. Scottish Executive Health Department. Scottish Diabetes Core Dataset. Scottish Executive, 2003.
9. Massi Benedetti M, Carinci F, Orsini Federici M. The Umbria Diabetes Register. Diabetes Res Clin Pract. 74(2):S200-4 2006
10. EUCID, available at: http://www.eucid.eu/
11. The Australian Diabetes, Obesity and Lifestyle (AusDiab) study, available at: http://www.diabetes.com.au/AusDiab2000datadictionary
12. Conversion of SI to Conventional Units: Blood Chemistries, available at: http://www.medal.org/visitor/www/Active/ch40/ch40.01/ch40.01.07.aspx
13. The International System of Units (SI), available at: http://www.bipm.org/en/si/
14. BIRO WP3 Common Dataset, available at: http://www.biro-project.eu/restricted/deliverables/D3_1%20Common%20Dataset.pdf
15. BIRO WP4 Data Dictionary, available at: http://www.biro-project.eu/restricted/deliverables/D4_1%20Data%20Dictionary%20v0_2.pdf
16. BIRO WP4 XML Metadata Dictionary, available at: http://www.biro-project.eu/restricted/deliverables/D4_2%20XML%20Metadata%20Dictionary%20v0_3.pdf
17. ESRI Shapefiles, available at: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf http://en.wikipedia.org/wiki/Shape_files
18. Nomenclature of Territorial Units for Statistics, available at: http://en.wikipedia.org/wiki/NUTS
19. Wikipedia Data Dictionary Definition, available at: http://en.wikipedia.org/wiki/Data_dictionary
20. Wikipedia Metadata Definition,available at: http://en.wikipedia.org/wiki/Metadata
21. BIRO WP7 Reports Template, available at: http://www.biro-project.eu/restricted/deliverables/D7_1%20Reports%20Template.pdf
22. Scottish Intercollegiate Guidelines Network: SIGN 25, available at: http://www.sign.ac.uk/
23. NCDDP (National Clinical Dataset Development Programme), available at: http://www.clinicaldatasets.scot.nhs.uk/ http://www.datadictionaryadmin.scot.nhs.uk/
24. NHS Clinical Datasets for England and Wales, available at: http://www.datadictionary.nhs.uk/
25. Australian Institute of Health and Welfare Metadata Online Registry, available at: http://meteor.aihw.gov.au/content/index.phtml/itemId/276533

Privacy Impact Assessment

Concetta Tania Di Iorio, Fabrizio Carinci

ABSTRACT

Introduction
The BIRO Information System involves the use of sensitive-medical data collected through diabetes registries within national boundaries and further processed for public health studies at the international level. Privacy impact assessment is a systematic and flexible process for evaluating a proposal/project in terms of its impact upon privacy, which has been specifically adapted to the BIRO context.

Objectives
To provide a definitive description of privacy risks, applicable privacy legislation and mitigation strategies adopted in the implementation and management of the BIRO Information System. To identify a general methodology supporting collaborative networks of regional disease registers and the routine evaluation of health information systems.

Materials and methods
A multidisciplinary team carried out a preliminary assessment through a systematic review of the privacy literature, followed by a general discussion on the data flow. Data flow analysis focused on alternatives identified in the first step. A Delphi consensus procedure defined the best alternative through the production of data flow tables (possible scenarios for the collection, use and disclosure of personal information/data, with related options) information flow questionnaire (marks for each scenario/option); overall consensus table (ranking all alternative architectures, scenarios and options). Privacy analysis covered issues arising in data transfer from local centres to the central database. Potential privacy risks have been identified and thoroughly analysed through a summary table indicating mitigation strategies to be implemented. The level of risk was classified according to an ordinal scale of intensity.

Results
Three main candidate architectures were identified: “individual patient data, de-identified through a pseudonym”; “aggregation by group of patients, with Centre’s identifiers available in de-identified form, securely encrypted”; and “Aggregation by Region”. Data flow analysis selected the second one as the best solution in terms of privacy protection, information content, scientific soundness and feasibility. Privacy analysis performed a detailed assessment of the various aspects involved in the adoption of the final BIRO architecture. The transfer of information occurring in the system, based upon the exchange of de-identified data and targeted mitigation strategies, identifies a low level of privacy risk.

Discussion
BIRO participating centres apply anonymisation procedures before any transfer to the BIRO central database, where aggregate records are processed solely for statistical and scientific purposes. According to Recital 26 of the EU Data Protection Directive, anonymisation allows personal data processing without consent, placing BIRO outside the scope of the data protection principles therein contained. The system processes only statistical objects stored as aggregate comma delimited files: there is no possibility, according to the state of the art, to identify a patient, either directly or indirectly, with a reasonable effort. Aggregate data processed by the local database engine are sent to the central statistical engine through “ad hoc” communication software ensuring secure information exchange and compliance with security requirements enshrined in EU and international data protection norms. Therefore, further processing by the global statistical engine cannot pose any privacy risk, either directly or indirectly. Trans-border data flow envisaged in BIRO is legally viable according to the EU legislation. Publication of project results is performed to avoid any direct/indirect identification of data subjects and/or local centres.

Conclusions
Privacy impact assessment shows that the selected BIRO architecture fulfils privacy protection requirements by addressing and resolving broad privacy concerns from different angles. The architecture of the system flexibly affords the best privacy protection in the construction of an efficient model for the continuous production of European diabetes reports. The methodology identified can be usefully applied in other fields of health information, particularly where disease registers are involved as primary units for data collection and statistical analysis.

3.2.1. Introduction

There is no unique definition of Privacy Impact Assessment (PIA) in the literature. It has been defined as a "process whereby a conscious and systematic effort is made to assess the privacy impacts of options that may be open in regard to a proposal. An alternative definition might be that a PIA is an assessment of any actual or potential effects that the activity or proposal may have on individual privacy and the ways in which any adverse effects may be mitigated."[1]

Moreover, PIA is usually conceived as a "protean document in the sense that it is likely to continue to evolve over time with the continued development of a particular system."[2]

Hence, there is a general consensus that a PIA is not just an end-product or a statement or practice. PIA is better conceived as a process rather than an outcome, which should be open-ended and regularised throughout the life-cycle of a programme / project.

With regard to different jurisdictions that have employed PIAs as structured means to assess privacy risks in government/private programs or projects, the following definitions are of utmost significance, since they highlight a bulk of common features: PIA has been defined as an "assessment of actual or potential effects on privacy, and how they can be mitigated" (Australia), "a systematic process for evaluating a proposal in terms of its impact upon privacy" (New Zealand), a "framework to ensure that privacy is considered throughout the design or re-design of a programme... [and to] identify the extent to which it complies with all appropriate statutes.
This is done to "mitigate privacy risks and promote fully informed policy" (Canada), an analysis of how information in identifiable form is collected, stored, protected, shared and managed…[to] ensure that system owners and developers have consciously incorporated privacy protection throughout the entire life cycle of a system (USA)[3].

According to the above definitions,
PIAs should be designed to:

- conduct a prospective identification of privacy issues or risks before systems and programmes are put in place, or modified
- assess the impacts in terms broader than those of legal compliance
- be process rather than output oriented
- be systematic

Legal compliance is, therefore, only one of the several criteria that need to be addressed in a larger process of risk assessment.

Those larger questions include the "moral and ethical issues posed by whatever is being proposed"[4].

Many projects might be technically compliant with the law, but they may raise significant concerns, even resistance, in certain societies or in public.

3.2.2. Materials and Methods

The adoption of PIA in B.I.R.O. seemed convenient and cost effective, as it allowed for privacy risks and concerns to be minimised by design. Ex-post adjustments were inherently excluded through the incorporation of mitigation strategies directly in the system design, whenever privacy risks could not be fully avoided.

A multidisciplinary dedicated "PIA Team" (PT) was formed, led by a facilitator (PF) expert in international privacy legislation, and including at least a representative from each partner institution.

The procedure involved four consecutive steps: preliminary privacy impact assessment, data flow analysis, privacy analysis and PIA report.

The preliminary part included a discussion on the data flow, focusing on the physical/logical separation of personal information/data. It involved a systematic review of the privacy literature, whose search strategy included use of Ovid Medline with criteria: {privacy AND [(registr* OR register) OR (health information system*) OR (health database*)]}, and limits [human AND English Language AND yr = 2001-2006].

A total of 64 biomedical and 11 law articles were identified after exclusion of papers more related to quality of care, privacy laws on research, genetic discrimination and patient recruitment strategies. A second search was performed on Law Journals using the same criteria.

A core set of fourteen papers was selected by comparing abstracts against main project objectives. Papers were reviewed by the PT to complete a comprehensive report of the first step and identify a short list of possible candidate architectures.
The second step involved a data flow analysis for each of the alternatives identified.

A Delphi Consensus Procedure was undertaken by the PT to define the best alternative by producing the following materials:

- data flow tables (DFT), including the possible scenarios for the collection, use and disclosure of personal information/data, with a number of possible options (Tables 3.2.1-3.2.3)
- an information flow questionnaire (IFQ), to assign marks to each scenario/option
- an overall consensus table (OCT), ranking scenarios/options (Table 3.2.4)

Materials were assembled using the procedure presented in Figure 3.2.1.

DFTs were initially prepared by the PF and revised by the whole PT. They were finally approved and used to compile the IFQ. The IFQ provided a series of scenarios, broken down into separate sub-options, for any of which marks were assigned on the basis of a set of three essential criteria: privacy, information

Table 3.2.1. CANDIDATE ARCHITECTURE 1: INDIVIDUAL PATIENT DATA

Description of personal information / data clusters	*Collected by*	*Type of format*	*Used by*	*Purpose of collection*	*Transmission to BIRO: de-identification*	*Security mechanisms for data transmission*	*Format of BIRO Database*	*Disclosed to*	*Storage and retention site*
SCENARIO 1: Health Service Medical Record[1]	Clinical Centres, Coordinating Centre[2]	OPTION 1 Longitudinal data collection OPTION 2 Multiple measurements averaged over time interval3	Local Health Authority, Coordinating Centre	Disease Management Program	Pseudonym used for data linkage[4], multiple measurements per patient OPTION 1. Centre IDs retained OPTION 2. Centre IDs de-identified[5]	OPTION 1. Password access for local administrator prompting client program to send encrypted bundles to BIRO[6] OPTION 2. Client program automatically sending encrypted data (agent)[7]	OPTION 1. Full information on all medical records OPTION 2. Averaged over time[8]	OPTION 1. BIRO database administrator OPTION 2. All local database administrators[9]	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)[10]
SCENARIO 2: Administrative Data Service Episode[11]	Local Health Authority[12]		Local Health Authority	Policy and Planning					
SCENARIO 3: Epidemiological measurement of multiple individual characteristics[13]	Research Organization[14]		Research Centre	Epidemiological Study					
SCENARIO 4.1: Health Service Medical Record + Administrative Data Service Episode	Population-based Regional/ National Diabetes Register[15]		Local Health Authority, Research Centre, Regional/National Government	Disease Management, Policy and Planning, Research					
SCENARIO 4.2: 4.1 + Epidemiological measurement of multiple individual characteristics									

Table 3.2.2. CANDIDATE ARCHITECTURE 2: AGGREGATION BY GROUP OF PATIENTS

Scenario 1: Grouping condition directly set by statistical object (e.g. ordered frequency distribution of LOS by CENTRE to compute variability of medians)[16]

Description of personal information / Data clusters	*Collected by*	*Type of format*	*Used by*	*Purpose of collection*	*Transmission to BIRO: de-identification*	*Security mechanisms for data transmission*	*Format of BIRO Database*	*Disclosed to*	*Storage or retention site*
NO aggregation size limit OR min aggregation N=5 patients per cell[17] OR min aggregation N=5, only applicable for high critical privacy variables e.g. service centre, geographical site etc[18]	BIRO partner	One Record for each aggregation level	BIRO partner (local engine), BIRO Consortium (central engine)	Computation of single BIRO statistical object for local and SEDIS reporting[19]	OPTION 1. All DATE fields transmitted as in original OPTION 2. DATE fields approximated to time interval (e.g. months)[20]	OPTION 1. Password access for local administrator prompting client program to send encrypted bundles to BIRO OPTION 2. Client program automatically sending encrypted data (agent)	Separate sets of aggregated tables linkable by predefined statistical criteria	OPTION 1. BIRO database administrator OPTION 2. All local database administrators[21]	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)[22]
Aggregation across service centres[23] OR data aggregated at the level of Service Centre									
Aggregation of Multidimensional patterns (e.g. risk adjustment) NOT allowed[24] OR generally allowed[25] OR allowed with min N=5 condition applied[26]									

Table 3.2.3. CANDIDATE ARCHITECTURE 3: AGGREGATION BY REGION

Scenario 1: Grouping condition directly set by statistical object (e.g. ordered frequency distribution of LOS by REGION)[1]

Description of personal information / Data clusters	*Collected by*	*Type of format*	*Used by*	*Purpose of collection*	*Transmission to BIRO: de-identification*	*Security mechanisms for data transmission*	*Format of BIRO Database*	*Disclosed to*	*Storage or retention site*
Aggregation without restrictions OR with restrictions applied on specific stratification criteria (e.g. geographical variable, centres etc) Geographical mapping available[5] OR Unavailable Variability of Centres' Outcomes Available[6] OR Unavailable Aggregation by multidimensional patterns (e.g. risk adjustment) NOT allowed OR allowed without restrictions applied on specific stratification criteria OR allowed with restrictions applied on specific stratification criteria[7]	BIRO partner	One Record for each aggregation level by REGION	BIRO partner (local engine), BIRO Consortium (central engine)	Computation of single BIRO statistical object for local and SEDIS reporting	OPTION 1. All DATE fields transmitted as in original OPTION 2. DATE fields approximated to time interval (e.g. months)[2]	OPTION 1. Password access for local administrator prompting client program to send encrypted bundles to BIRO OPTION 2. Client program automatically sending encrypted data (agent)	Separate sets of aggregated tables linkable by predefined statistical criteria	OPTION 1. BIRO database administrator OPTION 2. All local database administrators[3]	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)[4]

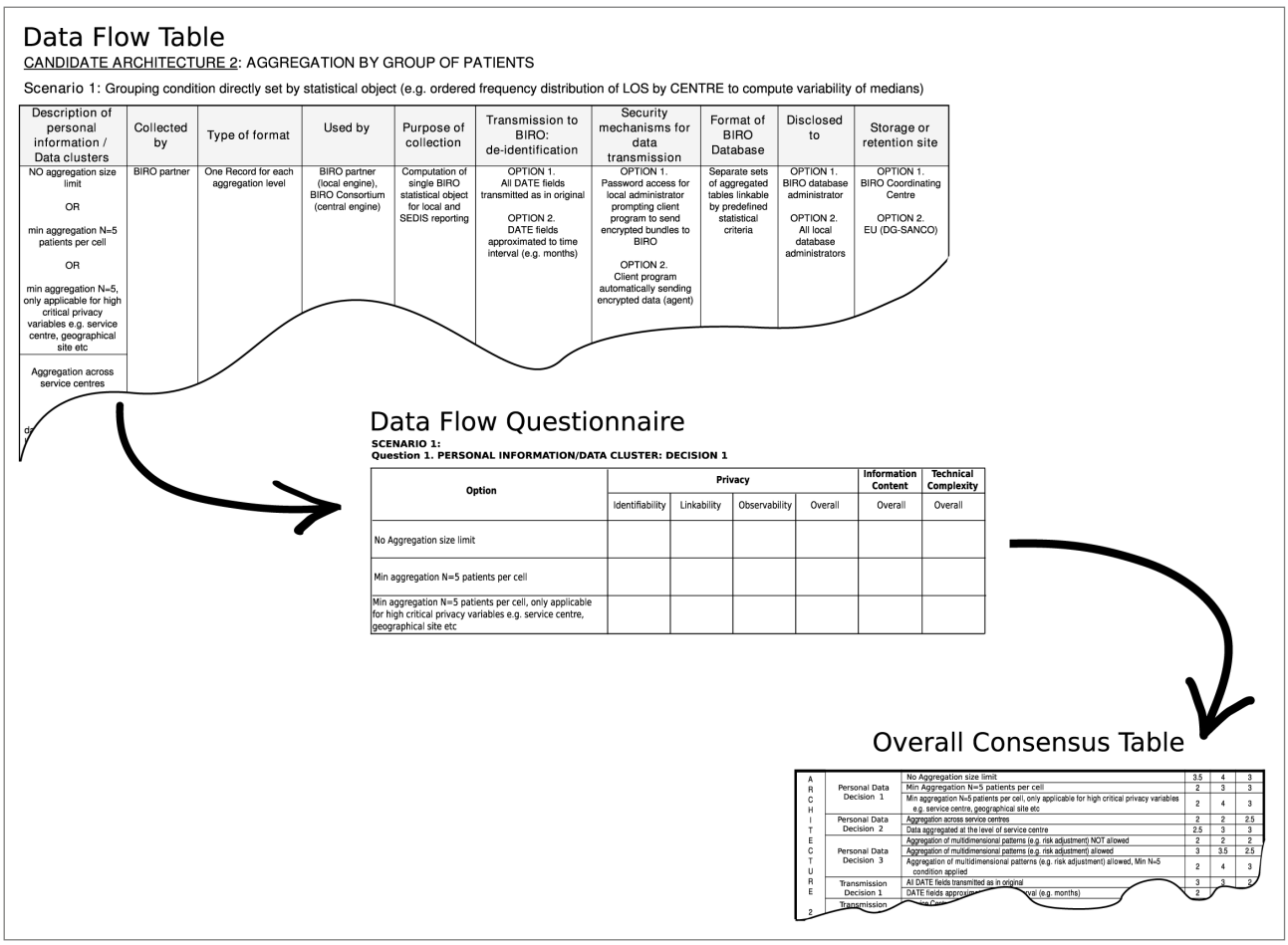


Figure 3.2.1: Method for the selection of the best architecture

Table 3.2.4 Overall Consensus Table

A.	Category	Option	P.	I.C.	T.C.
ARCHITECTURE 1	SCENARIO 1	One record for each service episode, centre IDs retained	5	5	3
		One record for each service episode, Centre IDs De-Identified	4	4	3
		Multiple measurements averaged over time interval, centre IDs retained	4.5	4	3
		Multiple measurements averaged over time interval, entre IDs De-Identified	4	3	3
	SCENARIO 2	Population-based longitudinal records, linked across administrative datasets, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained	5	5	3
		Population-based longitudinal records, linked across administrative datasets, Centre IDs De-Identified	4	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained	4	4	3
		Multiple measurements averaged over time interval, Centre IDs De-Identified	3	3	3
	SCENARIO 3	Longitudinal collection of clinical characteristics, Pseudonym used for data linkage, multiple measurements per patients	4	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients	3	3	3
	SCENARIO 4.1	Longitudinal data collection across relational data-warehouse, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	5	5	3
		Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified	4	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	4	4	3
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	4	3	3
	SCENARIO 4.2	Longitudinal data collection across relational data-warehouse, Pseudonym used for data linkage overmultiple datasets, all relational structure sent to BIRO	5	5	3
		Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified	4.5	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	4.5	4	3
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	4	4	3
ARCHITECTURE 2	Personal Data Decision 1	No Aggregation Size Limit	3.5	4	3
		Min aggregation N=5 patients per cell	2	3	3
		Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc	**2**	**4**	**3**
	Personal Data Decision 2	Aggregation across service centres	2	2	2.5
		Data aggregated at the level of service centre	**2.5**	**3**	**3**
	Personal Data Decision 3	Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed	2	2	2
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed	3	3.5	2.5
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed, Min N=5 condition applied	**2**	**4**	**3**
	Transmission Decision 1	All DATE fields transmitted as in original	3	3	2
		DATE fields approximated to time interval (e.g. months)	**2**	**3**	**2**
	Transmission Decision 2	Service Centre ID transmitted	3.5	3	2
		Pseudonym used for service centre	**2**	**2.5**	**2**
ARCHITECTURE 3	Personal Data Decision 1	NO restrictions on specific stratification criteria (e.g. geographical variable, centres, etc)	2	1	1
		Restrictions applied on specific stratification criteria (e.g. geographical variable, centres, etc)	1	1	2
	Personal Data Decision 2	Geographical mapping available	2	3	2
		Geographical mapping unavailable	1	1	1
	Personal Data Decision 3	Variability of centres outcomes available	2	3	3
		Variability of centres outcomes unavailable	1	1	1
	Personal Data Decision 4	Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed	1	1	1
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed WITHOUT restrictions applied on specific stratification criteria	3	3	2
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed WITH restrictions applied on specific stratification criteria	2	2	3
ALL	Security	**Password access for local administrator prompting client program to send encrypted bundles to BIRO**	**2**	**0**	**2**
		Client program automatically sending encrypted data (agent)	1	0	4
	Format	Full information on all medical records	4	5	3
		Averaged over time	**2**	**3**	**2**
	Disclosure	**BIRO database administrator**	**1**	**0**	**1**
		All local database administrators / registry managers	3	0	2
	Storage/Retention	**BIRO Coordinating Centre**	**2**	**0**	**2**
		EU/DG-SANCO	1	0*	3

content for diabetes, and technical complexity (feasibility).

Scores ranged from 0 (not applicable) to 5 (high level). The score on privacy was split into three separate criteria[5]:

- identifiability, a measure of how much the information available is personally identifiable on a continuum ranging from full anonymity (no name) to full verinymity (true name).
- linkability, a measure of the degree to which data elements can be used to reconstruct the true name of the subject.
- observability, a measure of the degree to which any other factor relative to data processing (time, location and data contents) can potentially affect identifiability and/or linkability (effect modifier).

An overall privacy score was assigned as an average of the three privacy dimensions, according to a scale of increasing "threat to privacy".

The score for the information content criterion was based on the information provided by the specific scenario/option in terms of relevance for diabetes, while the technical complexity score was based on the feasibility of the implementation of the specific scenario/option.

The overall mark for each option was based on the average of the three dimensions described above.

The IFQ was distributed to the PT and each member was asked to assign independent marks to each variable. The distribution, median and mean of scores were taken (with privacy scale reverted to higher privacy protection) and a final overall score assigned to each option.

All results were included in the OCT, presenting options ranked by overall scores, with ties ranked by increasing threat to privacy. The best architecture was defined as the mix of best options for all dimensions examined.

The final step involved an analysis of the selected architecture and the compilation of all materials/results into an overall report.

3.2.3. Results

The accomplishment of PIA tasks provided essential input for the development of all major components of the B.I.R.O. system.

Three main candidate architectures were identified, with differing levels of data sharing.

The first alternative required the transmission of "individual patient data, de-identified through a pseudonym", secured by an encryption algorithm and privacy protective communication technologies.

The second alternative envisaged data shared as "aggregation by group of patients, with Centre's IDs available in de-identified form, securely encrypted", transferred using privacy protective solutions.

The third alternative was based on "Aggregation by Region", optimised to impede reverse engineering, with the usual secure data transfer.

Details of the three alternatives were used to compile the DFTs and DFQ.

The Delphi panel selected the best alternative by ranking the three alternative scenarios, including options for their implementation.

The resulting B.I.R.O. system architecture is shown in Figure 3.2.2, whose criteria were duly taken into account for implementation.

Statistical properties (e.g. those of the arithmetic mean, percentiles, etc.) were exploited to transmit target objects in separate bundles over the network, so that international reports avoid many potential risks and restrictions imposed by privacy legislation, with no exchange of individual records.

Specialized communication software has been developed to securely transmit statistical objects as encrypted compressed folders containing comma-delimited text files (.csv).

Security has been addressed comprehensively according to ISO/OSI 7498-2. For authentication, digital certificates trusted by a common certification authority were exchanged and installed in sender and receiver. Access control was configuredsuch that only trusted identities were authorized to connect to services. Security was also provided by using encryption, and data integrity as well as non-repudiation were provided by digital signatures.

Web services were selected as the core technology for communication for their compliance with standards set by the open World Wide Web consortium: SOAP (Simple Object Access Protocol) for messaging, HTTP (Hypertext Transfer Protocol) for Internet transport and XML (eXtensible Markup Language) together with its security extensions XMLenc (encryption) and XMLsig (digital signatures). Apache Axis 2, together with Apache Rampart provided by Java 2 Enterprise Edition, were chosen for pilot development and configuration of sending and receiving applications.

Encryption and digital signatures were applied on two layers. Firstly, transport layer security using HTTPS, i.e. HTTP protocol together with SSL (Secure Sockets Layer), was used to protect the entire data stream exchanged between sender and receiver. Secondly, on the data layer, individual chunks of data were encrypted and digitally signed, giving the application full control over further utilization, storage and processing of digital signatures and other security related information.

The whole B.I.R.O. process is controlled by integrated software linking the different modules through a simple

graphical user interface (GUI). A ”local” module is used to allow users exporting local data to XML files, to add them to a local database, and to produce local reports and statistical objects for the central B.I.R.O. System. A “central” module is used by the server administrator to load statistical objects received from partial analyses in the form of csv files, and to run the overall analysis for the global B.I.R.O. report.

The B.I.R.O. architecture requires for the Central Engine to be managed by a unique administrator, ensuring compliance with all national and international security rules in the maintenance of the server, as specified in the Preliminary PIA Report[6].

Results are stored in a server database that will be connected to a web portal in charge of delivering online results to the masses, bundled with proper data definitions and methodological references.

3.2.4 Discussion

3.2.4.1 Privacy analysis: legislative framework

Of all the human rights in the international catalogue, the right to privacy is perhaps the most difficult to define[7].

Definitions of privacy vary widely according to contexts and environments. Nevertheless, privacy is usually seen as the way of drawing the line of how far a society can intrude into a person’s private life.

Privacy has been defined as the “right to be left alone”[8]; or as “the right of the individual to be protected against intrusion into his personal life or affairs, or those of his family, by direct physical means or by publication of information[9].

Although there is a lack of a single definition of privacy, it is a right generally recognized around the world and crystallised in many international instruments.

The 1948 Universal Declaration of Human Rights was the first international binding instrument to recognise privacy as a human right, specifically protecting territorial and communication’s privacy[10]. Article 12 states: “No one should be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks on his honour or reputation. Everyone has the right to the protection of the law against such interferences or attacks”.

In addition, numerous international human rights treaties specifically recognize privacy as a right. The International Covenant on Civil and Political Rights (ICCPR - art. 17)[11]; the UN Convention on Migrant Workers (Article 14)[12], and the UN Convention on Protection of the Child (Article 16)[13] adopt the same language. On the regional level, various treaties make these rights legally enforceable.

For instance, Article 8 of the European Convention for

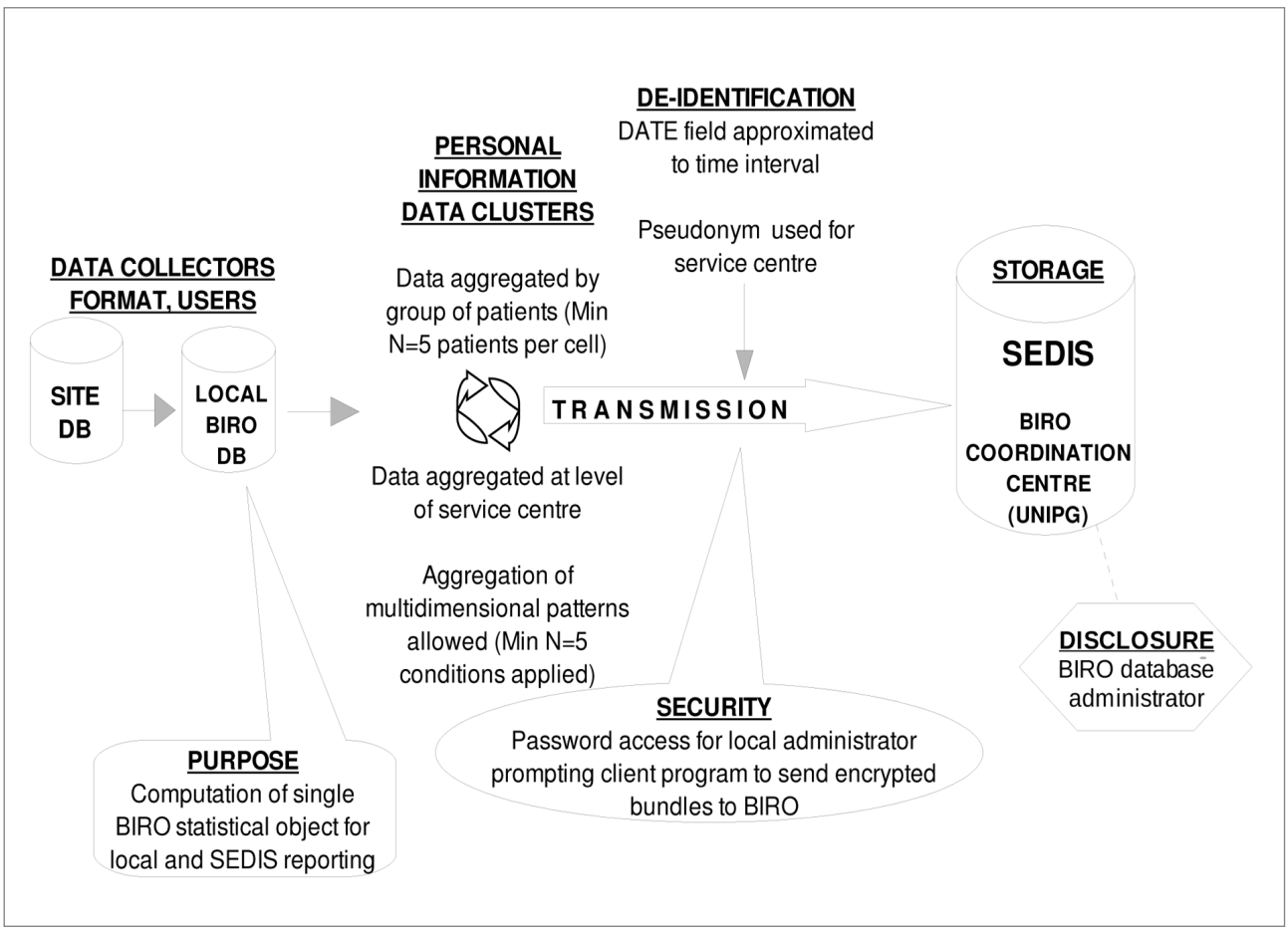


Figure 3.2.2: BIRO Data Architecture

Table 3.2.5. Privacy Contingency Risks

Element	Nature of risks	Level of risks			Comments	Mitigating Mechanisms
		Low	Medium	High		
Individual data: Pseudonym used for patients' IDs + Data is Aggregated (N=5 patient per cell)	Individual privacy	X			Pose an indirect risk to individual's privacy	Non-Reversible De-identification
Pseudonym used for Centres IDs	Non-Individual Privacy		X		Pose an indirect risk to Centres' privacy	Reversible De-identification + Reporting System: percentage
Data Transmission	Security Measures	X			Pose an indirect risk to individual's privacy	Encryption
Access to the BIRO network	Security Measures		X		Pose an indirect risk to individual's privacy	Secure applications Hacking tests
Global Statistical Analysis	Individual privacy + Non-Individual Privacy + Security Measures	X			Pose an indirect risk to individual's privacy and centres privacy	Non-reversible de-identification + Encryption

the Protection of Human Rights and Fundamental

Freedoms (1950)[14] states that "Everyone has the right to respect for his private and family life, his home and his correspondence. There shall be no interference by a public authority with the exercise of this right except as in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health of morals, or for the protection of the rights and freedoms of others".

The Convention created the European Commission of Human Rights and the European Court of Human Rights to oversee enforcement. Both have been active in the enforcement of privacy rights, and have consistently viewed Article 8's protections expansively and interpreted the restrictions narrowly[15].

The Court has reviewed Member States' laws and imposed sanctions on numerous countries[16]; and has also reviewed cases of individuals' access to their personal information in government files to ensure that adequate procedures exist[17]. In the evolution of data protection, the interest in the right of privacy increased in the 1960s and 1970s with the advent of information technology.
The surveillance potential of powerful computer systems has increased the demand for specific rules governing the collection and handling of personal information.

Two crucial international instruments in the evolution of data protection are the Council of Europe's (1981) Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data[18], and the Organization for Economic Cooperation and Development's (OECD) Guidelines Governing the Protection of Privacy and Transborder Data Flows of Personal Data[19], which set out specific rules covering the handling of electronic data.

These rules describe personal information as data that have accorded protection at every step: from collection to storage and dissemination.

As a matter of fact, the above-mentioned agreements have had a profound effect on the enactment of laws around the world. Nearly thirty countries have signed the COE Convention; and the OECD guidelines have been widely used in national legislations, even outside the OECD member countries. The development of

privacy protection in the EU took a step forward with the Council of Europe Convention on Human rights and Biomedicine (Oviedo 1997), which reinforced the principles that everyone is entitled to the right to privacy and confidentiality of personal medical data and the right to be informed about his/her health[20].

Finally, the Charter of Fundamental rights of the European Union (2000/C 364/01)21 specifically provides protection of personal data. Art 8 states: "Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. Compliance with these rules shall be subject to control by an independent authority".

The Charter of Fundamental Rights has been fully incorporated in the European Constitution (forming its part II)[22], signed in Rome on the 29th of October 2004. Although the Parliament, the Council and the Commission solemnly proclaimed the Charter on the 8th of December 2000, the Charter was not part of the Union's Treaties and therefore it had no binding legal force.

The Constitution thus achieved a major breakthrough, which allows the Union to have its own catalogue of rights, binding for all European countries and enforceable through the Court of Justice, which will in fact ensure that the Charter will be adhered to.

It is worth noting that the content of the Charter is broader than that of the European Convention for the

Protection of Human Rights and Fundamental Freedoms (ECHR), signed in Rome on 4 November 1950 and ratified by all the Member States of the Union.

Whereas the ECHR is limited to civil and political rights, the Charter of Fundamental Rights covers other areas such as the right to good administration, the social rights of workers, the protection of personal data and bioethics.

Finally, The Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research (2005)[23] further reinforced the duty of confidentiality in the handling of personal information in health research and reaffirmed the obligation to treat them according to the rules relating to the protection of private life.

In line with all the aforementioned instruments, the EU has adopted a privacy model that embraces comprehensive laws. The model is based on a general and abstract law that governs all aspects of the handling of personal information: from collection to use and dissemination, by both the public and private sectors.

The 1995 Data Protection Directive (95/46/EC)[24] sets up a common level of privacy among European countries, ensuring compliance through the establishment of a regulatory body.

The Directive not only reinforced current data protection laws, but also established a range of new rights and basic principles, namely: the right to know where the data originated, the right to have inaccurate data rectified, a right of recourse in the event of unlawful processing, and the right to withhold permission to use data in some circumstances.

The Directive contains strengthened protections over the use of sensitive data.

Art.7 of the Directive establishes a set of criteria of "legitimate processing". Processing, in order to be legitimate, has to take place: either with the unambiguous consent of the data subject, or where this is necessary for the performance of a contract with the data subject, for compliance with a legal obligation, or for the performance of a government task, just to mention a few examples.

More stringent conditions apply to the processing of special categories of sensitive data, such as medical data. Here, the processing of sensitive data is considered, in principle, not legitimate and member states has to prohibit their processing, unless special conditions verify.

According to art. 8, the processing of sensitive data is allowed when:

the data subject has given his explicit consent to the processing of those data, or processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law in so far as it is authorized by national law providing for adequate safeguards; or processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects; or the processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims.

Importantly, the prohibition of Article 8(1) shall, according to Article 8(3), also not apply where the data are required: for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional subject under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy.

Moreover, Member States may, according to Article 8 (4), for reasons of substantial public interest, lay down exemptions, in addition to those laid down, either by national law or by decision of the supervisory authority. Article 8(3) is extremely important for the health sector, since justifies the collection, use, and processing of health data, for the specified purposes, without the patient's consent.

Although the free and informed consent will be necessary if, for instance, those data would be further used for research purposes or any other secondary use. The reference to professional secrecy contained in Article 8(3) is crucial for obtaining a more effective protection of privacy in the handling of sensitive health data.

Although the issues surrounding the confidentiality of health data are not fully dealt with in the Directive, the referral to the obligation of confidentiality in the Directive represents a step forward towards an eventual harmonization of European legislation. At least, it imposes to Member States, in a binding form, the duty of confidentiality to any person involved in the processing of personal sensitive data, such as health data.

The duty of confidentiality has its origins in the duty of professional secrecy incumbent on health professionals either through a law or code of conduct. The principle of confidentiality of medical information, derived by the Hippocratic Oath, can be considered one of the oldest principles applying to data protection. Although privacy and confidentiality are conceptually distinct, they are strictly interrelated and need to be consistently implemented among European countries in order to enhance the protection of privacy when sensitive data are involved: as a matter of fact, confidentiality could rather be conceived as a means to protect the right to privacy.

In order to conduct scientific research without falling under the binding rules of the Directive, data should be rendered anonymous. Recital 26 of the EU Data protection Directive in fact states that "principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable".

Recital 26 thus places outside the scope of the Directive the discipline of data processed for research purposes when both direct and indirect identification is avoided. Direct identification should be interpreted as identification from the data itself and indirect identification as identification from the data itself matched with any other data or means that are reasonably likely to be used, such as an identification number or to one or more factors specific to the subject's physical, physiological, mental, economic, cultural or social identity[25]. For instance, coded and encrypted data are not considered anonymous "per se". If decoding or de-encrypt techniques are still possible without an unreasonable effort. In this circumstance, data shall be still subjected to the Directive rules[26].

Importantly, the 1995 Directive imposes an obligation on member states to ensure that the personal information relating to European citizens has the same level of protection when it is exported to, and processed in, countries outside the EU. As a result, countries refusing to adopt adequate privacy protections may find themselves unable to conduct certain types of information flows with Europe, particularly if they involve sensitive data.

In line with the EU Data Protection Directive, the Council of Europe enacted, in 1997, a Recommendation on the Protection of Medical Data: Council of Europe Recommendation No. R (97) 52[7]. The recommendation acknowledges that medical data requires even more protection than other non-sensitive personal data, reaffirming that the respect of rights and fundamental freedoms, and in particular of the right to privacy has to be guaranteed during the collection and processing of medical data.

For those reasons, Principle 3.2 recalls the requirement in Article 6 of the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981) for appropriate safeguards in the law, in so far as the various stages of collection and processing of medical data are concerned.

According to the Recommendation, the processing of medical data is, in principle, prohibited, unless appropriate safeguards are provided by domestic law.

One of such safeguards is that only health-care professionals, bound by rules of confidentiality, should collect and process medical data, or where necessary persons acting on behalf of health-care professionals, as long as such persons are subject to the same rules.

Since the definition of health professional may vary across different countries, the recommendation provides for the possibility that personnel not directly responsible for health care may collect and process medical data; but only on the condition that this category of professionals must abide by confidentiality rules comparable with those imposed on health-care professionals, or that domestic law provides for appropriate safeguards which are as efficient as confidentiality rules, that is, they are efficient enough to guarantee respect of privacy of the data subject. Trough this Recommendation, the duty of confidentiality has been in fact strengthened within European countries.

Once again, with a view to the sensitive nature of medical data, Principle 4.1 recalls the provisions in Article 5 of the Convention: the collection and processing of medical data must be fair and lawful, and for specific purposes only.

The principle of fair collection is made more explicit in Principle 4.2: medical data must, in normal conditions, be obtained from the data subject himself/herself.

This principle therefore concerns the "disclosure" of these data by the data subject himself/herself, and not "communication" of medical data by a third party (for example, the doctor).

Principle 4.3 lays down the rules governing the collection or processing of medical data. The latter may be collected or processed: if it is provided for by law, there is a contractual obligation to do so, if this is necessary for the establishment of a legal claim or if the data subject has given his/her consent. Principle 4.3 does not constitute a derogation from Principle 3.2, but sets conditions for the legitimacy of the collection or processing.

Medical data may also be collected from the data subject or from other sources if this is provided for by the law for one of the purposes set out in Principle 4.3(a): for public health reasons, the prevention of a real danger or the suppression of a specific criminal offence, or another important public interest.

Furthermore, medical data may be collected and processed if permitted by law for the purposes set out in Principle 4.3 (b): for preventive medical purposes or for diagnostic or therapeutic purposes (in this case data may also be processed for the management of medical service operating in the interest of the patient), or to safeguard the vital interests of a data subject, or with a view to respecting specific contractual obligations, or with a view to the establishment, exercise or defence of a legal claim. Thus, Principle 4.3 (b) reaffirms the rules set forth in the EU Data Protection Directive.

In accordance with principle 4.3 (c), medical data may also be collected and processed if the data subject has given his/her consent for one or more purposes in so far as domestic law does not provide otherwise.

Medical data may therefore be collected without consent, if the law provides for this, "for the purposes of" (that is, in the interest of) public health; this purpose is in line with the derogation for reasons of public safety in Article 9 of the Convention.

It should also be noted that the words "in the interest of public health" include the management of health services.

One of the means to ensure that medical data are obtained and processed fairly and lawfully is to inform the data subject, whose data are collected, of a number of elements (information to be given to the data subject). These elements are listed in Principle 5.1.

The provision of information is indispensable when the data subject is required to give his/her "informed" consent. But even in cases where his/her consent is not required - that is, when the collection and processing of medical data follow an obligation under the law or under a contract, are provided for or authorised by law, or when the consent requirement is dispensed with-the recommendation provides that the data subject is entitled to relevant information.

Although Principle 5.1 should be interpreted strictly, two kinds of derogation are admitted.

First of all, Principle 5.6 allows for derogations to be made for certain reasons of public interest, for protection of the data subject or a third person, or in medical emergencies.

Secondly, information on the various elements listed in the principle has to be supplied only in so far as it is relevant.

Principle 5.1 identifies the following elements on which the data subject must be informed:

- the existence of a file containing his/her medical data and the type of data collected or to be collected;
- the purpose or purposes for which they are or will be processed;
- where applicable, the individuals or bodies from whom they are or will be collected;
- the persons or bodies to whom and the purposes for which they may be communicated;
- the possibility;
- if any, for the data subject to refuse his consent, to withdraw it and the consequences of such withdrawal;
- the identity of the controller and of his/her representative, if any, as well as the conditions under which the rights of access and of rectification may be exercised.

One of the conditions on which medical data may be collected and processed is that the data subject has given his/her consent, in so far as he/she is capable of doing so. As these data are regarded as sensitive data, Principle 6.1 requires that the consent be "free, express and informed".

Consent is "informed" if the data subject is informed in particular of the purposes involved and the identity of the data controller. Consent is "free" if the data subject has the possibility to refuse his/her consent, to withdraw it or to modify the terms and conditions of consent. Consent can be expressed orally or in writings.

However, under certain conditions, medical data could be processed without the data subject's free, express and informed consent. These conditions are listed exhaustively in the recommendation.

As regards the collection of medical data in the course of a consultation or treatment for preventive, diagnostic or therapeutic purposes by a doctor, and which the data subject has freely chosen, the consent of the patient may not need to be expressed if the data were indeed to be processed only for the provision of care to the patient. This is also valid for processing medical data in the context of the management of a medical service operating in his/her interest.

The recommendation reaffirms the right of access: every person has to be enabled to have access to his/her medical data, either directly or through a health-care professional. Importantly, Article 8(1) of the recommendation states that the information must be provided to patients “in understandable form”. Access to medical data may be refused, limited or delayed only if the law provides for this.

The data subject has also the right to rectification:

patients may ask for rectification of erroneous data concerning him/her and, in case of refusal, he/she has to be able to appeal.
In general, medical data shall be kept no longer than necessary to achieve the purpose for which they were collected and processed (conservation).

Although the recommendation does not refer to it explicitly, the requirement in Article 5 of the Convention that personal data undergoing automatic processing should be adequate, relevant and not excessive applies equally to medical research. It means that only the data necessary for the purposes of such research should be used.

The primary means of protecting medical data to be used for scientific research purposes, is to make them anonymous. For this reason, researchers as well as public authorities concerned are urged to develop anonymisation techniques, which should be continuously updated and kept efficient.

The nature or objectives of certain research projects sometimes make it impossible to use anonymous data. In such cases, under Principle 12.2, personal data may be used if the purposes of the research project are legitimate and one of the listed conditions is fulfilled.

Firstly, personal data may be used for medical research if the data subject has been duly informed of the research project - or at least if the information requirements have been respected - and has given his/her consent for that particular project, or, at least, for the purposes of medical research.

Secondly, in the case of a legally incapacitated person, this consent must have been given in accordance with Principle 6.4, and the research project must have a connection with the medical condition or disease of the data subject (sub-paragraph b). This is provided to avoid that consent given on behalf of a legally incapacitated person might be motivated by material interests.

Thirdly, cases may arise where the data subject cannot be found or where for other reasons it is apparently impossible to obtain consent from the data subject himself/herself (for example, in the case of an epidemic). When in such cases the interests of the research project are such that they justify the consent requirement to be waived - for example in the case of an important public interest - and unless the data subject has explicitly refused any disclosure, then the authorization to use personal data may be given by the body or bodies designated by domestic law and competent in the area of personal data.

Such authorization should, however, not be given globally, but case-by-case; moreover, the medical data should be used only for the medical research project defined by that body, and not for another project of the same nature (sub-paragraph c).

The authorization, by the designated body, of communication of medical data for the purposes of a medical research project also depends on other factors implicit in the spirit of the recommendation in the present principle, or explicitly set out in other principles:
- the existence of alternative methods for the research envisaged;
- the relevance of an important public interest of the aim of the research, for example in the field of epidemiology, of drug control or of the clinical evaluation of medicines;
- the security measures envisaged to protect privacy;
- the necessity of interfering in the privacy of the data subject.

Under sub-paragraph (c), it would not be necessary to make the reasonable efforts in all cases; the person in charge must, however, consider whether with reasonable efforts it would be practicable to contact all data subjects. If this seems possible, then the efforts must be made. Furthermore, it was understood that to seek the consent of the data subject for medical research would be an unreasonable demand for the research institute, and would rather be the responsibility of the person or body envisaging disclosure of medical data.

According to Article 12(3), subject to complementary provisions determined by domestic law, health-care professionals entitled to carry out their own medical research are allowed to use the medical data which they hold, as long as the data subject has been informed of this possibility and has not objected.

Finally, personal data used for scientific research must not be published in a form that enables the data subjects to be identified, unless they have given their consent for the publication and publication is permitted by domestic law.

Some considerations could be made on the basis of the above legislation and regulations.

In all EU and International legislative Instruments, the right to privacy is not considered an absolute right. It has in fact to be weighed against other matters that provide benefit to the society. All the exemption to the prohibition of processing operations that involve personal data relative to health care and health research constitute clear examples of the non-absolute nature of the right to privacy.

Therefore, the protection of privacy is conceived as value that should not unnecessarily jeopardize health research.

The interest of societies in enhancing the health of populations is in fact strictly related to the possibility of conducting appropriate research in the health sector and the availability of personal data is fundamental for this purpose.

Considering that privacy protection and health research might conflict on the increasing demand of researchers to access data in identifiable form, appropriate methodologies and techniques should be implemented. PIAs are a valuable means to address this issue,

providing a balanced trade-off between privacy protection and the efficient and effective conduction of research projects and programs.

3.2.4.2 Privacy analysis: privacy protection in the context of BIRO engineering

The BIRO Information System involves the use of sensitive-medical data collected through diabetes registries within national boundaries and further processed for public health studies at international level. It has to be noted that the collection of data takes place at national level; and the investigation of privacy compliance of registries is out of the scope of the present assessment.

The privacy analysis covers any privacy issue that might arise in the transfer of data from the BIRO Centres to the central database, hosted by the University of Perugia, Italy.

At a general level, the kind of processing that takes place in the BIRO centres should be subject to Article 8 (3) of the Data Protection Directive[28].

Each centre collects, in fact, information relating to an identified or identifiable natural person for the purpose of setting up diabetes registries. Hence, it can be asserted that those data are collected and processed for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health care services. According to the EU Data Protection Directive, consent from the data subject may not be required in this case.

The norm constitutes an exemption to the prohibition of processing sensitive data, which is set forth by Article 8(3) of the Directive.

In this case, the exemption is justified by the need to protect the competing interests of society to a better health care. However, domestic laws may provide more stringent rules.

The further processing of those data, other than the care of the patients and the management of health care services, would instead not be covered by Article 8 (3) exemption. Hence, consent should be required for any secondary use of those data.

However, each centre of the BIRO consortium provides for the anonymisation of data before transferring them to the BIRO central database, where data are processed for statistical and scientific purposes.

The way data are rendered anonymous is central to determine if true anonymisation is actually envisaged in the BIRO System.

Ex Recital 26 of the Data Protection Directive, anonymisation allows the processing of personal data without consent, placing anonymous data outside the scope of the data protection principles therein contained. Anonymisation could be therefore seen as a means to determine the boundaries of privacy protection principles. When data is truly anonymous, the interest of the data subject to maintain his/her data private and confidential is in fact protected “ipso iure”; hence, the processing should be considered legitimate.

Data is rendered anonymous, according to Directive, only if “the data subject is no longer identifiable”. The Directive specifies that an “identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity”. The same Recital specifies that, in order to determine whether a person is identifiable, “account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person”.

Consequently, when the data subject could be identified with reasonable means directly from the data itself or indirectly through the combination of other means, data cannot be considered anonymous and, therefore, fall under the Directive principles, including the need to gain expressed consent from the data subject.

Data could be instead considered anonymous when “it would be reasonably impossible for the researcher or any other person to re-identify the data”.

The identification of the data subject through “reasonable means” is a vague concept that involves a value judgement. However, the reference to the state of the art in decoding and/or other similar techniques is usually considered decisive in valuing whether data is truly anonymous or not. For instance, coded and encrypted data are to be considered anonymous for the purpose of the EU Data Protection Directive if data cannot be decoded and de-encrypted with a reasonable effort.

In the context of BIRO, the local centres will use pseudonyms for patients IDs and data will be then stripped of their identifier and aggregated: at least n. 5 patients per cell are to be used. As a matter of fact, the BIRO System processes statistical objects, which basically are tables that contain statistical aggregations of local data (arithmetic mean, percentile, variance, linear and logistic regression, bar plot data, histogram data, box pot data, etc), stored as flat text comma delimited files (CSV).

Hence, there is no possibility, according to the state of the art, to identify, either directly or indirectly, a patient through a reasonable effort.

Although the privacy of legal persons, such as the BIRO Centres, does not receive protection within EU and International legislation, the PIA Team acknowledged that the availability of Centres’ IDs could pose broader privacy concerns.

Project’s results could reveal information about participating Centres that might jeopardize their reputation. Hence, this factor could not positively impact on data sharing and eventually discourage participation in the project.

Moreover, when dealing with very small Centres, even doctors or patients could be indirectly identified, if specific information is disclosed together with Centres' IDs. In consideration of the above concerns, Centres' IDs have been protected through the use of a pseudonym, together with to a reporting system based on percentages rather than on absolute numbers. Accordingly, the size of single Centres would be hidden, avoiding their indirect identification by third parties.

Although personal data is rendered truly anonymous and there is no need to justify the processing of those data without obtaining patients' consent, the further processing of personal data for statistical or scientific research purposes is generally considered, even within the EU Directive, compatible with the purposes for which the data have previously being collected. This principle is expressed, among the others, in the provision of Article 11 (2) of the EU Directive.

While Articles 10 and 11 impose the data controller, as a general rule, to give some kind of information to the data subject (for instance: the right to know the identity of the controller, the purpose of the processing and any further information), Paragraph 2 of Article 11 exempts the data controller from providing such information when the processing is performed for statistical or scientific research purposes, if the provision of such information proves impossible or would involve a disproportionate effort.

The case of BIRO would fall within the scope of the latter case. Considering its very large sample size, the effort to provide information to patients should herein be easily considered disproportionate. Consequently, the information to be provided to data subject could be waived by the single centres, unless domestic law provided differently, even if the kind of processing would be considered as falling under the EU Data Protection Directive rules.

The exemptions provided by the Directive are also in line with the principles contained in the Convention on the Protection of Individuals with regard to Automatic Processing of Personal Data (1981), which envisages the possibility of restricting the exercise of the data subject's rights with regard to data processing operations which pose no risk (Article 9, par. 3). Examples of no or minimal risk operations are, in particular, the use of data for statistical work, in so far as those data is presented in aggregate form and stripped of their identifiers, as in the case of BIRO. Similarly, scientific research is included in this category.

The aggregated data, in the form of statistical objects, once processed through the local database engine, are to be sent to the central statistical engine, which will perform global analysis.

A communication software has been developed to ensuring a secure data and information exchange transmission between the regional information systems and the central SEDIS.

To facilitate secure data transmission in the BIRO infrastructure, an applicable technology has been selected and successfully used in a pilot implementation. This is a foundation for further integration in data exchange workflows required in the shared European diabetes information system, as fully explained in paragraph 2.

Considering the security mechanisms implemented in the BIRO system, it can be asserted that the security requirements enshrined in EU and international data protection norms and regulations are fully fulfilled, considering the actual state of the art.

According to the BIRO data flow and architecture, statistical analysis will be then performed at global level. Considering that data have been rendered anonymous by local BIRO centres and transmitted to SEDIS in a secure environment, the further processing performed by the global statistical engine cannot pose any privacy risk either directly or indirectly.

The last issue that could be considered in the privacy analysis of the BIRO project is relative to the transborder data flow. In fact, data is to be sent to a central database, which is located outside the single national boundaries, except for the Italian partner (Coordinator).

The BIRO System, as already demonstrated, processes only anonymous data; therefore, privacy rules should not bound its implementation.

Nevertheless, the free flow of information, regardless of frontiers, is also a principle enshrined in Article 10 of the European Human Rights Convention. Accordingly, art 12 of the Convention on the Protection of Individuals with regard to Automatic Processing of Personal Data (1981)[29] and Article 25 of the EU Data Protection Directive (1995) discipline the transfer of data from one country to another.

The main rule contained in Article 12 (2) of the Convention, is that, in principle, obstacles to transborder data follows are not permitted between Contracting States in the form of prohibitions or special authorisations of data transfers. The rationale for this provision is that all Contracting States, having subscribed to the common core of data protection provisions set out in Chapter II, offer a certain minimum level of privacy protection.

In addiction, Article 12 (2) states that prohibiting or subjecting to special authorizations transborder flows of personal data is allowed only "for the sole purpose of the protection of privacy". The norm adds an important clarification, namely that a Contracting State may not invoke this convention to justify interference with transborder data flows for reasons which have nothing to do with the protection of privacy.

However, paragraph 2 of this article does not affect the possibility for a Party to lay down in its domestic data protection law provisions that, in particular cases, do not permit certain transfers of personal data, irrespective of whether such transfers take place within its territory or across the borders.

The Council of Europe Recommendation on the Protection of Medical Data[30], resembles the Convention and establishes that the transborder flow of medical data to a state which has ratified the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, and which disposes of legislation which provides at least equivalent protection of medical data, should not be subjected to special conditions concerning the protection of privacy.

Where the protection of medical data can be considered to be in line with the principle of equivalent protection laid down in the Convention, no restriction should be placed on the trans-border flow of medical data to a state which has not ratified the convention, but which has legal provisions which ensure protection in accordance with the principles of that convention and the recommendation.

Unless otherwise provided for by domestic law, the transborder flow of medical data to a state which does not ensure protection in accordance with the convention and with this recommendation, should not as a rule occur, unless necessary measures, including those of a contractual nature, to respect the principles of the convention and this recommendation, have been taken, and the data subject has the possibility to object to the transfer; or the data subject has given his consent.

According to the EU Directive, the cross border flow of personal data is allowed only when an adequate level of privacy protection is envisaged in the countries involved in the processing operations.

Following the same reasoning applied to the interpretation of the Convention, countries that have implemented the Directive are automatically allowed to transborder data flows: complying with the Directive ensures, “ipso iure”, an adequate level of protection.

Although not required for the implementation of BIRO, which process only anonymized data, the Centres involved in the BIRO project belong to European countries that have fully implemented the EU Data Protection Directive, and ratified the Convention. Hence, an adequate level of privacy protection is fully guaranteed across the countries involved. This means that the exchange of data envisaged in the project would be legally viable even if the data processing operations concerned fell under the binding principles of the EU legislation.

Finally, publication of project results will be performed in a form that does not enable not only the data subjects, but also the local centres to be ever identified.

3.2.4.3 Privacy risks and Mitigation Strategies

The potential privacy risks envisaged in the BIRO project could be summarized as follow:

- Data cannot be considered truly anonymous
- Data transmission from local to central database cannot be considered secure
- Performance of global analysis based on non-truly anonymous data could indirectly reveal patients’ identities; for instance through the publication of results.
- Access to central server may be hacked and reversely used to access individual local server and break into personal information stored in computerized registries

The Potential privacy risks have been analysed through a summary table, which allows estimating the best privacy protective alternative in data processing.

The level of risk has been classified as follow:

- Low: There is a possibility that the risk will materialize but there are mitigating factors
- Moderate: There is a strong possibility that the risk will materialize if no corrective measures are taken
- High: There is a near certainty that the risk will materialize if no corrective measures are taken

Anonymization is a crucial factor in the development and implementation of the BIRO project. In order to carry out research on anonymous data outside the application of the Data Protection Directive, data have to be acquired from authorized controllers, local B.I.R.O. Centres, who had already anonymised the data irreversibly; in other words, the data subject re-identification through a reasonable effort has to be impeded before the transfer of data from the local Centres to the Central Database (SEDIS).

Different elements of anonymisation had to be then verified:

- data controller authority to collect and process those data
- purposes of processing
- efficiency of the anonymisation process, according to the state of the art

The local B.I.R.O. Centres collect and process health data according to different national legislations, which grant them authority to collect and process data through diabetes registries and/or databases.

The king of processing performed by local Centres is legitimate according to art. 8 (3) of the EU Data Protection Directive: each center in fact collects information related to an identified or identifiable natural person for the purpose of setting up diabetes registries. Hence, data are to be considered collected and processed for purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health care services; which is one of the purposes considered legitimate for the collection of sensitive data ex art. 8 of the Directive.

The anonymisation techniques used and implemented in BIRO guarantee an irreversible anonymisation. The B.I.R.O. centres, in fact, send only aggregate records to the central server. For the most sensitive variables, aggregated records are not transmitted if groups contain less than five patients. Statistical objects are sent as tables stored in compressed bundles of flat text comma delimited files (CSV). Hence, there is no possibility,

either directly or indirectly, that a patient could be ever identified with a “reasonable effort”.

In broader terms, the privacy of clinical centres has also been considered in the project. The relative privacy risk has been mitigated through the use of pseudonyms for Centre IDs and a reporting system of project results that shows information in percentage rather than in absolute numbers; thus, publication of project results does not reveal, for instance, the size of local Centres, impeding their indirect identification.

Security of transmission: aggregated statistical objects are sent to the central statistical engine to carry out global analysis.

A dedicated communication software has been specifically developed to ensure secure information exchange between the regional systems and the central SEDIS. Global reporting does not pose any direct or indirect risk to privacy, as anonymous data sent by B.I.RO. Centres is transmitted to SEDIS in a secure environment, and further processed in aggregate form.

This task will be performed through updates of the BIRO system.

Relative to the access, security mechanisms are implemented using standard procedures at the strictest level. Once the application will be completely tested, it will be possible to conduct experiments to check the level of security using different hacking techniques.

At a general level, the BIRO Information System processes only de-identified data. Hence, the level of risk can be considered, in most of the cases described, low.

As highlighted in the privacy summary table (Table 3.2.5), efficient mitigation strategies have been implemented in the context of BIRO. Consequently, the aforementioned potential privacy risk could be considered fully avoided and/or removed.

3.2.5 Conclusions

The selected BIRO architecture, considering the characteristics described in the present privacy impact assessment, not only fulfils any privacy requirements, but also foresees the implementation of a system that encompasses privacy concerns at a more general level. Providing for the anonymisation of Centres IDs, which is not a privacy requirement according to EU and international legislation and regulations, could be seen as a means to favouring the respect of privacy beyond the boundaries of individual’s privacy. For instance, addressing issues surrounding professional and institutional integrity in the conduct of health research. These are, in fact, issues to which various stakeholders might be sensible and that could guarantee a better implementation and, eventually enlargement, of the project.

The BIRO architecture, which set up an international health information system that links data sourced by different diabetes registries, flexibly affords the best privacy protection in the construction of an efficient model for the continuous production of European reports in the field of diabetes.

The privacy impact assessment method developed and applied in B.I.R.O. may represent a general tool that can be used to design trans-border health information systems.

Explanatory Notes for Tables 3.2.1,3.2.2,3.2.3

1. Data collected during medical examinations according to a structured procedure within a health service framework e.g. disease management program, systematically organized by means of an electronic database.
2. Clinical centres may be coordinated by a local institution in the framework of a structured program e.g. disease management.
3. For simplicity, data relative to the same subject can be amalgamated over a period of time in various ways. For instance, one may just retain the last measurement of Hba1c or compute the average of different measurements over n months. All other original data for the same variable are not retained. The process is systematically repeated, and the individual record updated or a new individual record appended to the previous for each new time interval.
4. Individual identifier is replaced by a unique, fake identifier created via an algorithm applied by the local database administrator.
5. Same process applied to de-idetified the individual subject is used for clinical centres. Other characteristics that can lead to identify any centre can be blinded, e.g. absolute frequencies are not retained and only percentages are sent to the BIRO central engine
6. Database administrator may decide when to send structured encrypted data bundles to the BIRO server, using ad hoc client software.
7. The client program automatically sends data packets to the BIRO central engine, based on a routine that activates according to a schedule agreed by the database administrator.
8. Information on individual data may be stored averaged over a predetermined time interval.
9. Privileges to access pooled data may be extended to all local BIRO database administrators.
10.European Commission may be in charge of the maintenance of the permanent BIRO Central server
11.Data originated by administrative data flows e.g. hospital discharges, pharmaceutical, mortality data etc.
12. Local government ruling collection of administrative data. In the framework of the present document, a region is intended as a geographical area or even a cluster of geographical areas characterized by homogeneous criteria for data collection. For instance, Tayside may be recognised as a specific region.
However, Scotland applies the same basic set of definitions for data collection, so the BIRO Consortium may even consider the wider geographical area as a single region.
13. Clinical, demographic and socio-economic characteristics of subjects studied in a epidemiological investigation.
14. Institution conducting the epidemiological investigation.
15. Typically, a regional population-based register involves linkage of different data flows, including general administrative data and medical records more targeted at the diabetes population.
16. Aggregated tables strictly relate to the construction of a statistical quantity. For this reason we can also call them as “statistical objects”, as each table is required to apply a particular statistical procedure. For instance, computing the average may only require the total sum of a specific variable, e.g. Length of Stay (LOS), plus the total number of observations related to that sum. A “bundled” table including both entities is a statistical object that can lead to the actual statistical parameter in a subsequent step (central server), where the formula AvLOS=Total (LOS)/n(OBS) is applied. The step is not always so immediate. To compute the median LOS, one requires the entire frequency distribution of LOS at each site/region, i.e. n(OBS) for each level of LOS. The median for all sites/regions is computed from the sum of all frequency distributions collected.
17. Small groups of subjects may lead to the identification of subjects/centres/regions etc. For instance the number of subjects aged 90+ or living in a specific geographical area may be so small and well known that all characteristics stored in tables may be indirectly linked to the specific individual/centre.
18. Since the criterion may be too strict for all variables included in the database, it may be only applied to specific characteristics that are more sensitive to privacy issues.
19. Tables can be used either to carry out reports for the individual region and/or to compute overall results for the BIRO collaboration.
20. Dates pose a specific threat to privacy, as it can be very unlikely that same service or individual characteristic occurs at the same time for different individuals. Therefore it can be an option to approximate dates by weeks or months.
21. Privileges to access pooled data may be extended to all local BIRO database administrators.
22. European Commission may be in charge of the maintenance of the permanent BIRO Central server
23. Publication/exchange of tables stratified by health service centre - as in the case of league tables of performance indicators - is a specific condition affecting “institutional privacy” towards which policy makers can be particularly sensitive. A sharp decision in this regard may involve the restriction to publish all results without using centres as a specific level of aggregation.
24. Risk adjustment techniques may work even without exchanging individual data using different solutions (e.g. pooling multidimensional patterns in logistic regression). However, patterns may lead to very fine stratifications that can pose threats to privacy via indirect identification (low frequencies in specific cells of crosstabulations).
25. Risk adjustment techniques may work even without exchanging individual data using different solutions (e.g. pooling multidimensional patterns in logistic regression). However, patterns may lead to very fine stratifications that can pose threats to privacy via indirect identification (low frequencies in specific cells of crosstabulations).
26. Min N condition may provide a solution to control privacy in sparse cells.
27. Aggregated tables strictly relate to the construction of a statistical quantity. For this reason we can also call them as “statistical objects”, as each table is required to apply a particular statistical procedure. For instance, computing the average may only require the total sum of a specific variable, e.g. Length of Stay.

(LOS), plus the total number of observations related to that sum. A “bundled” table including both entities is a statistical object that can lead to the actual statistical parameter in a subsequent step (central server), where the formula AvLOS=Total (LOS)/n(OBS) is applied. The step is not always so immediate. To compute the median LOS, one requires the entire frequency distribution of LOS at each site/region, i.e. n(OBS) for each level of LOS. The median for all sites/regions is computed from the sum of all frequency distributions collected.

28. Dates pose a specific threat to privacy, as it can be very unlikely that same service or individual characteristic occurs at the same time for different individuals. Therefore it can be an option to approximate dates by weeks or months.

29. Privileges to access pooled data may be extended to all local BIRO database administrators.

30. European Commission may be in charge of the maintenance of the permanent BIRO Central server

31. Geographical characteristics can be highly informative and useful for both epidemiological and policy purposes, but they are prone to privacy issues, as they can link to both the individual and the health service centre.

32. Even though centres’ tables are not made available, one may choose to exchange/publish overall variability of target indicators across centres. For instance, range of performance indicators, or standard deviations. However, these can disclose elements of performance across the region that policy makers may regard as jeopardising institutional privacy.

33. At the level of region, min N=5 may not be considered relevant, so other criteria may be applied.

References

1. Blair Stewart, “Privacy Impact Assessments,” *Privacy Law and Policy Reporter (1996), vol. 39*, available at: http://www.austlii.edu.au/au/journals/PLPR/1996/39.htm
2. David H. Flaherty, “Privacy Impact Assessments: An Essential Tool for Data Protection,” in S. Perrin, H. Black, D.H. Flaherty and T. M. Rankin *The Personal Information Protection and Electronic Documents Act* (Toronto: Irwin Law, 2001), p. 272
3. ICO (2007a) 'Privacy Impact Assessments: International Study of their Application and Effects' Information Commissioner's Office, Wilmslow, I.K., December 2007; available at: http://www.ico.gov.uk/Home/about_us/research/data_protection.aspx
4. Flaherty, Privacy Impact Assessments, p. 266
5. Peter Hope-Tindall, Privacy Impact Assessment – Obligation or Opportunity: The Choice is Ours! 2000-2002 - data Privacy Partners Ltd. Prepared for CSE ITS Conference - Ottawa, Ontario - May 16th, 2002
6. The B.I.R.O. Privacy Impact Assessment Team, Preliminary Privacy Impact Assessment Report, Perugia 2006; available at: http://www.biro-project.eu/index.html
7. James Michael, Privacy and Human Rights 1. (UNESCO) 1994; available at: http://webjcli.ncl.ac.uk/articles1/davies1.html
8. Samuel Warren, Louis Brandeis. The Right to Privacy. Harvard Law Review 1890; 4:193–220
9. David Calcutt QC (Chairman), Report of the Committee on Privacy and Related Matters. London: Cmnd. 11027, 1990
10. Universal Declaration of Human Rights, adopted and proclaimed by General Assembly resolution 217 A (III) of December 10, 1948; available at: http://www.un.org/Overview/rights.html
11. International Covenant on Civil and Political Rights, adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of December 16, 1966, entry into force March 23rd 1976; available at: http://www.unhchr.ch/html/menu3/b/a_ccpr.htm
12. International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families, adopted by General Assembly resolution 45/158 of December 18, 1990; available at: http://www.unhchr.ch/html/menu3/b/m_mwctoc.htm
13. Convention on the Rights of the Child, adopted and opened for signature, ratification and accession by General Assembly resolution 44/25 of November 20, 1989, entry into force September 2, 1990; available at: http://www.unhchr.ch/html/menu3/b/k2crc.htm
14. Council of Europe, Convention for the Protection of Human Rights and Fundamental Freedoms, (ETS no: 005) open for signature November 4, 1950, entry into force September 3, 1950; available at: http://conventions.coe.int/Treaty/EN/cadreprincipal.htm
15. Strossen Nadine, Recent United States and Intl. Judicial Protection of Individual Rights: A comparative Legal Process Analysis and Proposed Synthesis. Hastings Law Journal 1990; 41: 805
16. European Court of Human Rights, Case of Klass and Others: Judgement of 6 September 1978, Series A No. 28 (1979). Malone v. Commissioner of Police, Series A82 (1984); available at: http://hudoc.echr.coe.int/hudoc/ViewRoot.asp?Item=5&Action=Html&X=1007095902&Notice=0&Noticemode=&RelatedMode=0;
17. European Court of Human Rights, Leander v. Sweden, series A No 116 (1987); available at: http://hudoc.echr.coe.int/hudoc/ViewRoot.asp?Item=0&Action=Html&X=1007101431&Notice=0&Noticemode=&RelatedMode=0
18. Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data Convention. Strasbourg, 1981; available at: http://www.coe.fr/eng/legaltxt/108e.htm
19. OECD, Guidelines Governing the Protection of Privacy and Transborder Data Flows of Personal Data. Paris, 1981; available at: http://www.oecd.org/dsti/sti/it/secur/prod/PRIV-EN.HTM
20. Council of Europe Convention on Human rights and Biomedicine (Oviedo 1997); available at: http://conventions.coe.int/Treaty/EN/Treaties/Html/164.htm
21. Charter of Fundamental Rights of the European Union (2000/C 364/01); available at: http://ec.europa.eu/justice_home/unit/charte/index_en.html
22. Official Journal of the European Union C 310 Volume 47 of 16 December 2004; available at: http://eur-lex.europa.eu/JOHtml.do?uri=OJ:C:2004:310:SOM:EN:HTML
23. Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research. Strasbourg, 25.I.2005; available at: http://conventions.coe.int/Treaty/EN/Treaties/Html/195.htm
24. Directive 95/46/EC; available at: http://ec.europa.eu/justice_home/fsj/privacy/law/index_en.htm
25. Roy McClelland at all. European Standards on Confidentiality and Privacy in Healthcare, EuroSOCAP Project (2003-2006); available at: http://www.eurosocap.org/Downloads/European-Standards-on-Confidentiality-and-Privacy-in-Healthcare.pdf

26. Roy McClelland at all. European Standards on Confidentiality and Privacy in Healthcare, EuroSOCAP Project (2003-2006); available at: http://www.eurosocap.org/Downloads/European-Standards-on-Confidentiality-and-Privacy-in-Healthcare.pdf
27. Council of Europe Recommendation No R (97) 5; available at: http://www.coe.int/t/e/legal_affairs/legal_co-operation/data_protection/documents/international_legal_instruments/Rec(97)5_EN.pdf
28. Directive 95/46/EC; available at: http://ec.europa.eu/justice_home/fsj/privacy/law/index_en.htm
29. Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data Convention. Strasbourg, 1981; available at: http://www.coe.fr/eng/legaltxt/108e.htm
30. Council of Europe Recommendation No R (97) 5; available at: http://www.coe.int/t/e/legal_affairs/legal_co-operation/data_protection/documents/international_legal_instruments/Rec(97)5_EN.pdf

The BIRO System

Valentina Baglioni, Pietro Palladino, Peter Beck, Philipp Perner

ABSTRACT

Introduction
The BIRO approach is based on the strategic development of a specialised system that involves standardized processing of electronic medical records and routine data exchange across collaborating centres. The local application of the system provides the method for each centre to extract relevant data from local data sources, performing the same standard statistical analysis and producing a comparable report for selected BIRO indicators. Further to that, there is a need to safely send aggregated data towards a central server that would compile and process all aggregate data to derive an overall statistical report.

Objectives
To develop the BIRO framework and integrate all the different components that would allow implementation of the general architecture and practical use in real life situations, using data from diabetes registers.

Materials and methods
A local BIRO system has been developed to manage the core tasks foreseen by the BIRO architecture. It establishes a connection with the local database where clinical data are stored, extracting relevant records according to specifications given by the common dataset through the “BIRO Adaptor”, resulting in XML BIRO export files. Data concerning a single patient, the profile and the clinical episodes clustered by episode date are then loaded into a BIRO local database using the “BIRO Database Manager”. The system then triggers execution of the BIRO statistical engine to produce the local report and partial results files that are sent to the central BIRO system via Communication software. By compiling results from different centres, the central server produces results for the whole population starting from partial results coming from multiple sites. A central component of the statistical engine deploys the overall BIRO statistical report. In an attempt to transfer BIRO technology easily and effectively, a graphical user interface, the “BIROBox”, has been developed to provide users with a simple instrument capable of managing all functions and the different steps required for the application of the local BIRO system.

Results
The BIRO architecture, designed in accordance with results obtained by the Privacy Impact Assessment, allows to get international results for diabetes indicators without transferring sensitive data out of the boundaries of the local system. Powerful statistical analyses are made possible through the use of a statistical engine realised in R. An integrated system has been built around the BIRO dataset and report specifications, mainly adopting open source tools and Java as a main programming language and cost effective solution. The use of the local BIRO System does not require any changes in the way data are gathered, but it offers additional tools that any centre can routinely use to compare own practice against other units. The BIRO architecture can be potentially used in a recursive fashion to organize the same network within countries and/or regions.

Conclusions
After three years of development, the BIRO system has been finally released according to the original plans of the project. It successfully proposes a self-sustainable solution that can become a fundamental protocol to collect and share diabetes information in Europe. Following the ongoing testing phase, the product will become widely available in the public domain.

3.3.1 Introduction

Building a collaborative, multilevel information system such as the one described by the BIRO project is a challenge for reasons strictly connected to its compelling design.

The architecture of the BIRO project (Figure 3.3.1) envisages the construction of two different subsystems, namely the “Local BIRO System” and the “Central BIRO System”. This means that many procedures implemented on one end must be coupled with similar ones on the other end, to allow for results to be compared across sites and from each site to the European average.

By design, the Local BIRO System (Figure 3.3.2) involves a regional dimension, i.e it will implemented and used within each region or even a centre joining the BIRO Consortium. The system should be at the same time efficient and practical enough to ensure that users with different capacity and skills can apply it with an equal chance to succeed. Original data can be of different formats and content, so users should be made capable of transforming/exporting data to a common format with reasonable effort.

Through the Local BIRO System, and its statistical engine, the BIRO participating centre can extract relevant data from local data sources, perform statistical analysis and automatically create a complete report of BIRO indicators. The report must be ready for use and amenable to be distributed to multiple stakeholders in the health care sector who must be able to read and interpret information herein contained.
The report needs also to be always updated through the repeated application of the entire computational process. The system should be restarted at any time, i.e. whenever a more complete local data set is made available.

The local BIRO system must also ensure the highest attainable level of privacy and protection. Although favoured by a scheme that sends only aggregate data to the Central BIRO System, the implementation must embed state of the art technology for security, systematically avoiding any major threat that would be negatively perceived by participating centres.

Finally, the Central BIRO System (Figure 3.3.3) the implementation must embed state of the art technology for security (Figure 3.3.4), systematically all aggregates sent by different centres, storing tables into a central BIRO database. An overall statistical report must be obtained through the application of routines that should match the set of procedures applied by the statistical engine locally, to produce the same standardized report that will be published on the BIRO web portal.

Development of the BIRO system must deliver an integrated package allowing accomplishment of the above tasks, extending the use of the approach in different directions. Information infrastructure must be flexible enough to allow both horizontal and vertical expansion through a recursive structure. New regions interested in joining the BIRO Consortium must be able to do so with minimal effort; linking data from multiple heterogeneous sources even within individual countries, or regions, must be relatively straightforward, allowing for the scheme to be replicated.

Here we describe the main components in the implementation of the BIRO system, focusing on the engineering, database management, communication soft-

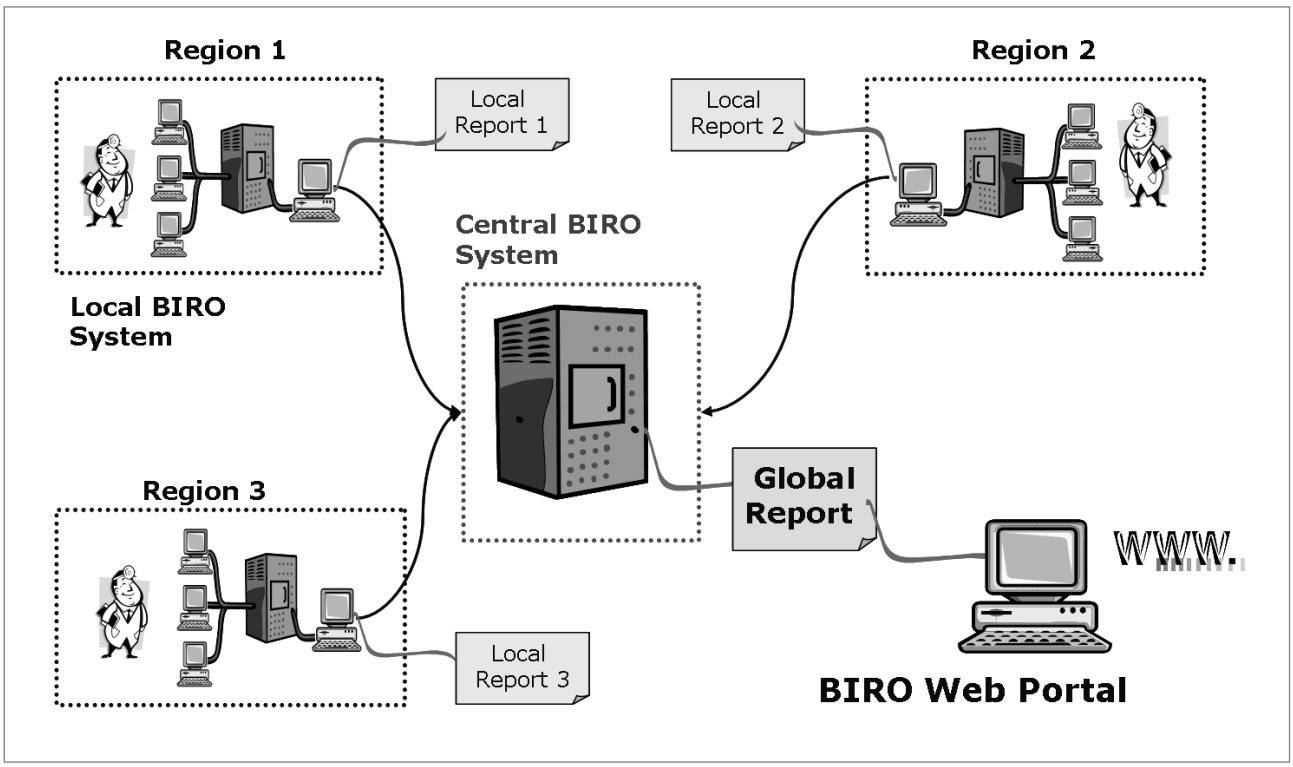


Figure 3.3.1: BIRO architecture

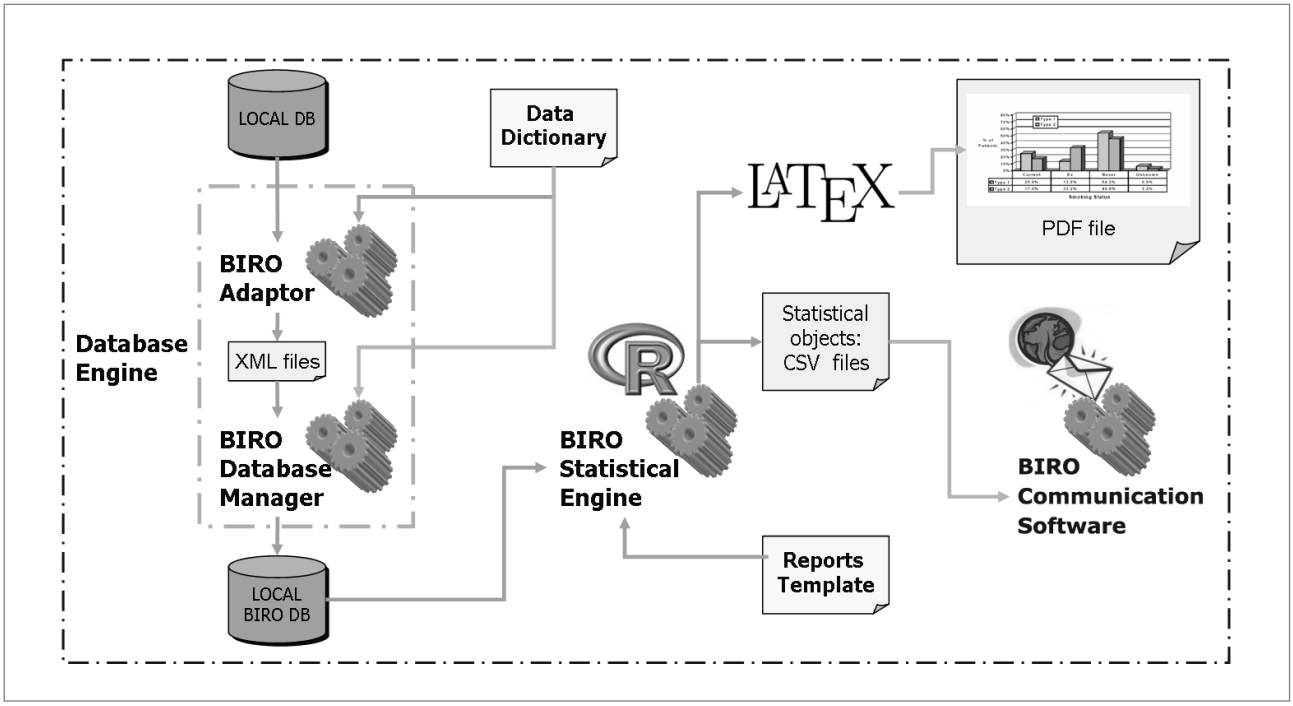


Figure 3.3.2: The Local BIRO System

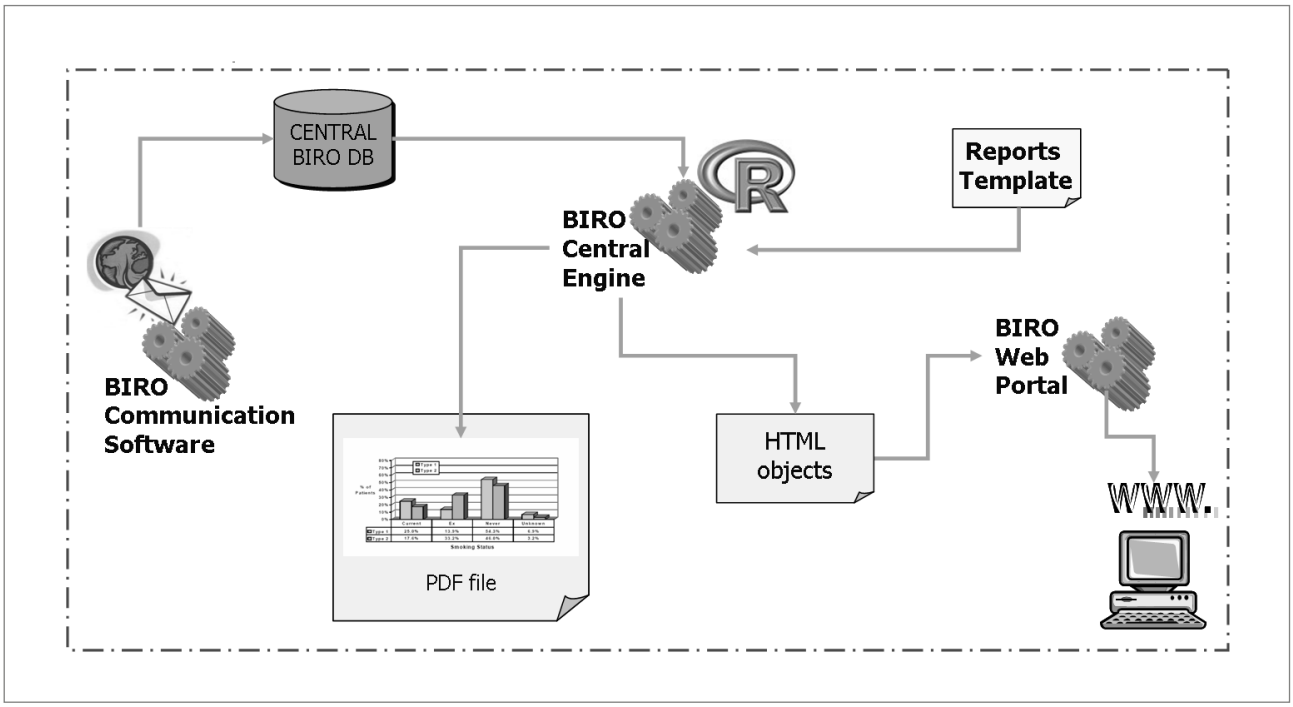


Figure 3.3.3: The Central BIRO System

ware, and the integration of different modules with a visual interface. Details on the statistical engine at the core of the BIRO system are provided more specifically in Chapter 3.4.

3.3.2 Materials and Methods

Several components have been designed to fulfil the scope of the BIRO System: the Statistical Engine, the Database Engine, including the Adaptor and Database Manager, the Communication Software and the BIROBox.

BIRO Statistical Engine

The Statistical Engine is split into a local and central component performing statistical analysis for report delivery at both ends. Details of methods and results obtained are included in Chapter 3.4.

BIRO Database Engine

The Database Engine provides an interface between an undefined original data source compatible with the scope of BIRO (e.g. a diabetes register), and the BIRO

environment. Heterogeneous input data must be transformed into a standard BIRO format, allowing the Statistical Engine to work properly, in connection with a common database architecture.
A software tool must be developed to be directly used by each centre to perform the following tasks: connection with different kinds of local data sources, extraction of relevant data for BIRO purposes, mapping of local dataset into a “BIRO compliant” dataset, export of mapped data in the form of xml files, construction and population of a local BIRO database.
The Database Engine is closely linked to the Common Dataset and the BIRO Data Dictionary, offering a detailed description of the BIRO dataset format, the schema for xml export files, and the architecture of the BIRO Database.
Different tasks are managed by two different software components: the BIRO Adaptor and BIRO Database Manager.

The *BIRO Adaptor* is a small piece of software allowing the user to connect to a data source, to retrieve data using the SQL Query language, to organize data into BIRO structures, and to write down an XML file using the BIRO XML Schema. It is completely written in Java language.
Using the JDBC layer, this tool is able to connect and retrieve data from many different database management systems provided we have the JDBC Java driver (available for many DBMS such as PostgreSQL, MySQL, SQL Server, Oracle).

The BIRO *Database Manager* is required to parse the BIRO Export XML files produced by BIRO Adaptor and store data into the BIRO Database.

The data transfer from the XML document towards the database is planned to occur in two steps (Figure 3.3.5).
Firstly, XML files produced by BIRO Adaptor are parsed and data translated in some appropriate Java Objects. This procedure is called “unmarshalling”, i.e. creating a mapping between elements of the XML document and members of a class to be represented in memory. The reverse process, i.e. to serialize a Java Object as XML, is called “marshalling”. This phase includes two phases: parsing BIRO XML Schema and creation of an abstract Object Model including a java class for each element in the schema (Patient, Profile, EpisodeData, Data, etc.); and parsing the BIRO Export XML files compliant with the BIRO XML Schema to produce instances (of Patient, Profile, EpisodeData, Data, etc.). The second step, “storing” is performed to transform hierarchically structured Java Objects into columns and records of the local BIRO database. Even in storing there are two phases: the abstract Object Model is used to create the structure of the BIRO Database (tables, columns, primary keys, foreign keys, etc.); and data contained in the fields of Java Objects are inserted as records into the appropriate tables of BIRO Database (Figure 3.3.6).

Tools to implement the BIRO Database Manager are free and open source.

The unmarshalling process is accomplished trough a data-binding framework called Castor[4], an Open Source data-binding framework for Java. It is the shortest path between Java objects, XML documents and relational tables. In particular, Castor XML and the Source Code Generator have been widely used within the Database Manager. Castor XML performs automatic XML data binding for class definitions which adhere to the Java Beans design pattern. Java to XML mapping automates transformation of Java objects to and from XML documents and provides Java object validation. This way it is possible to write XML based mapping file to specify XML bindings for existing object models. The Source Code Generator can produce Java class definitions, XML binding information, and validation code based on a provided XML schema. The Source Code Generator supports a "customization" binding file used in conjunction with an XML Schema for greater control over the generated source code.

Storing uses a persistence framework called Hibernate[5], a powerful, high performance object-relational mapping (ORM) library for the Java language, providing a framework for mapping an object-oriented domain model to a traditional relational database. Hibernate's primary feature is the mapping from Java classes to database tables (and from Java data types to SQL data types). Hibernate also provides data query and retrieval facilities. Hibernate generates the SQL calls and relieves the developer from manual result set handling and object conversion, keeping the application portable to all SQL databases, with database portability delivered at very little performance overhead.

Other alternatives have been considered for saving information of an XML documents into a database, e.g. W3C DOM API[6] and SAX API[7]. Through them the user deals directly with the structural components (elements, attributes) of XML files, manipulating XML at low level and extracting pieces of data.

Frameworks operate at a higher level of abstraction therefore all the XML manipulations are transparent

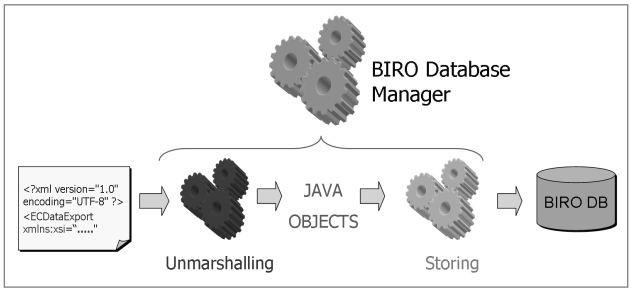


Figure 3.3.5: Unmarshalling and storing sections

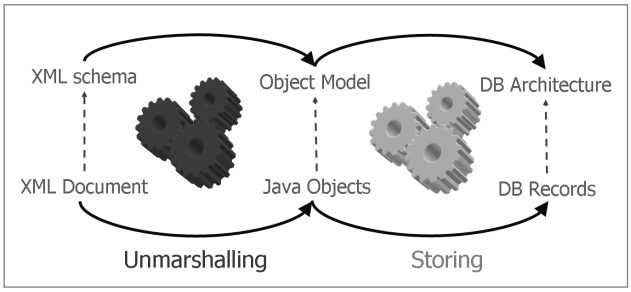


Figure 3.3.6: unmarshalling and storing section in detail

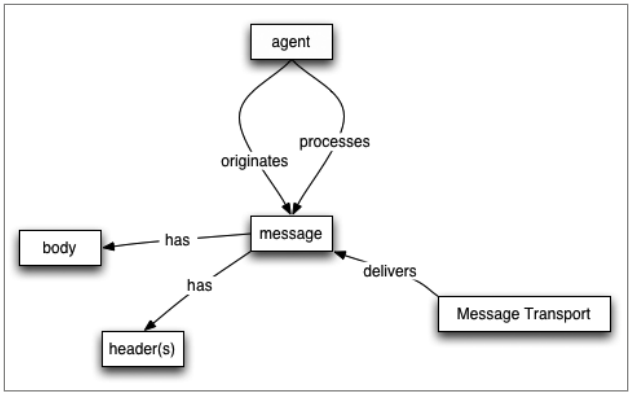


Figure 3.3.7: Message Oriented Model

to the user. Using them requires a very little effort; creating Java Objects allow dealing with the logical model of data instead of their XML representation. Data conversion and validation is done automatically. If needed, changes to the XML schema can be done easily even when the schema is growing large. The only drawback of frameworks is that creating Java objects requires long time therefore performance is worse than working on XML at low level.

BIRO Communication Software

Web-Services have been regarded as the most suitable approach for communications occurring in the BIRO system. Several protocols/standards have been selected for the scope for which some basic information is reported in the following paragraphs.

SOAP (since version 1.2 no longer an acronym for "Simple Object Access Protocol") is a lightweight protocol intended for exchanging structured information in a decentralized, distributed environment. It uses XML technologies to define an extensible messaging framework, providing a message construct that can be exchanged over a variety of underlying protocols. The framework has been designed to be independent of any particular programming model and other implementation specific semantics.
A *Web service* is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format, i.e. the Web Service Description Language (WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards[8].

A web service by definition is described with 6 major elements:

- Types: Provides data type definitions used to describe the messages exchanged.
- Message: Represents an abstract definition of the data being transmitted. A message consists of logical parts, each of which is associated with a definition within some type system.
- PortType: A set of abstract operations. Each operation refers to an input message and output messages.
- Binding: Concrete protocol and data format specifications for the operations and messages defined by a particular portType.
- Port: Address for a binding, thus defining a single communication endpoint.
- Service: Aggregation of a set of related ports.

Web services provide a standard means of interoperating between different software applications, running on a variety of platforms and frameworks.

The web-service architecture can involve different dimensions:

- Service Oriented Model
- Message Oriented Model
- Policy Model - focusing on policies, security and quality of service. A policy is a constraint concerning allowable actions of agents, persons or organisations.
- Resource Oriented Model - focusing on resources, i.e. everything that is owned by anyone and therefore is able to have policy constraints.

The perspective of the Service Oriented Model and the Message Oriented Model deserves to be described in more detail.

A briefer look on the *message oriented model* (MOM, see Figure 3.3.7) reveals the underlying actors and elements, which focus on the structure of a message, the relationship between sender and receiver and the transmission.

- *Agent*: an agent is a computational resource which is owned by either a person or an organisation. Agents can be clients (service-requester) or servers (service-provider).
- *Message*: a message is a basic unit of data sent from one agent to another. A message consists of a message envelope, which contains the message body and zero or more message headers.
- *Body*: a message body represents the primary application-specific content sent from the client-agent to the service-agent.
- *Header*: a message header represents information about a message which is needed for modular processing of a message. Headers usually describe extended Web-Service functionality like security, transaction or routing.
- *Message Transport*: The message transport is the mechanism used by agents to deliver messages. Possible mechanisms can be HTTP, TCP or SMTP. The responsibility of the message transport is the actual delivery of a SOAP message from a sender to a receiver.

The *service oriented model* (SOM, see Figure 3.3.8) reveals the underlying actors and elements, focusing on the architecture that relates to service and action. It builds on the MOM but rather describes the relationships between agents and services than on structures of messages.

- *Agent*: an agent is a computational resource which is owned by either a person or an organisation. Agents

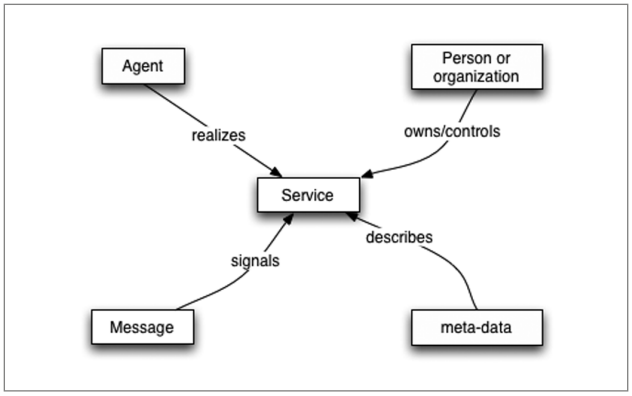


Figure 3.3.8: Service Oriented Model

can be clients (service-requester) or servers (service-provider).
- *Message*: a message is a basic unit of data sent from one agent to another. A message consists of a message envelope, which contains the message body and zero or more message headers.
- *Person or organization*: A person or organization is the owner of agents that provide or request web-services.
- *Meta-data*: the meta-data which describes the service can be regarded as partial machine-readable description of the semantics of a Web-service. In particular WSDL, an xml-based language, provides a model for describing web-services, especially the expected messages and data types of the message's elements.

Web services follow the concept of message-oriented software, because the main components for interactions are messages sent from a webservice requester agent to a webservice provider agent

Web Services Security (WSS), developed by OASIS (Organization for the Advancement of Structured Information Standards), defines a standard set of SOAP-extensions that can be used when building secure

Web services, although the standard itself just provides an abstract message security model, which includes three main security mechanisms:
- Security tokens (Authentication)
- Message integrity (Signature)
- Message confidentiality (Encryption)

These mechanisms can be used independently or in a combination with each other. Possible implementations are security models like Public Key Infrastructure like X.509 certificates, Kerberos authentication protocol and SSL Transport Layer Security.[9]
The standard is designed to make use of existing security enhancing technologies, like XMLSignature[10] and XML Encryption[11].

Various specific standards define the usage of concrete security implementations
- X.509 Certificate Token Profile defines the usage of X.509-certificates as described in PKCS#12[12] (Public-Key Cryptography Standards).
- Username Token Profile describes the usage of a username/password combination for authentication.
- SAML Token Profile makes use of the xml-based standard Security Assertion Markup Language Token, a Single Sign On solution mostly for business-to-business or business-to-costumer transactions.
- Kerberos Token Profile makes use of Kerberos protocol for network authentication
- Rights Expression Language (REL) Token Profile uses the international standard ISO/IEC 21000-5 relating to digital media resources.

Multiple frameworks, which implement the web-service architecture standard, are available either for free as open source or under special licenses as purchasable software products.
As BIRO tries to use open source software and to keep licensing fees as low as possible, only available free and open source frameworks were investigated for suitability in the BIRO environment.

As far as the *Web-service* is concerned, considering the requirements of a lightweight SOAP framework providing mechanisms of security as stated in the requirements analysis, Apache Axis2[13] was considered as most suitable and reliable for the usage in the BIRO system. It is a core engine for creating web services and transmitting SOAP messages according to the implementation of the W3C's SOAP submission. Axis2 is an open-source framework written in Java and licensed under the Apache Software License. Axis2 supports several specifications of the web service architecture, e.g. SOAP protocol 1.1 and 1.2 as well as Message Transmission Optimization Mechanism (MTOM), XML optimized Packaging (XOP) and SOAP with attachments (SwA). Moreover the webservice description language 1.1, native implementations for WS-Addressing and WS-Policy as well as the SOAP with Attachments API for Java is supported. Axis2 provides messaging over HTTP, SMTP (mail), JMS (Java Message Service) or directly over TCP.

In terms of *security*, we have examined Apache Rampart by the Apache Software Foundation as a possibility to integrate OASIS' WSS-specification in Apache AXIS2.

Apache Rampart is based on Apache's own project Web-Service Security for Java (WSS4J[14]), i.e. the implementation of the WS-Security specification. WSS4J is primarily a Java library that can be used to sign and verify SOAP messages with WS-Security information.
WSS4J implements the specific security-standards SOAP Message Security, X.509 Certificate Token Profile and Username Token Profile.
As the security-implementation Apache Rampart ships as module, it can be integrated on different scopes. Security-bindings use different tokens to secure the message exchange. As mentioned above AXIS2 implements different specific standards as claimed by OASIS. Possible items for securing a message are *"UsernameToken"*, *"Timestamp"*, *"Signature"* or *"Encryption"*. Security-bindings need additional information for proper configuration of the required security level, e.g. for Signature or Encryption an X.509-certificate is needed for the Public-Key Encryption.

From the point of view of *Service Provider and Service Requester*, we have taken into account Public Key Infrastructure (PKI) as a term used for a framework that enables secure exchange of information based on public key cryptography.
PKI allows identities (of people, organizations, etc.) to be bound to digital certificates and provides a means of verifying the authenticity of certificates. PKI encompasses keys, certificates, public key encryption, and trusted Certification Authorities (CAs) who generate and digitally sign certificates.

The BIRO Box

The initial plan for the BIRO Adaptor and BIRO Database Manager was to execute tools only from console, with options chosen through configuration files. However, the increase of mapping choices and configuration features have rapidly made a graphical user interface (GUI) necessary. Moreover, the different components developed in BIRO could be more easily integrated through a common “Box”.

The design of the BIROBox GUI was finally undertaken to solve the above problems. The Box was rapidly identified as the optimal solution to interface all tools in the BIRO framework, including the statistical engine and communication software.

Therefore, the BIRObox (Figure 3.3.9) is what the real user would see of the local BIRO System, an intuitive user interface, through which it is possible to configure and run all functions.

3.3.3 Results

Development of different components of the BIRO system occurred separately and was then integrated through the design and implementation of the BIROBox. Here we describe the results obtained separately for all components.

BIRO Adaptor

From a programmer's perspective, the BIRO Adaptor has been fully realised in Java language (Figure 3.3.10) through the development of several classes that are here briefly described.

BIROAdaptor2 is the core class of BIRO Adaptor architecture: it establishes a connection with the local database using the appropriate driver set in the BIROAdaptorConfigurationFile and manages the connection through the specific DBStreamer class that is also responsible of executing the queries that extract data from the local database. An output file is initialized through the FileWriter class, which opens the necessary stream towards the zip file that will contain the XML files relative to individual medical records. Finally, this class manages data writing into each XML file through the following actions: requesting the DBStreamer to execute the query retrieving data from the mergetable; scanning the ResultSet record by record and for each different subject for which a new XML file is created. The FileWriter is responsible for writing nodes of XML files related to profile fields, activity fields and episode data fields.

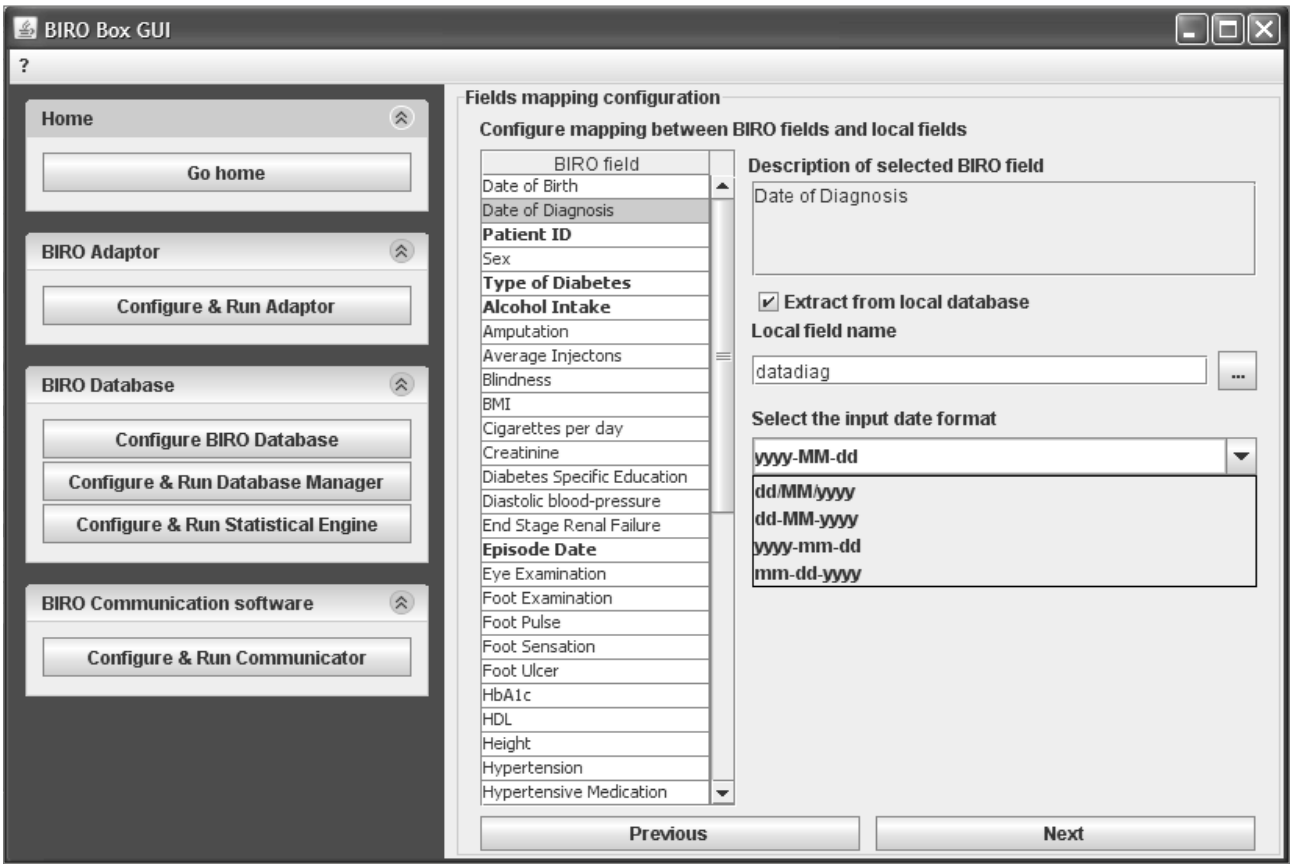


Figure 3.3.9: Screenshot of BIROBox

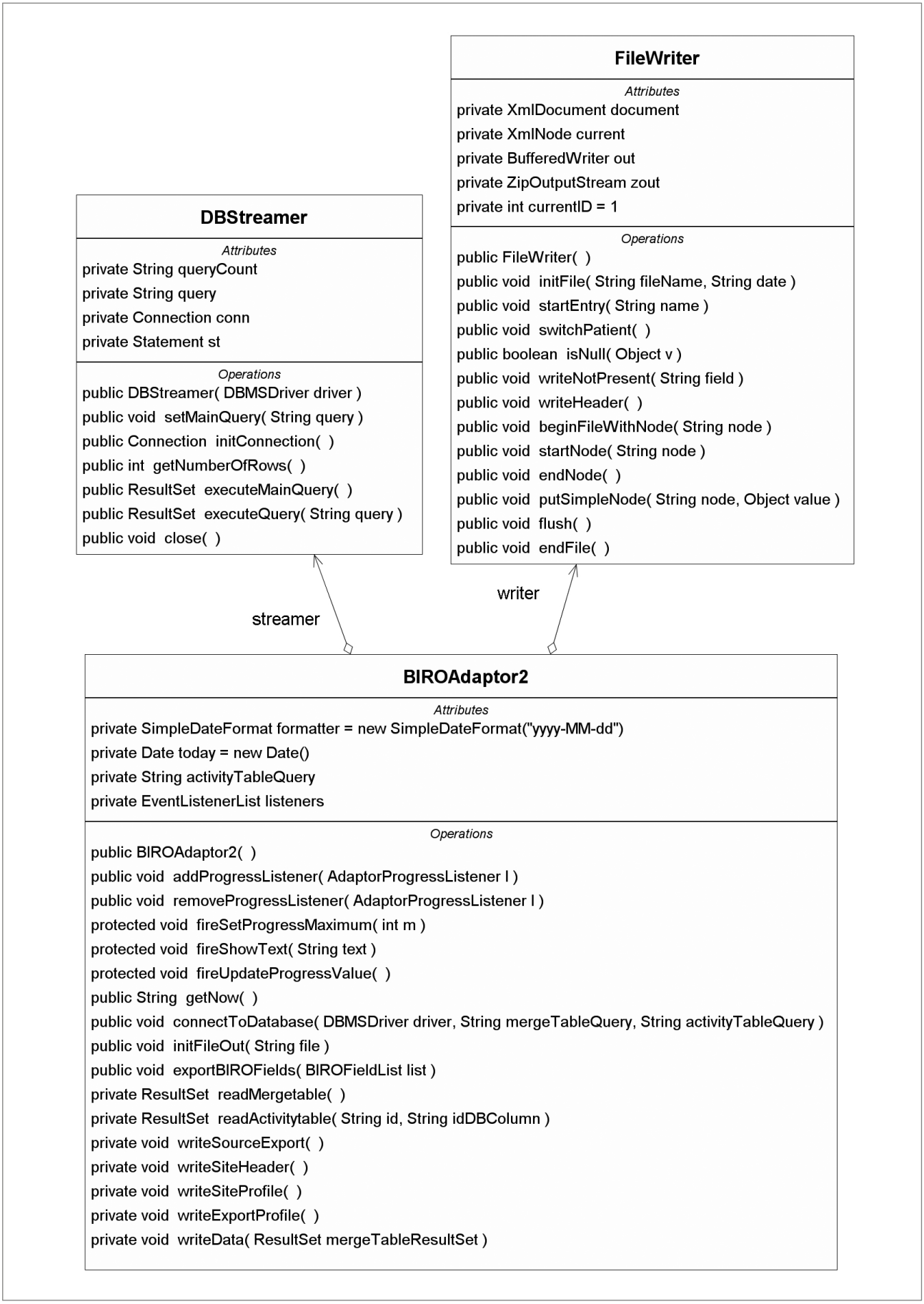


Figure 3.3.10: UML representation of BIROAdaptor architecture

The BIROAdaptorConfiguration class (Figure 3.3.11) includes the following information for the proper functioning of BIROAdaptor:

- dBMSDriver: all the details to be used to connect to the local data source
- exportFilePath: a String representing the complete path chosen by the user to save the XML export ZIP file
- mergeTableQuery and activityTableQuery: two String representing the SQL query to be executed to retrieve the mergetable and the activitytable from the local source. There may be simple select queries if those tables are already present in the local database, or more complex ones, if records are to be retrieved by joining different tables.
- biroFieldList: the list of BIRO fields (Figure 3.3.12) with all the necessary information on mapping saved by the user

The AbstractBIROField class represents characteristics and methods common to all BIRO fields. Every different kind of BIRO field (EnumeratedBIROField, DateBIROField, SimpleBIROField, NumericBIROField) has a different implementation of the “getValueToWrite” method, which returns the mapped value to be written into an XML node. A Java class is created for each BIRO field. Each class extends the appropriate field type and contains the name, description and specific characteristics depending on the field type (supported unit of measurement for NumericBIROField, default value for EnumeratedBIROFields, etc.).

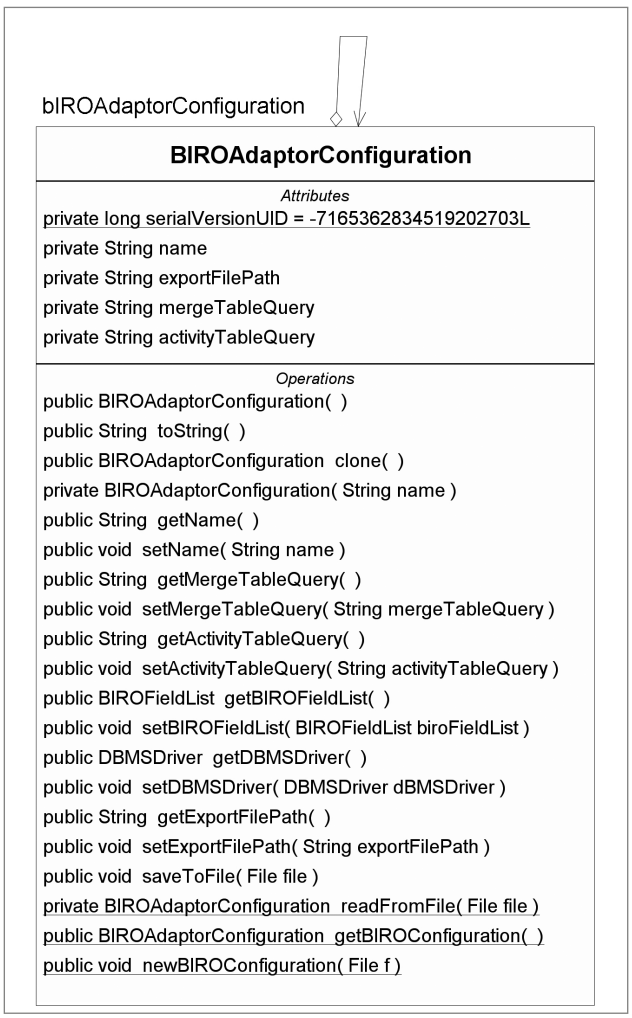


Figure 3.3.11: UML representation of BIROAdaptorConfiguration class

From the user perspective, the Adaptor implements JDBC connections for the most common DBMS, so that the only information needed to access the local database are host, port, database name, database username and password. If any of the provided drivers is suitable for the local DBMS, then it is also possible to include a new JDBC driver. However, in such situation is preferable to have an expert user involved, since the system requires specification of technical details e.g. URL pattern, driver class name and driver file name. A more practical option is to read straight, flat text files as input for the Adaptor.
Once the connection has been established, the Adaptor requires details on how data are stored in the database or in the text file in order to be able to retrieve them.

The BIRO Adaptor normally expects to work mainly with two tables.
The first one, hereinafter called *mergetable*, should have the following structure: {patient_ID, episode_date, [data field],...}where the couple patient_ID and episode_date represents the primary key of the table. In other words, each row of the mergetable should represent a specific episode of a specific patient. The second table, celled *activitytable,* should contain information about the movement of patients with respect to the centre, i.e. dates of entry and exit from the centre and the related reasons (diagnosis, transfer toward/from another centre, death, loss to follow-up). The *activitytable* is very important for the calculation of indicators, since statistical routines should not include patients not “active” in that centre for a particular time interval.

Other cases are possible, but must lead to the creation of the required tables. For instance, the user can submit to the Adaptor an SQL query joining two or more tables to obtain a temporary view. As an alternative, the user may also pre-process data to create temporary tables in the database before running the tool.
The definition of the merge query or the pre-processing of the database is the only technical work that BIRO users are asked to produce. Once the merge is ready, the following steps are performed automatically without any difficult intervention of the user.

Some mapping rules between the local dataset and the BIRO standard dataset have to be defined prior to run the Adaptor. For each BIRO field, the user should specify: if it is present in the local dataset; if it can be extracted or not; and the local name.
Depending on the BIRO field type, the user should provide the Adaptor with some additional details: the local date format for date fields, the local unit of measurement for numeric fields; the correspondence between local enumerated values and BIRO values in case of enumerated BIRO fields.
Finally, the user may establish the time range (start date, end date) for clinical episodes to be extracted. At this point, the Adaptor will extract only data within a specified time range, then it will automatically do all

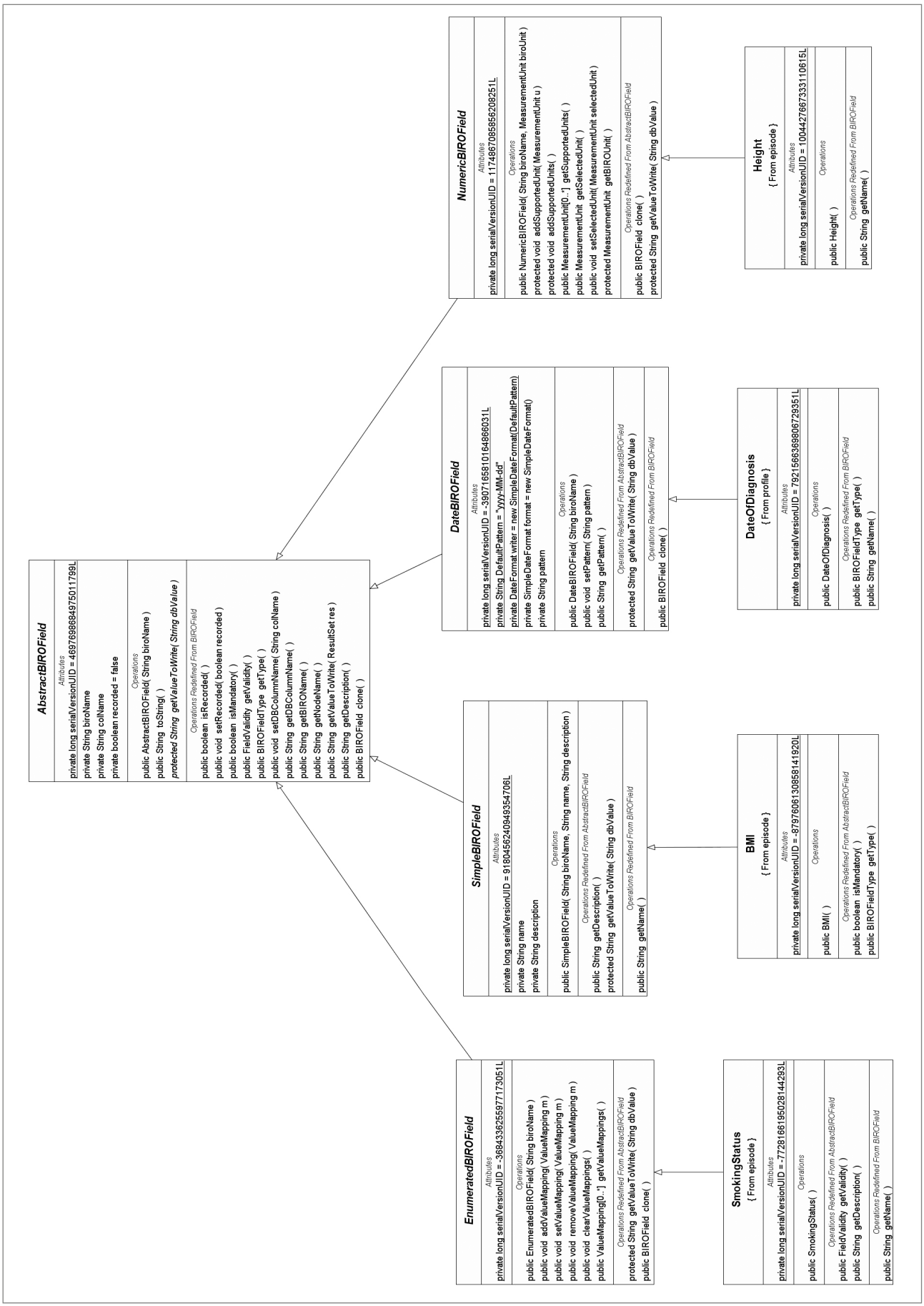


Figure 3.3.12: BIRO Fields implementation within BIROAdaptor

Box 3.3.1: XML Schema - Site Header Extract

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" version="0.4">
<xsd:element name="ECDataSourceExport">
         <xsd:complexType>
               <xsd:sequence>
                        <xsd:element name="SiteHeader">
                                 <xsd:complexType>
                                  <xsd:sequence>
                                  <xsd:element name="DateHeaderInformationChecked"
type="xsd:date"/>
                                  </xsd:element>
                     <xsd:element name="DS_ID" type="DataSource" id="BIRO002">
                    </xsd:element>
                    <xsd:element name="DS_WEBSITE" type="xsd:string" id="BIRO106" minOccurs="0">
                    </xsd:element>
                    <xsd:element name="DS_ADDRESS_1" type="xsd:string" id="BIRO107">
                    </xsd:element>
                    <xsd:element name="DS_ADDRESS_2" type="xsd:string" id="BIRO108">
                    </xsd:element>
                    <xsd:element name="DS_ADDRESS_3" type="xsd:string" id="BIRO109"
                    minOccurs="0">
                    </xsd:element>
                    <xsd:element name="DS_ADDRESS_4" type="xsd:string" id="BIRO110"
                    minOccurs="0">
                    </xsd:element>
                    <xsd:element name="DS_POST_CODE" type="xsd:string" id="BIRO111"
                    minOccurs="0">
                    </xsd:element>
                    <xsd:element name="DS_COUNTRY" type="xsd:string" id="BIRO101">
                    </xsd:element>
                    <xsd:element name="DS_C_CONTACT" type="xsd:string" id="BIRO112">
                    </xsd:element>
                    <xsd:element name="DS_C_EMAIL" type="xsd:string" id="BIRO113">
                    </xsd:element>
                    <xsd:element name="DS_T_CONTACT" type="xsd:string" id="BIRO114">
                    </xsd:element>
                    <xsd:element name="DS_T_EMAIL" type="xsd:string" id="BIRO115">
                    </xsd:element>
                    <xsd:element name="HeaderComments" type="xsd:string" minOccurs="0"/>
                    </xsd:sequence>
             </xsd:complexType>
            </xsd:element>
              </xsd:sequence>
        </xsd:ComplexType>
</xsd:element>
</xsd:schema>
```

conversions needed, and then write down the corresponding XML files packed in a single zip file.

BIRO Database Manager

The BIRODatabaseManager presents a very simple architecture composed by four major classes.

The BIROClassGenerator is a stand alone application that reads the XML schema and automatically generates the source code of Java classes related to the elements in the XML schema. It is very useful for the development and maintenance of BIRO Database Manager because if changes to the XML schema are needed, the structure of Java classes can be immediately updated accordingly. The BIROClassGenerator uses the Castor feature named SourceGenerator. It requires as input the BIRO XML schema and a binding file, which is an XML file used to specify some customized constraints regarding class names and data types to be respected while generating Java source code.

In Box 3.3.1 and Figure 3.3.13 there are, respectively, an abstract of the XML schema regarding Site Header information and the UML representation of the generated SiteHeader Java class. SourceGenerator transforms all the elements in attributes of SiteHeader class, then it creates the related getter and setter methods and finally it adds methods for marshalling, unmarshalling and validation.

The mapping between Java Object Model and Relational Model is specified by means of one or more XML documents named mapping files. The mapping file is designed to be Java centric: for each class and each attribute, the document specifies the associated table and column and constraints. These files are used during the storing section to create the database architecture and to store at the right place the unmarshalled Java Objects. An example of the mapping file for SiteHeader class is reported in Box 3.3.2 and Box 3.3.3 contains a description of the SiteHeader table as it will appear within the BIRO Local database.

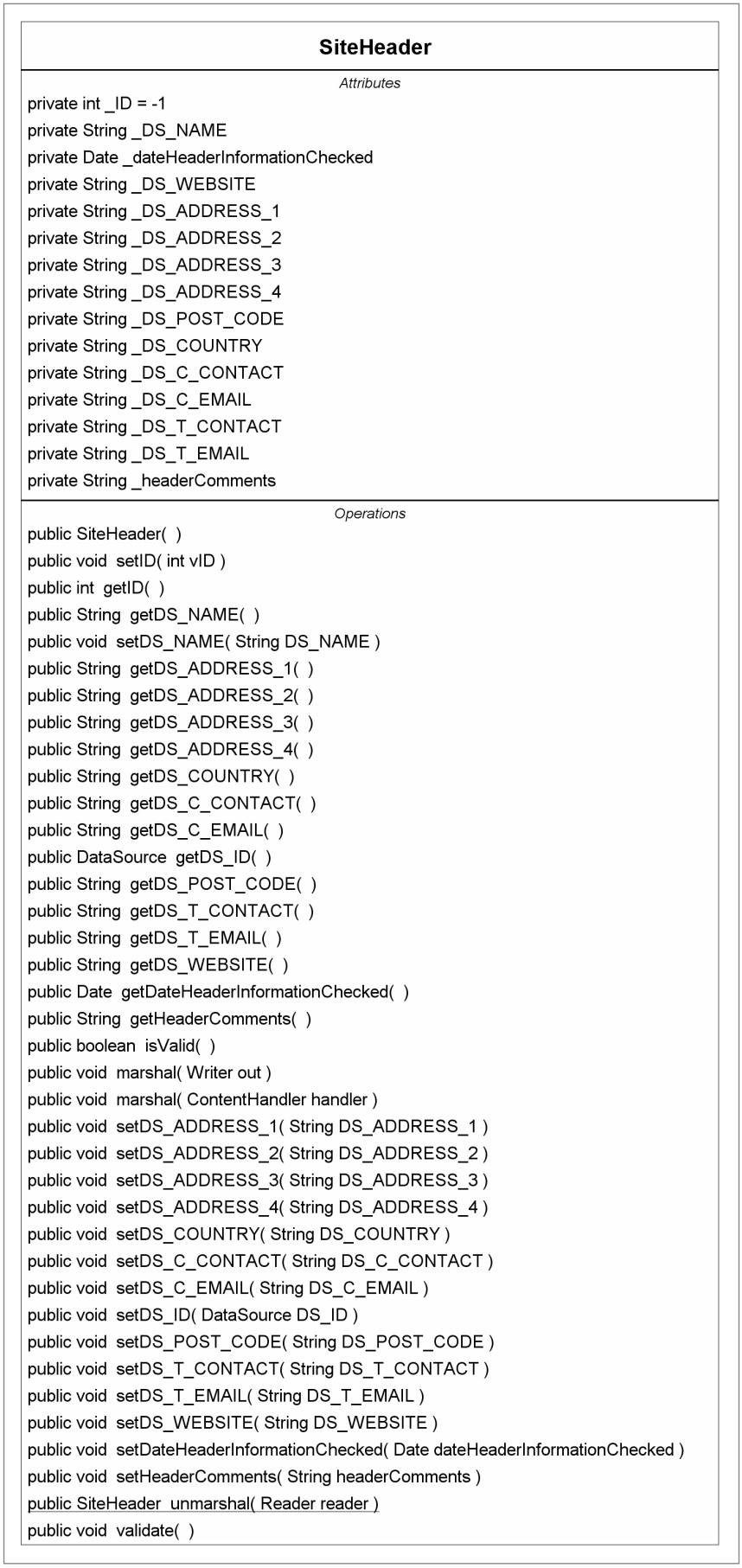


Figure 3.3.13: UML diagram for SiteHeader class

BOX 3.3.2: Example of mapping file for SiteHeader class

```
<?xml version="1.0"?>
<!DOCTYPE hibernate-mapping PUBLIC
               "-//Hibernate/Hibernate Mapping DTD 3.0//EN"
               "http://hibernate.sourceforge.net/hibernate-mapping-3.0.dtd">
<hibernate-mapping>
    <class name="export.SiteHeader" table="SITE_HEADER">
        <id name="ID" column="SITE_HEADER_ID"><generator class="native"/></id>
        <property name="hibernateDateHeaderInformationChecked" type="date"
column="DATE_HEADER_INFORMATION_CHECKED"/>
        <property name="DS_ID" type="export.types.DataSourceUserType" column="DS_ID"/>
        <property name="DS_WEBSITE" type="string" column="DS_WEBSITE"/>
        <property name="DS_ADDRESS_1" type="string" column="DS_ADDRESS_1"/>
        <property name="DS_ADDRESS_2" type="string" column="DS_ADDRESS_2"/>
        <property name="DS_ADDRESS_3" type="string" column="DS_ADDRESS_3"/>
        <property name="DS_ADDRESS_4" type="string" column="DS_ADDRESS_4"/>
        <property name="DS_POST_CODE" type="string" column="DS_POST_CODE"/>
        <property name="DS_COUNTRY" type="string" column="DS_COUNTRY"/>
        <property name="DS_C_CONTACT" type="string" column="DS_C_CONTACT"/>
        <property name="DS_C_EMAIL" type="string" column="DS_C_EMAIL"/>
        <property name="DS_T_CONTACT" type="string" column="DS_T_CONTACT"/>
        <property name="DS_T_EMAIL" type="string" column="DS_T_EMAIL"/>
        <property name="headerComments" type="string" column="HEADER_COMMENTS"/>

    </class>
</hibernate-mapping>
```

Box 3.3.3: Example of data definition code for site_header table inside BIRO Local Database

```
CREATE TABLE site_header
(
  site_header_id integer NOT NULL,
  date_header_information_checked date,
  ds_id character varying(255),
  ds_website character varying(255),
  ds_address_1 character varying(255),
  ds_address_2 character varying(255),
  ds_address_3 character varying(255),
  ds_address_4 character varying(255),
  ds_post_code character varying(255),
  ds_country character varying(255),
  ds_c_contact character varying(255),
  ds_c_email character varying(255),
  ds_t_contact character varying(255),
  ds_t_email character varying(255),
  header_comments character varying(255),
  CONSTRAINT site_header_pkey PRIMARY KEY (site_header_id)
)
```

BIRODatabaseManager is the main class. It reads the zipped folder containing the BIRO Export XML files and sends each file to the UnmarshallingAndStoring-Manager which calls the appropriate Castor utilities to unmarshall the file and then calls the appropriate Hibernate utilities to store the Java objects into the database. HibernateUtil class handles the Database-Manager configuration, composed by the mapping files and the database connection settings.

BIRO Communication Software

The BIRO Web-Service implementation was performed using the WSDL-first design approach. For this purpose a WSDL-file according to the W3C Standard Web Service Description Language[15] was generated, which describes the web-service, the operations, the corresponding messages, their contents, service endpoints and SOAP bindings.

The web-service consists of one PortType, which is linked to the service via SOAP 1.1 and SOAP 1.2 message-bindings. The PortType consists of one abstract operation “transferData” which consists of the input message, i.e. request, “transferDate” and the output message, i.e. response, “transferDataResponse”.

An illustration of the service, the portTypes including the operations and the request and response message can be seen in Figures 3.3.14-16. The complete description of the Web-Service (wsdl-File) is shown in Box 3.3.5.

AXIS2 provides tools and techniques for Code-Generation from WSDL-files. AXIS2 uses an internal Code-Generation-Engine, which creates an XML-file. This XML-file is transformed by an XSLT-engine by using different templates, i.e. XSL-stylesheets. The generated code provides programming-language-

Box 3.3.5 Web-Service Description of the B.I.R.O. Communication Service

```
<?xml version="1.0" encoding="UTF-8"?>
<wsdl:definitions xmlns:http="http://schemas.xmlsoap.org/wsdl/http/"
 xmlns:mime="http://schemas.xmlsoap.org/wsdl/mime/"
 xmlns:soap12="http://schemas.xmlsoap.org/wsdl/soap12/"
 xmlns:types="http://TransferService.communication.biro.eu/types"
 xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
 xmlns:biroComm="http://TransferService.communication.biro.eu"
 xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/"
 targetNamespace="http://TransferService.communication.biro.eu" name="BiroCommunicationService">
 <wsdl:documentation>BiroCommunicationService</wsdl:documentation>
  <wsdl:types>
   <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" attributeFormDefault="qualified"
    elementFormDefault="qualified"
    targetNamespace="http://TransferService.communication.biro.eu/types">
    <xs:element name="transferData">
     <xs:complexType><xs:sequence>
     <xs:element name="data" nillable="true" type="xs:base64Binary"/>
     <xs:element name="datasource" nillable="true" type="xs:string"/>
     <xs:element name="timeframeStart" nillable="true" type="xs:date"/>
     <xs:element name="timeframeEnd" nillable="true" type="xs:date"/>
     </xs:sequence></xs:complexType></xs:element>
     <xs:element name="transferDataResponse">
     <xs:complexType><xs:sequence>
     <xs:element name="return" nillable="true" type="xs:string"/>
     <xs:element name="serverMessage" nillable="true" type="xs:string"/>
    </xs:sequence></xs:complexType>
   </xs:element></xs:schema>
  </wsdl:types>
  <wsdl:message name="transferDataMessage">
   <wsdl:part name="part1" element="types:transferData"/>
   </wsdl:message><wsdl:message name="transferDataResponseMessage">
    <wsdl:part name="part1" element="types:transferDataResponse"/></wsdl:message>
    <wsdl:portType name="BiroCommunicationServicePortType">
     <wsdl:operation name="transferData">
     <wsdl:input xmlns:wsaw="http://www.w3.org/2006/05/addressing/wsdl"
      message="biroComm:transferDataMessage" wsaw:Action="urn:transferData"/>
     <wsdl:output xmlns:wsaw="http://www.w3.org/2006/05/addressing/wsdl"
      message="biroComm:transferDataResponseMessage"
      wsaw:Action="http://TransferService.communication.biro.eu/
                   BiroCommunicationServicePortType/transferDataResponse"/>
     </wsdl:operation></wsdl:portType>
     <wsdl:binding name="BiroCommunicationServiceSOAP11Binding"
      type="biroComm:BiroCommunicationServicePortType">
      <soap:binding transport="http://schemas.xmlsoap.org/soap/http" style="document"/>
     <wsdl:operation name="transferData">
     <soap:operation soapAction="urn:transferData" style="document"/>
     <wsdl:input><soap:body use="literal"/></wsdl:input>
     <wsdl:output><soap:body use="literal"/></wsdl:output></wsdl:operation></wsdl:binding>
     <wsdl:binding name="BiroCommunicationServiceSOAP12Binding"
      type="biroComm:BiroCommunicationServicePortType">
      <soap12:binding transport="http://schemas.xmlsoap.org/soap/http" style="document"/>
      <wsdl:operation name="transferData">
      <soap12:operation soapAction="urn:transferData" style="document"/>
       <wsdl:input><soap12:body use="literal"/></wsdl:input>
       <wsdl:output><soap12:body use="literal"/></wsdl:output>
      </wsdl:operation></wsdl:binding>
      <wsdl:service name="BiroCommunicationService">
      <wsdl:port name="BiroCommunicationServiceSOAP11port_http"
        binding="biroComm:BiroCommunicationServiceSOAP11Binding">
       <soap:address location="BiroCommunicationService"/>
      </wsdl:port>
       <wsdl:port name="BiroCommunicationServiceSOAP12port_http"
      binding="biroComm:BiroCommunicationServiceSOAP12Binding">
    <soap:address location="http://localhost:8081/axis2/services/BiroCommunicationService"/>
       </wsdl:port>
        <wsdl:port name="BiroCommunicationServiceSOAP12port_http"
       binding="biroComm:BiroCommunicationServiceSOAP12Binding">
    <soap12:address location="http://localhost:8081/axis2/services/BiroCommunicationService"/>
     </wsdl:port></wsdl:service></wsdl:definitions>
```

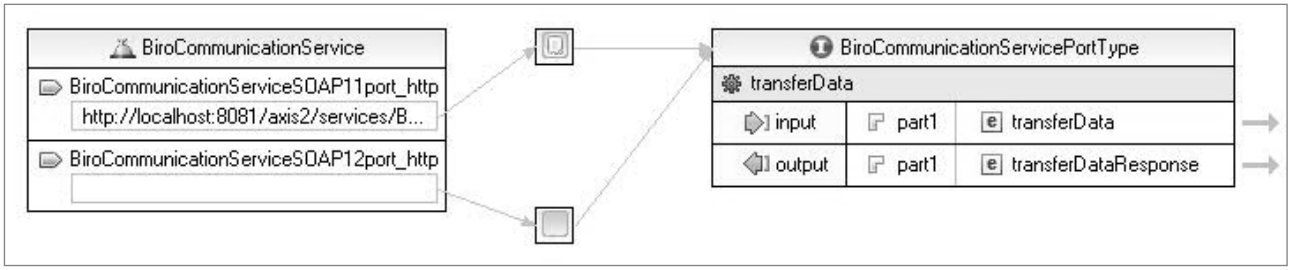


Figure 3.3.14: B.I.R.O. Communication Service and corresponding PortType

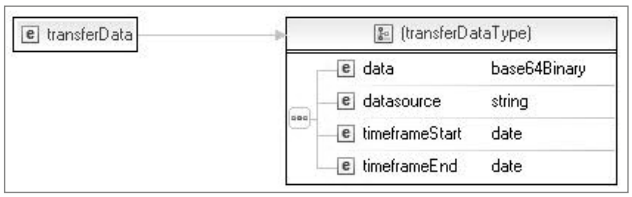


Figure 3.3.15: Request message of service "transferData" with corresponding datatypes to transfer.

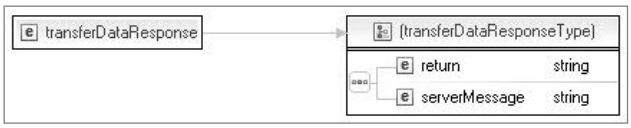


Figure 3.3.16: Response-message of service "transferDataResponse" with corresponding datatypes to transfer.

specific implementations for the web-service - for the service-provider as well as for the service-requester. Implementations for different programming-languages can be achieved by exchanging the templates used by the XSL-Transformation.

The complete process of the code-generation in AXIS2 can be seen in Figure 3.3.17.
The code-generation, using the previously explained WSDL-file, can be done via the command line with the following statement:

"%AXIS2_HOME%\bin\wsdl2java.bat -o java/adb/ -p eu.biro.communication.transferservice -d adb -t - ss -sd -g -uri wsdl/biroCommService.wsdl"

-o <output Location>	Output file location. This is where the files would be copied once the code generation is done. If this option is omitted the generated files would be copied to the working directory.
-p <package name>	The target package name. If omitted, a default package (formed using the target namespace of the WSDL) will be used.
-d <databinding>	Specifies the Databinding framework. Valid values are xmlbeans, adb, jibx, and none. Default is adb.
-t	Generates a test case. In the case of Java it would be a JUnit test case.
-ss	Generates server-side code (i.e. skeletons). Default is off.
-sd	Generates the service descriptor (i.e. server.xml). Default is off. Only valid with -ss, the server-side code-generation option.
-g	Generates all the classes. This option is valid only with the -ss (server side code generation) option. When on, the client code (stubs) will also be generated along with the skeleton.
-uri <Location of WSDL>	WSDL file location. This should point to a WSDL file in the local file system.

After this step the client- and service-stubs have to be filled with the appropriate business-logic.

Apache AXIS2 creates SOAP-messages during runtime using the concept of handlers, which can be used separately on global-, service and message-level, namely the scopes. The final handler chain is calculated combining the engaged handlers from all scopes. Rampart can be used as configurable and custom loadable handler on every scope. An illustration of the concept of handlers for the processing of messages can be seen in Figure 3.3.18.

AXIS2 can be configured via the mechanisms of an internal deployment model. This model consists of three entities for configuration, i.e. global-configuration, service-configuration and module-configuration. The global configuration provides information for the client and the server for global parameters (timeouts, attachments, message optimization, caching ...), transport receivers, transport senders and phases with engaged handlers within the *"axis2.xml"* xml-File. The service-configuration provides information for service level parameters, modules engaged on service-level, service specific message receivers and operations inside a service. The service-configuration is contained inside the *"service.xml"* file, which must be deployed within the service-archive.
The module-configuration is the configuration within the *"module.xml"* of the module-archive. It provides information for module parameters and operations defined in the module.

For *security*, it was decided to implement security-features not only for parts of the message, but for the whole content of a transferred message. That's why security-features for the BIRO project are configured on service-level, i.e. in the *"service.xml"*.
When configuring security on service level, the security module "Rampart" has to be engaged via the xml-

syntax <module ref="rampart"/> in the AXIS2-configuration. Two parameters, namely *"InflowSecurity"* and *"OutflowSecurity"*, are used to configure the handler, which provides the logic in the security phase.

The *"OutflowSecurity"*-parameter of the client has to comply with the *"InflowSecurity"*-parameter of the service, and vice versa.

For the BIRO project, at the current level of implementation JAVA-keystores (jks) are used to store and manage keys. The internal format of the keystore depends on the implementation of the used cryptographic provider. In general keystores from different crypto-providers are not compatible, but they all have a common architecture, i.e. they manage key-elements and trusted entries. Key-elements consist of certificate-chains and sensitive cryptographic information, i.e. secret or private keys. Trusted entries are entries of trusted parties. This means that an owner of a keystore can assume that the owner claiming to possess a certificate of trusted entry is actually the owner of the

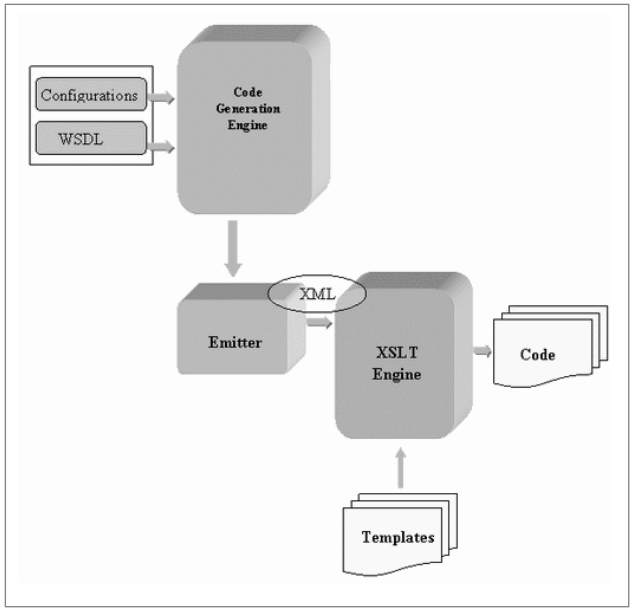


Figure 3.3.17: Process of Code-Generation within Apache AXIS2

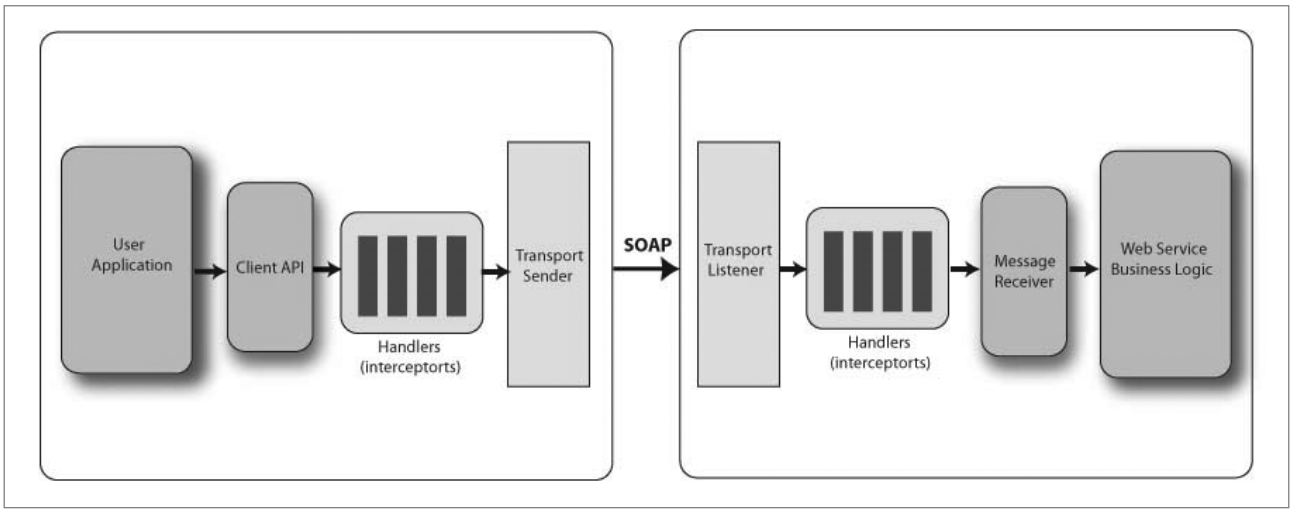


Figure 3.3.18: Processing model, showing the concept of handlers in Apache AXIS2.

certified public key. For BIRO, JAVA-keystores with self-signed certificates following the X.509 v.3-certificate-architecture were used.

For *Service Provider and Service Requester*, as BIRO tries to provide a solution for an infrastructure that can be used in disease-independent health-networks, every partner has to prove its authenticity.

Due to the fact that the BIRO system is a closed network within the internet, the maintainers of the BIRO system play the role of the certificate authority themselves. Users who want to join the system can create their own public-private keypair, export a certificate and deliver it to the central BIRO system. The central-server imports the local certificate and delivers its own certificate to the local BIRO system.

Therefore, every local BIRO system will have the central-server's certificate in its keystore, whereas the central-server will have every local-server's certificate in its keystore.

In its first implementation, the Communication-Software was designed as piece of software between two J2EE Application Servers, set up as the sender and the receiver respectively. As Application-Servers the light-weight Apache Tomcat Servlet Containers were chosen. The BIRO Local Engine was set up on one the first server which is in charge of providing a front-end for BIRO users to manage the upload of BIRO datasets.

In the front-end a user can upload a file which will be transmitted to the server by invoking the web-service asynchronously. On the server-side the transmitted files are stored in proper file-structure, which is understandable for the central-engine.

During the BIRO project, claims for an all-in-one solution providing functionality of all software-steps in the project's processes came up.
As a result, the server-to-server communication was replaced by packaging the communication software on the client-side into a library, i.e. a java-archive (jar), which can easily be integrated into the BIROBox. The

client library is invoked by the BIROBox and communicates with the central-server of the BIRO system. A configuration-file is needed to setup the communication-software with the proper parameters like service-endpoint (domain- or IP-location of the BIRO central engine), location of the client's AXIS2 configuration and the location of the client's keystore and corresponding properties.

The BIRO Box

The BIROBox has been fully developed in Java, with a straightforward architecture.
All main functions in the local BIRO System can be accessed by selecting the appropriate button of the button panel located at the left of the window. The bigger area on the right of the window is designed to contain the configuration panels (one or more than one according to the needs) for each function.

Each BIRO function (data extraction from local source, creation and population of local BIRO database, print of the statistical report, data sending to the central BIRO system) can be executed separate from others, and steps are triggered by the local user.

Alternatively, it is possible to run all tools at once automatically, although the step-by-step approach is preferable since it provides more liberty in the management of the local system. This way, one would import database XML files into the local BIRO database even if those files are produced by a different centre; or if a database is already in the BIRO format, one would choose to skip the execution of the Adaptor and Database Manager; in case the user is interested only in printing the local report (but not sending data to the central system) he/she would launch the Statistical Engine at any time without running the Communication Software.

3.3.4 Discussion

In terms of developing a sustainable information system, the BIRO Consortium focused its attention mainly on creating a solution that could be applicable in real life situations, where diabetes data are used and stored on a daily basis.

Clearly, much of the success of our initial trial was based on the applicability of the most visible and direct component: the BIRObox.

The BIROBox was initially tested on a sample dataset extracted from the Umbria register, then other databases from Cyprus, Malta and Romania (BIRO Technology Transfer Meeting, Bergen, Norway, 15th- 17th January 2009, see Chapter 4.1).
These tests highlighted many positive aspects of the BIRO software and were the occasion to collect suggestions regarding possible improvements on the software from the whole Consortium.
The fact that BIRO was delivered in the form of a single setup file was very appreciated: with a simple double click, the tool allowed to install everything produced by the system, including documentation.

The BIROBox also allowed to fill forms or choose options, as well as inspecting the local data source, or looking up a list of all local tables within the chosen database, and local fields within the selected table.

Test users noticed that they inevitably needed to know how to harmonise the local dataset with the BIRO standard, highlighting the fact that using the Box cannot circumvent the need for a detailed knowledge of the system.

A special attention was dedicated to the possibility for the BIROBox to connect with many different data sources. A survey has been conducted within the BIRO Consortium in order to know the technology in use for each local data source. All drivers for the most common DBMS have been added to the driver list of BIRO Adaptor.

Although most database drivers were included in the list of those usable for BIRO, the need for a "custom" driver was made clear, and consequently added to the box. However, such customization makes use from a non technical person unfeasible.

The alternative of use of CSV files has been very positively evaluated as a means of bypassing the problem of creating a customized driver, using this format as a *lingua franca* for the transport of files from any format to BIRO.

Finally, there was a good performance of the BIRO Adaptor in terms of execution time: about 5 minutes to transform nearly 100,000 patient's episodes from the local data source to the XML export. Obviously, the process duration showed to depend both on the number of episodes recorded
and the number of BIRO fields exported.

As suggested by the Consortium, the following aspects of the BIRObox would need to be revised in the future:

- popup help screens are not available for each form, especially those regarding data source properties and warning messages; it is important to prevent the user from inserting wrong entries;
- it is not possible to check the correctness of data mapping in real time: if the user makes a mistake when mapping local field to BIRO field (e.g. a wrong date format or a wrong unit of measurement is chosen), it will be noticed only when running Adaptor. BIROBox should immediately check mapping and send warnings to the user.
- Adaptor stops execution if a record contains wrong or unexpected data; modification is needed so that wrong values are simply discarded and each error is reported into an error log file. This could prevent multiple stops in case of poor quality datasets.
- all BIRO fields are hard coded within the Adaptor source code as java classes: this means that adding new fields to the BIRO Dataset would require modifying and recompiling the whole source code. In a future revision, software must allow more flexibility with respect to the changes on data dictionary.
- although for each numeric BIRO field the user can

select the unit of measurement used locally (within a predefined set) and map it automatically into the official BIRO unit, the user cannot add any other unit of measurement. This is because the mapping rules between units are hard coded. The list of units of measurement should be expanded, or let the user specify conversion algorithms.
- the Database Manager process requires long time to be completed because XML files are transformed into Java objects and stored into the local database one-by-one. Speeding up the process would be very important for the user.

Communication software represented another important testbed for the engineering of a robust network application.

The component was tested on datasets created by the BIROBox from the Romanian database, in its first and preliminary version, i.e. the server-to-server communication.

The tests with the datasets using full security enhancing mechanisms, i.e. encryption of the message including digital signature, highlighted that the amount of time needed for the invocation of the web-service is rather extensive.

In consideration of the rather low frequency of web-service invocation, this inconvenience of the system can be rather neglected. In contrast to this limitation, the problem of file size limitations is much severer. The AXIS2-framework causes a Java-Exception when digitally encrypting and signing SOAP-messages with attachments greater than 2 Megabytes. This limitation is caused by an OutOfMemoryError during reading the SOAP-attachment from a Buffer.

One way to avoid this exception can be to split up attachments greater than 2 MB into smaller pieces and to send them in different SOAP messages sequentially. Another possibility could be to just sign and encrypt the message and leave the attachment in base64-binary coding without encryption or signature. For future versions of the software, these limitations should be resolved and a new solution must provide best fit for BIRO datasets.

3.3.5 Conclusions

The common keyword that can be used to describe the BIRO development model is “integration”.

Through an intensive development phase, the proposed BIRO architectural model has been translated into a system that is able to extract different kinds of clinical data stored in diabetes registers from different sources and centres in different parts of Europe. Data are first conveniently transformed into a format that allows making standard comparisons, then aggregated, both at the local level and at the central level. Integration performed within the BIRO framework is not simply data collection extended to a broad number of partners, it is a fine aggregation of comparable data submitted to a sophisticated statistical analysis system transforming clinical records into usable information for policy and practice.

The BIRO system is the integrated result of multiple efforts described in different chapters: from the clinical review of diabetes indicators to the juridical review of European laws regarding privacy concerns, from the development of BIRO software tools to the production of the European report, everything is deeply interconnected.

From the technical point of view, many software tools, from existing ones to those developed ad hoc, have been merged in order to provide a complex product that is able to manage many tasks, from data extraction to the web publication.

Professionals with different backgrounds interacted to develop a European shared diabetes infrastructure in almost ideal terms, respecting their own views, perspectives, cultural backgrounds, and health care systems.

The result is a system that is not invasive and does not require changing much in data management as BIRO standards come directly from actual daily practice. In the end, the BIRO system is just an additional tool that we have made available, hopefully a good one that can be used to connect to peers for health improvement.

The system has been developed to respect the higher levels of privacy through the architecture based on privacy impact assessment, realized in detail, and security, through appropriate communication software.

The system integrates data and information: statistical reports are not closed within the Consortium but they can be made open to the public through the connection to the web portal.

BIRO software is free, open source, available to anyone interested that can easily download and test it.

The forthcoming EUBIROD project will represent a major challenge for the planned improvements and the extensive application of the system. Success in its future application will represent a major step for the BIRO architecture to become a key source of up-to-date information on diabetes on an international scale.

References

1. R software for statistical computing and graphics, available at: http://www.r-project.org/
2. LATEX typesetting system, available at: http://www.latex-project.org/
3. Information-security (2009, February 27), Wikipedia - the free encyclopaedia. available at: http://en.wikipedia.org/wiki/Information_security, retrieved 2009, February 20
4. Castor project, available at: http://www.castor.org/
5. Hibernate object/relational persistence and query service, available at: http://www.hibernate.org/
6. W3C Document Object Model (DOM) available at: http://www.w3.org/DOM/
7. Simple API for XML (SAX) at http://www.saxproject.org/
8. W3C Working Group Note on Web Services Architecture (2004, February 11). Retrieved 2009, February 27, available at: http://www.w3.org/TR/ws-arch/#whatis
9. WS-Security Core Specifications (2006, February 1), available at: http://docs.oasis-open.org/wss/v1.1/wss-v1.1-spec-os-SOAPMessageSecurity.pdf
10. W3C Recommendation on XML Signature Syntax and Processing (2008, June 10). Retrieved on March 01, available at: http://www.w3.org/TR/xmldsig-core/
11. W3C Recommendation on XML Encryption Syntax and Processing (2008, June 10). Retrieved on March 01, from http://www.w3.org/TR/ xmlenc-core/
12. Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile (2008, May), available at: http://tools.ietf.org/rfc/rfc5280.txt
13. AXIS2, Web Service Framework, Apache Software Foundation, available at: http://ws.apache.org/axis2
14. WSS4J, Apache WS-security Implementation. Apache Software Foundation, available at: http://ws.apache.org/wss4j
15. W3C Note on Web Service Description Language (2001, March 15), available at: http://www.w3.org/TR/wsdl

Distributed Statistical Analysis Software

Fabrizio Carinci, Luca Rossi

ABSTRACT

Introduction
Sustainable solutions for the routine provision of strategic data across Europe require highly collaborative frameworks. The BIRO approach is based on a collectively agreed application of two consecutive data processing steps, locally and centrally, each one involving key statistical procedures. A “Statistical Engine” is specifically required to derive aggregate tables from databases held at the regional level that will be sent towards a central BIRO server.

Objectives
To run the same specialised statistical software in each partner region exporting local data to a standardized database, formatted according to common criteria. To implement and disseminate the use of advanced statistical methods to collect and analyse population-based data by developing open source statistical software that will allow users replicating and further extending the approach.

Materials and Methods
The software has been adopted as a development platform for all BIRO statistical software. A statistical engine connects to the local database using R Postgres drivers. Through the notion of “statistical object”, tables are created to store aggregates of local data (e.g. the arithmetic mean, percentile, variance, etc.) as flat text comma delimited files. A taxonomy defines all objects being implemented. The BIRO template has been used as a guide for data processing and consequent transfer to the central server, where the central statistical component runs the overall analysis for the delivery of the global report.

Results
The BIRO statistical and central engines have been successfully developed and tested on both Vista and Linux. Average hardware allowed completing a full local BIRO report from a test sample of more than 5,000 patients and several thousands episodes in about 7 minutes. The central engine, using aggregate data from N=5 centres, corresponding to over 43,000 subjects and 273,000 episodes, completed the entire process of statistical analysis and production of a full overall report in 22 minutes. Installation of the software is identical regardless of the hardware, requiring R>1.8, Latex, Java 6.0 and PostgreSQL, plus various additional libraries/packages included in the distribution packages. All R functions are released under the GPL license and made available to partners of the Consortium bundled with all other BIRO components.

Discussion
The statistical engine provides a platform for accurate benchmarking that currently does not exist in its innovative form at the point of health care provision. The system may serve multiple users, from the European Union, to the local physician. The engine may improve the validity and completeness of information available: existing registers may be optimised on the basis of common standards, and new ones can be created with a fostered structure. The system should represent a component of a progressive approach, through which statistical functions can be constantly improved. Users, once inducted to using the software, can apply it independently and submit better aggregate data to the central server, safeguarding sensitive data as a result of the application of rigorous rules set by the BIRO privacy impact assessment.

Conclusions
The application of the BIRO distributed statistical analysis framework can help evaluating clinical practice more rapidly and effectively both within and across regions. Prevention strategies and health services may be planned more carefully on the basis of factual information, empowering clinicians with more accurate and structured information. The free availability of a modern statistical component can help disseminating the BIRO approach across Europe.

3.4.1 Introduction

Performance evaluation has become a fundamental step for continuous quality improvement in modern health systems[1].

The current pressure on cost containment has increased the availability of administrative data, now more standardized and widely available than ever before. Worldwide, there is an abundance of technical specifications that can be shared to organize powerful health data warehouses, through which complex analyses can be carried out quite rapidly and effectively.

In many regions, linked databases are routinely organized, including a client master index, hospital discharges, pharmaceutical prescriptions, specialist services, pathology tests, mortality register, etc. Through these new goldmines of health information, the estimation of routine outcome indicators has become more common, complete and increasingly reliable.

However, a methodological question remains about the possible interoperability of the different approaches and the comparability of results that are routinely published by different institutions.

Benchmarking health systems, providing evidence of better performance, informing stakeholders and consumers with up-to-date and validated information, they have all become common keywords in policy making. Nevertheless, it is still very difficult to provide international estimates that can be continuously reported and properly classified.
Many challenges must be faced to make substantial progress in that direction.

There is a need for improving our ability to maintain and process massive linked datasets, taking into account the reliability of validated unique identifiers, which however cannot be exchanged over the network and analysed collaboratively for obvious issues of privacy protection.

There is a need for developing scientifically robust methods and statistical procedures that can be widely agreed, and possibly shared and used across Europe. There are indeed positive examples that can be used as starting points.

A breakthrough in outcomes evaluation came with the detailed publication of standardized methods for risk adjustment by the US Agency AHRQ[2]. Briefly, this method allows to compute a different equation for each outcome indicator, based upon a national (or international) standard, so that crude estimates will be risk-adjusted, and standardized rates can be obtained at the centre, region, or state level.

Such general equation is first produced at the global level, by pooling large samples out of state datasets for each particular outcome. The main statistical method used to derive such formula is logistic regression, which is ideal in the case of binary outcomes, as is the case of most outcome indicators.

Parameters are then passed back to individual states, enabling them to produce own standardized estimates independently, by multiplying the overall national rate by a “post-adjustment factor” that is equal to the ratio of observed over expected rate for that particular region. Confidence intervals can be extracted using algorithms that are fully published by the AHRQ, along with all software and specifications applied to produce a report (adjusting factors, coding, etc).
The same method has been recently applied in Europe[3] on a large regional data warehouse, for which over 130 indicators directly chosen by health professionals were automatically estimated by a “System for the Evaluation of Outcomes”.

The system carries on previous work[4] conducted by F.Carinci in the field of linked datasets and automatic statistical analysis, including the extensive development of linked diabetes registers as a means to derive accurate and sustainable information for policy[5].

From such work it was immediately clear that computing estimates from very large databases pooled out from different regions was both technically difficult and overtly controversial for political and organizational reasons. Different regions in Europe apply different standards, they use different database systems, and most importantly, they would probably never allow to share electronic medical records in their entirety, as statistical procedures commonly request.
A novel method based on a distributed mechanism was sketched out to overcome these problems.

Early experiments carried out in different Italian regions[6] and in the framework of a federal system at Monash University[7] paved the way for a different approach, based upon the fragmentation of databases across different units (clusters).

This solution exploits a natural feature of health datasets: data is dispersed across different levels, among multiple providers, governed by different organizations. It would be then easier not to change what appears normally to be a hierarchical structure (local health authorities, hospitals, wards, etc).

This way different sources of variations can be embedded in database management, and multilevel statistical methods, that are becoming increasingly popular in health services research, can be applied by design.

The experiences made with RISS and H+ software provided some initial scope for the substantial enhancement planned for the BIRO project.

Two initial pitfalls had to be overcome.

Firstly, software was initially based on the SAS[7] language, a powerful statistical system that is frequently used by regional governments in the management of large data warehouses. The problem with this system is that it is excessively expensive and not open source, making it difficult to distribute it across providers, and not easy to customise at the desired level of detail.

Therefore, open source tools were selected as the platform of choice for a new prototype[9].
Secondly, the target for the development of a statistical engine had to be more focused and policy oriented. This opportunity was found in the estimation of risk adjusted outcome indicators, a primary goal in performance evaluation that usually involves many diabetes indicators. In this case the method developed by the AHRQ provided an important endpoint for the BIRO project.

Diabetes represents an almost ideal framework to envisage statistical modeling in a distributed fashion, as it would be done for a meta-analysis of observational studies[10].

Predictive models in diabetes are widely used to take into account the association between a set of characteristics (risk factors) and outcomes defined for each target indicator, e.g. HbA1c.

A restricted set of statistically significant risk factors are usually selected among a list of case mix variables that can potentially confound the relationship between individual characteristics and the outcome of interest. Although these variables are usually subject level, they can also include structural, contextual, and population based factors[11].

Statistical models are used to carry out estimates of rates that are "adjusted" for possible imbalances in the composition of risk factors for the particular centre, region, or entire country. This way standardized rates can be obtained for comparison purposes.

A simple example may simply show how this process has been envisaged from initial experiences, taking the rate of high Hba1c as an intermediate outcome indicator of poor metabolic control in diabetes.

The SAS source code presented in Box 3.4.1 produces

Box 3.4.1 Source code for logistic model

```
/* Variables of diabetes dataset SAS "_MODEL_"                    */
/* hi_hba:  High level of Glycated Haemoglobin                    */
/* gender:  gender=males      (r.c.=females)                       */
/* cl_age2: age class=2      (r.c.=class 1)                       */
/* cl_age3: age class=3      (r.c.=class 1)                       */
/* cl_age4: age class=3      (r.c.=class 1)                       */

/* 1. Logistic model on total sample                              */
/* Analysis of the association between gender, age and high Hba1c */

proc logistic descending data=_model_;
model hi_hba=gender cl_age2 cl_age3 cl_age4
        / clparm=wald rl rsquare ctable lackfit pprob=.5 aggregate scale=1;
run;

/* Applying the resulting logistic equation (see box 3.4.2)       */
/* on each patient seen from a participating centre, and summing over */
/* all patients seen by that centre, it can be used to predict the average */
/* rate of patients with high Hba1c (diabetes indicator)          */

/* Using aggregate tables (counts) instead of subjects, the logistic equation */
/* works on smaller samples composed of all observed patterns of covariates */

proc sort data=_model_ out=_model_;
by idcentro gender cl_age2 cl_age3 cl_age4;
run;

proc freq data=_model_ noprint;
tables hi_hba / out=test (drop=percent);
by idcentro gender cl_age2 cl_age3 cl_age4;
run;

/* Using either model is absolutely indifferent in terms of final results */
/* meaning that centres can avoid sending individual data         */

/* The result is due to the mathematical properties of logistic regression */
/* 2. Applying the same logistic procedure only on aggregate counts from all */
/* centres                                                        */

/* <....more SAS code to append frequencies from all centres>     */

proc logistic descending data=in_sedis;
model hi_hba=gender cl_age2 cl_age3 cl_age4
        / clparm=wald rl rsquare ctable lackfit pprob=.5 aggregate scale=1;
weight count; /* estimate is weighted by frequencies */
run;

/* Results of 1., 2. are summarized in Box 3.4.3-4                */
```

Box 3.4.2. Output Logistic Model on all observations

The LOGISTIC Procedure
Model Information

Data Set	WORK._MODEL_
Response Variable	HI_HBA
Number of Response Levels	2
Number of Observations	17102
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	HI_HBA	Total Frequency
1	1	4856
2	0	12246

Probability modeled is HI_HBA=1.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.6862	0.1028	44.5243	<.0001
GENDER	1	-0.2297	0.0343	44.7555	<.0001
CL_AGE2	1	0.0916	0.1092	0.7027	0.4019
CL_AGE3	1	-0.1465	0.1040	1.9842	0.1589
CL_AGE4	1	-0.2491	0.1086	5.2637	0.0218

Box 3.4.3. Output Logistic Model on aggregate data

The LOGISTIC Procedure
Model Information

Data Set	WORK.IN_SEDIS
Response Variable	HI_HBA
Number of Response Levels	2
Number of Observations	16
Weight Variable	COUNT
Sum of Weights	17102
Model	binary logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	HI_HBA	Total Weight	Total Frequency
1	1	8	4856.000
2	0	8	12246.000

Probability modeled is HI_HBA=1.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.6862	0.1028	44.5243	<.0001
GENDER	1	0.2297	0.0343	44.7555	<.0001
CL_AGE2	1	0.0916	0.1092	0.7027	0.4019
CL_AGE3	1	-0.1465	0.1040	1.9842	0.1589
CL_AGE4	1	-0.2491	0.1086	5.2637	0.0218

Box 3.4.4. Observed/expected rates by centre using logistic regression

Centre	Den.	Num.	%Observed	% Expected	95% Lower	95% Upper
1	7699	2189	28.4	28.5	27.5	29.5
2	2360	1000	42.4	28.0	26.1	29.8
3	3422	916	26.8	28.4	26.9	29.9
4	1239	222	17.9	28.3	25.8	30.8
5	2382	529	22.2	28.4	26.6	30.2

two logistic equations leading to exactly identical results, due to the particular properties of logistic regression. Among the two options, only the first is based on individual data, while the second is produced on top of frequency counts from individual centres, as shown in Box 3.4.2 (individual data) and Box 3.4.3 (aggregate data).

Therefore, individual data are not strictly necessary to risk-adjust diabetes indicators by using logistic regression. This property can be exploited to enhance performance and facilitate data management.

According to the method used by the AHRQ, the logistic equation can be separately used to compute expected rates for each data source. In this case, a score is obtained for each patient (or group), through the application of the logistic equation to each distinct pattern of characteristics (weighted by the frequency of that group). The sum of the scores per centre is then used to estimate expected rates (Box 3.4.4), from which standardized rates can be obtained.

Such a simple case exemplifies the main motivation for the design of the BIRO project: through a well defined set of aggregate tables, it is possible to feed carefully standardized procedures leading to comparable results from disjoint samples, very rapidly and effectively.

However, logistic regression is a special case that does not apply to other problems in health care applications. For instance, it may be difficult to use fragmented data to compute time-dependent models (as in the case of survival analysis / proportional hazards regression). In that case, it would be necessary to collect very fine aggregate tables, or an anonymous sample of individual data (which would raise concerns on the transfer of sensitive data).

Alternatively, separate equations can be produced by different registers, pooled together by using some algorithm of weighted meta-analysis, as in the case of inverse-variance weighted average of risk parameters[12]. In this last case, obviously, there might be a problem of approximation and information loss that must be taken duly into account.

The plan for BIRO was to develop a main architecture that would progressively allow to further advance methods to deal with fragmented data, starting from the calculation of crude rates and the publication of a basic descriptive report.

The BIRO project envisages a system where partial results contribute to an overarching framework collecting aggregate data from different regions. The main responsibility of the statistical engine should be to deliver tables derived from each individual database at the

local level, e.g. a clinical site or regional register.

A central server is in charge of running "twin" statistical routines that process all aggregate tables to produce overall estimates of diabetes indicators.
As a result of this design, two specific work packages have been allocated by the BIRO project to develop statistical routines.

The plan was to reconstruct a novel, completely restructured Intranet Statistical System (ISS)[13], to harmonise distributed data management with the routine publication of results at the European level. Practical objectives were defined at the outset for these work packages.

The primary objective of the "*Statistical Engine*" is to run specialised, standardized software in each participating region on top of a well formatted database, based upon common definitions included in the BIRO concept/data dictionary.

Outputs must be standardized, based upon specifications provided by the reports template, whose scope is to present information according to an agreed standard for the local register and the European collaboration.

The exchange of aggregate data is at the basis of this process.

Secondary objectives of the statistical engine include:

- implementation and dissemination of modern database techniques and advanced statistical methods to collect and analyse population-based data stored in diabetes registries. Statistical models must include mainstream methods for the standardization of the diabetic population, including multivariate models, e.g. Logistic Regression, GEEs, Multilevel models, taking into account different sources of variation. Whenever appropriate, a meta-analytical approach is used to bypass data transfer of excessive micro data across countries.
- creation of a fully documented repository of open source statistical software that will allow user to replicate and further extend the application of specialised software.

The objective of the "Central Engine" is to join all aggregate tables collected from several data sources and to run specialised, standardised software for all participating regions, based upon common definitions included in the BIRO concept/data dictionary, to produce the overall BIRO report.

3.4.2 Materials and methods

The design of the statistical engine revolves around the application of R[14] software, playing a central role in the whole process of data loading, transformation and analysis (Figure 3.4.1)

The application of R is triggered by the user, either through a script command file or with the aid of a GUI interface. R software connects to the local database using proper Postgres drivers.

According to the specifications given by the *report template,* and the associated definitions of statistical

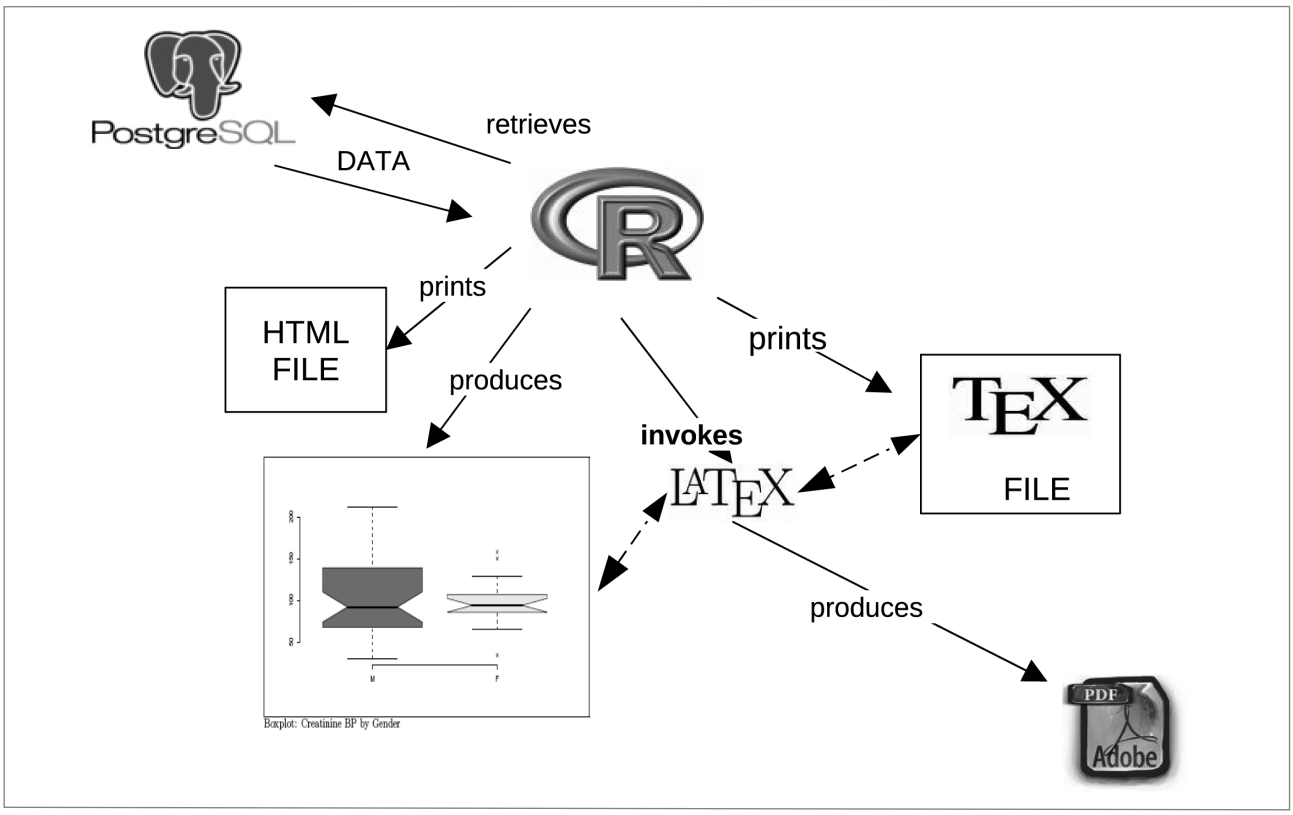


Figure 3.4.1 BIRO Statistical Engine Design

Table 3.4.1. Meta-data of BIRO Statistical Objects

Code	Sequential code based on the taxonomy of the statistical objects dictionary
Statistical Object	Name of the statistical object
Description	Short description of the principal content and output of the statistical object, and the main properties
Variables	Type of variables (categorical, continuous)
Properties	Mathematical and statistical properties in a distributed data environment
Local Component	OUTPUT OF THE LOCAL STATISTICAL ENGINE Technical characteristics of the statistical object that is produced from each data repository, to be sent to the central engine. Data section includes details on the format of the csv output.
Cumulative Component	CUMULATIVE DATASET PROCESSED BY CENTRAL ENGINE Technical characteristics of the procedure implemented in the central engine to produce the statistical object for the overall sample of connected repositories Data section includes details on the format of the csv output.
Output	STATISTICAL OUTPUT OF THE CENTRAL ENGINE Includes the list of components that will be computed and stored in the statistical object (ex: n, relative risks + confidence intervals, graph elements). Data section includes details on the format of the csv output. Defined codes are attributed to the list of electronic elements (e.g. XML tags, or csv tables)

Table 3.4.2. Statistical Objects Dictionary

SECTION 1. FREQUENCY TABLES

1.1 Univariate Frequency Distribution
1.2 Outliers
1.3 Contingency Table

SECTION 2. MEASURES OF LOCATION

2.1 Arithmetic Mean
2.2 Percentile
2.3 Range

SECTION 3. MEASURES OF DISPERSION

3.1 Variance
3.2 Interquartile Distance

SECTION 4. GRAPHICAL ELEMENTS

4.1 Bar plot
4.2 Histogram
4.3 Partial boxplot
4.4 Overall boxplot
4.5 Line plots
4.6 XY Plots
4.7 Webplot
4.8 Maps
4.9 Forest plot

SECTION 5. REGRESSION

5.1 Linear regression
5.2 Logistic regression
5.3 Meta-analysis

SECTION 6. STANDARDIZATION

6.1 Standardized rate
6.2 O-E

targets, the R functions being developed process the Postgres database to deliver results in the form of small CSV datasets. Such datasets are further processed to produce individual centre outputs and full local reports in the form of pdf and html files, by using different graphical drivers and the high quality typographical software Latex.

A compressed CSV folder is created to deliver all tables produced by each local run of the statistical engine, to be stored in a directory that is properly named with the current datetime and centre id, ready to be transmitted as a compressed file to the central server.

The statistical engine consists basically into a set of functions specifically designed to create and manipulate "*statistical objects"* according to an original definition provided by the BIRO approach.

A statistical object is defined as "*an element of a distributed information system that carries essential data in the form of embedded, partial aggregate components, required to compute a summary measure or relevant parameter for the whole population from multiple sites*".

The definition of statistical objects is central to the functioning of BIRO, as it allows using pre-determined datasets as basic elements of a statistical analysis ran on top of aggregate data to produce individual centre reports. Such partial results are then transmitted over the network for the production of global reports. This solution allows bypassing many possible risks and

Table 3.4.3 GLOSSARY OF OUTPUT COLUMNS

Variable Name	Description
x	Variable X
y	Variable Y
n	number of non missing values for each cell
sum_x	
n_x	is the total number of non missing values for variable x

Table 3.4.4 Example of BIRO Statistical Object: Percentile

Code	2.2
Statistical Object	Percentile (Measures of Location, e.g. Median=50%)
Description	Value that includes the desired percent (Median=Central) of observations in the weighted ordered list of a target variable. If the desired percentile lies between two values, then the percentile is equivalent to the arithmetic mean of the two adjacent values.
Variables	CONTINUOUS
Properties	The percentile of the overall sample is obtained from the complete ordered list, including n from all levels of the target variable in each local object
Local Component	Data vector composed of two quantities: value for each level of the target variable; total number of observations in the specific level DATA: <2.2.a>id, date, stratum, x, n
Cumulative Component	Sum of all ordered lists from each local object DATA: <2.2.a>id, date, stratum, x, n
Output	Single parameter value that includes the desired percent (Median=Central) of observations in the weighted ordered list of the target variable, obtained as a sum of all ordered lists from each local object DATA: <2.2.a>date, stratum, pcl_x Single parameter value by centre DATA: <2.2.b>id, date, stratum, pcl_x

restrictions imposed by the privacy legislation, as defined by the best architecture, avoiding to exchange individual records.

Basically, statistical objects are tables that contain statistical aggregations of local data (arithmetic mean, percentile, variance, linear and logistic regression, bar plot data, histogram data, box pot data, etc), stored as flat text comma delimited files (CSV).

Metadata for statistical objects are shown in Table 3.4.1. Statistical objects are organized according to a dictionary (Table 3.4.2) including as basic components of frequency tables, measures of location, measures of dispersion, graphical elements, regression, and standardization.Criteria agreed by the Delphi panel for the definition of the best architecture have been duly taken into account in the specifications of statistical objects. Prior to developing all statistical objects,

Table 3.4.5 Example of BIRO Statistical Object: Variance

Code	3.1
Statistical Object	Variance (Measure of Dispersion)
Description	Sum of squared deviations from the mean divided by the total number of observations minus one. Can be interpreted as the average squared distance from the mean.
Variables	CONTINUOUS
Properties	The overall variance is equal to the sum of two components: the variance "within" data repositories, expressed as the weighted average of the variances in each data repository, and the variance "between", expressed as the weighted average difference between the mean at each data repository and the overall mean. Formula for the overall variance: (weighted.mean(var,n)*(sum_n-length(unique(id))) +sum(((mean-weighted.mean(mean,n))^2)%*%n))/ (length(x)-1)
Local Component	List of values of the arithmetic mean, variance, total number of observations for each stratum of interest DATA: <3.1.a>id, date, stratum, mean, var, n
Cumulative Component	Appended list of values of the arithmetic mean, variance, total number of observations for each stratum of interest for all data repositories DATA: <3.1.a>id, date, stratum, mean, var, n
Output	Variance parameter and number of observations for each stratum DATA: <3.1.a>date, stratum, var, sum_n Variance by local data repository DATA: <3.1.a>date, stratum, var, sum_n

individual tables have been defined to provide descriptions, using a common glossary for each of the output columns (Table 3.4.3).

Examples of statistical objects for percentiles (Table 3.4.4), variance (Table 3.4.5), and overall boxplot (Table 3.4.6) show the contents of the objects delivered by the local statistical engine to the report output (local component) and the central server (cumulative component).

Table 3.4.6 Example of BIRO Statistical Object: Overall Boxplot

Code	4.4
Statistical Object	Overall Boxplot
Description	Simultaneous graphical representation of measures of location and dispersion to represent the statistical distribution of a continuous variable. The graph includes the mean, median, interquartile range, two derived measures of deviation from the centre of the distribution, defined as "wiskers", and extremely deviant observations, also known as "outliers". Wiskers are calculated using the following formulas: Upper wisker = 75% percentile + 1.5 (interquartile range) Lower wisker = 25% percentile – 1.5 (interquartile range) Outliers are presented as values of a target variable above and/or below wiskers.
Variables	CONTINUOUS
Properties	Overall boxplot is computed by appending individual frequency distributions, summing up all frequencies for the union of levels observed, and computing the graphical representation from the weighted cumulative distribution.
Local Component	Data matrix including all non-zero frequencies for each level of a target variable. Optionally the original variable can be rounded based on a desired interval (e.g. the one determined by privacy protection rules) Ex: 52=48=50 (>45,< 55) ~ 50; round to the nearest integer; day ~ month DATA: <4.4.a>id, date, stratum, x, n
Cumulative Component	Overall boxplot obtained from weighted cumulative distribution DATA: <4.4.a> date,stratum, mean, median, pcl_25x,pcl_75x, l_wisk, u_wisk, outlie_x
Output	<4.4.a> PNG, GIF, JPG files

3.4.3 Results

The statistical engine has been successfully developed with the following structure (pseudo-code):

```
_____________________________________________
Start
1. Setup environment
2. Compute Indicator Statistics
 For each indicator in the Report Template:
 Loop Start
    Reference Indicator
    IF i-th statistical procedure is TRUE then
 Apply Statistical Procedure
 Output production
       END
Loop End
3. Compile results
End
_____________________________________________
```

The loop is presented as a flow chart in Figure 3.4.2.

The first step relates to the definition of the workspace, data preparation, and output formatting (Box3.4.5). Execution starts with a fresh setup of the complete environment, including a check of the local OS version, any required installation of additional R packages, and the definition of global variables.
The BIRO database is formatted by applying definitions in the data dictionary: new variables are created using a predefined set of cutoffs, new tables are created by merging and linking the original datasets into a new format amenable to statistical analysis. Finally, html and tex (pdf) outputs are initialized and formatted where required.

A second step is required to compute all indicator statistics (Box 3.4.6).

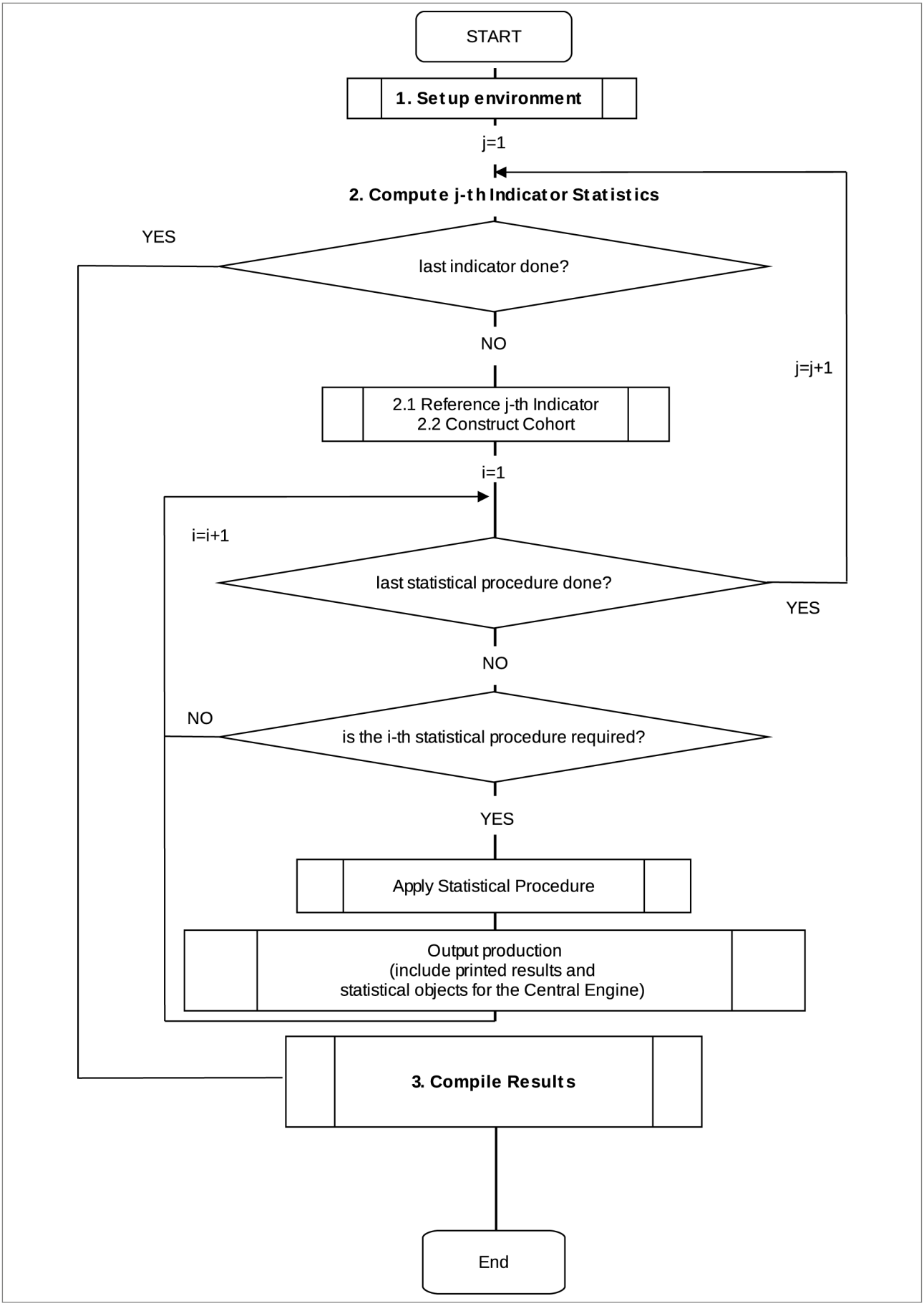


Figure 3.4.2 Statistical Engine Flow Chart

Box 3.4.5. Components and Files of the Statistical Engine (1): "Setup Environment"

BIRO_se_.r

BIRO_se_setup.r
• Clean active Workspace
• Load Environment Parameters (db drivers, directories, R libraries)
• Check for the existence of SEAfs directories, create missing ones
• Load Utilities
  o BIRO_aggregate.r
  o BIRO_demographics.r
• Load BIRO R libraries

BIRO_se_.r

BIRO_se_datastep.r
• Check for the existence of stored R data frames (Merge Table)
  o Connect to the BIRO database, transform it into R data frames and Merge Table, save transformations into .csv files
  o Load BIRO R data frames from transformed .csv files
• Apply thresholds, limits, levels for categorical variables from BIRO R libraries
• Apply date parameters stored in BIRO R libraries

BIRO_se_.r

BIRO_se_report.r
• Pre-production of Tex File
  o Open Tex file, write report cover page including authors, logo etc.

BIRO_se_.r

Box 3.4.6. Components and Files of the Statistical Engine (2): "Compute Indicator Statistics"

BIRO_se_.r

For each indicator in the Report Template:

Loop Start

  Reference indicator (read relevant parameters)

  BIRO_se_report.r
  o Open Tex Indicator Section, write section cover page

  BIRO_se_indicator_<section>.r
  o Construct indicator data frame (valid cohort)
  o Apply relevant indicator parameters to statistical procedure call

IF a possible statistical procedure among:

  o Measures of location (BIRO_se_location.r)
  o Measures of dispersion (BIRO_se_dispersion.r)
  o Contingency Tables (BIRO_se_tables.r)
  o Histograms (BIRO_se_histograms.r)
  o Boxplots (BIRO_se_boxplots.r)
  o Historical Trend (BIRO_se_trend.r)
  o Forest plot (BIRO_se_forest.r)
  o Trellis (BIRO_se_trellis.r)
  o Webplots (BIRO_se_webplots.r)
  o Maps (BIRO_se_maps.r)
  o Regression (BIRO_se_regression.r)
  o Standardization (BIRO_se_standardize.r)

IS TRUE then

  Call and apply i-th Statistical Procedure
  Output Production (see relevant .r files above)
  • Save i-th set of produced statistical object as .csv
  • Save i-th set of statistical objects as .csv
  • Save i-th set of statistical tables as .html
  • Save i-th set of produced graphs as .png

End

Loop End

The complete list of BIRO indicators is read from the report template, along with definitions included in the data dictionary. An indicator “cohort” is automatically constructed, based upon the agreed specifications relative to the particular category of patients that must be included in each indicator. Appropriate database and statistical procedures are executed to reproduce algorithms foreseen for each indicator, until the complete list of tasks is finalised and the set of planned outputs is entirely produced.

The loop ends when the complete list of indicators in the BIRO report template is produced (Box 3.4.7).

All results are compiled into an overall report that is produced in pdf and html format for the local centre site, including output files that include raw data, text listings (individual html tables) and graphical outputs. Results are stored in a directory with a unique timestamp, whose content is sent towards the central server by invoking a BIRO routine, where they are used by the central engine to produce European results from a part or all BIRO participating centres.

The complete list of functions specifically created to realise the statistical engine, along with their location in storage files, is presented in detail in the BIRO report “Statistical Engine”.

Central Engine

The *central engine* is designed to operate after all elements produced by the statistical engine have been transmitted to the central server.

Box 3.4.7. Components and Files of the Statistical Engine (3): “Compile Results”

BIRO_se_.r

BIRO_se_report.r
- Post-production of Tex file
  - close tex file and execute Latex to create report .pdf file

BIRO_se_.r

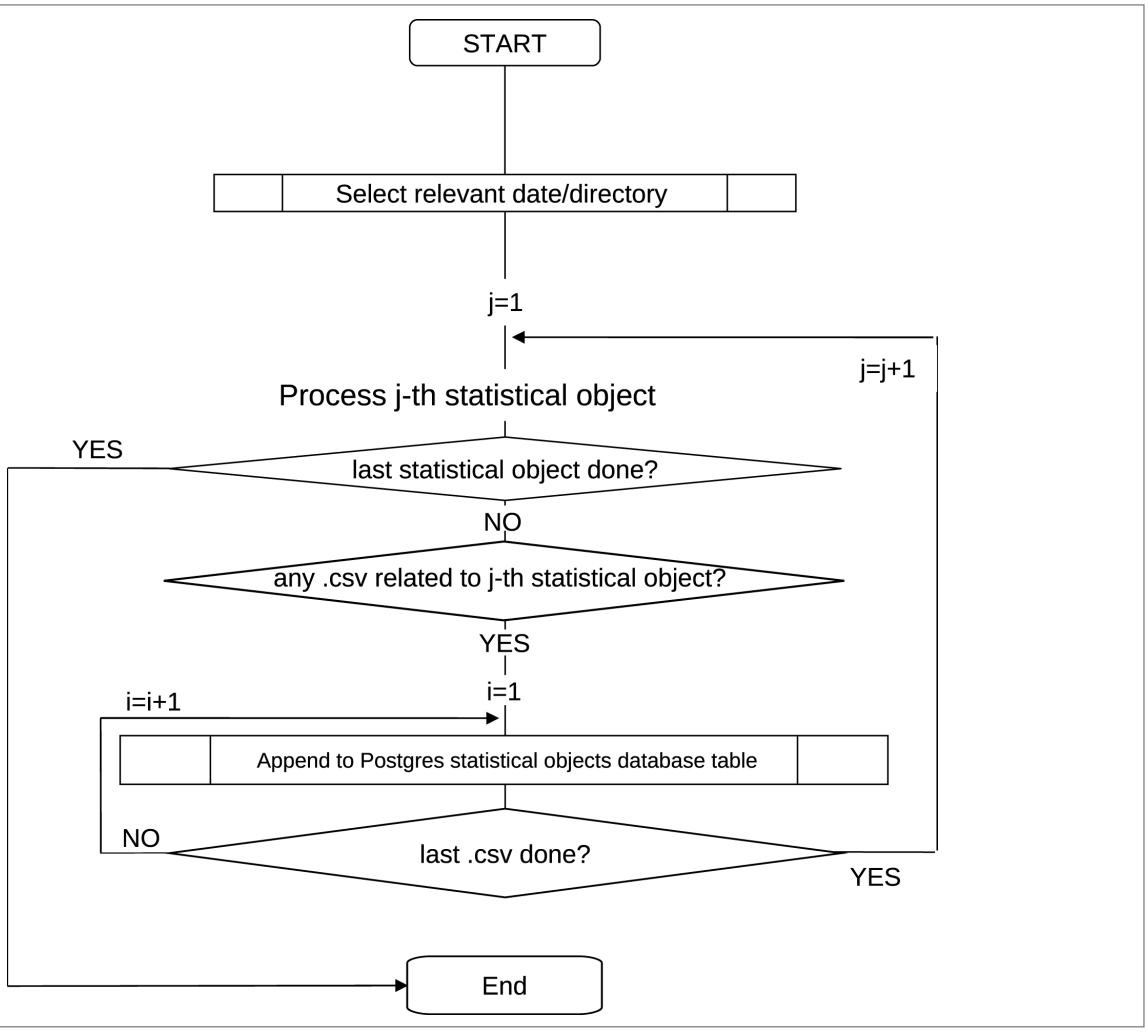


Figure 3.4.3. Pile-up database Flow Chart

Here we describe the statistical components of the central engine only in general: more details are provided in the specific BIRO report “Central Engine”.

The first part of the engine is dedicated to the creation of an overall database from multiple aggregate tables, also known as the “Pile-Up Database” (Figure 3.4.3).

A compressed directory of “partial results” in .csv format is uploaded by each participating centre, after the application of the statistical engine. The central engine processes a predefined list of statistical objects required for the production of a specific indicator, checking for their presence in the “partial” directory.

The engine appends each object to the specific table formed by all the same statistical objects that have been transferred by BIRO centres for a particular reference time interval. The database component of the central engine is invoked through a Java function that has been specifically developed to load .csv objects in a PostgreSQL database (CSV “Importer”). The loop is completed once all objects are allocated to a PostgreSQL table. The key component of the central engine operates once all tables from all centres have been allocated (Figure 3.4.4).

Basically, a set of routines that is “twin” to those in the statistical engine have been developed in a loop almost identical to the partial one. The major difference is that statistical procedures of the central engine operate on top of the PostgreSQL set of aggregate tables, based on slightly different algorithms running on top of the “cumulative component” (Table 3.4.1).

Outputs produced by the central engine are mostly identical both in format and content to those delivered by the statistical engine, with the exception of variations among regions that usually does not appear in local reports.

Geographical mapping may be required to identify clusters of observations highlighting abnormal values for target indicators in diabetes. Maps can be created by categorizing areas according to a limited number of classes, either by using simple statistical measures e.g. percentiles, or advanced methods e.g. cluster analysis, regression trees etc, or by directly specifying cut-offs for the definition of classes.

Geographical areas are normally represented by closed polygons linking bi-dimensional points whose coordinates are stored as latitude and longitude in geographical libraries that in many cases are freely available. Polygons may refer to entities of different size and nature e.g. cities, provinces, states, or entire continents. A fundamental problem hampering the uniform geographical representation of epidemiological data across Europe is the heterogeneity of such areas across different states. To map different definitions to a uniform representation, a specific taxonomy is required.

The European Union has developed the Nomenclature of Territorial Units for Statistics (NUTS)[15] as a geocode standard for referencing the administrative divisions of countries for statistical purposes. A NUTS code begins with a two-letter code referencing the country and is available beyond the EU, with a two-letter code for a continent, two numbers for the country, and for the USA, Canada and Australia the states, provinces, and territories, separate numbers.

NUTS regions are based on the existing national administrative subdivisions. In countries where only one or two regional subdivisions exist, or where the size of existing subdivisions is too small, a second and/or third level is created. This may be on the first level (ex. France, Italy, Greece, and Spain), on the second (ex. Germany) and/or third level (ex. Belgium).

In smaller countries, where the entire country would be placed on the NUTS 2 or even NUTS 3 level (ex. Luxembourg, Cyprus, Ireland), levels 1, 2 and/or 3 are identical to the level above and/or to the entire country. Indicative thresholds are 3-7 millions for NUTS 1, 800,000-3 millions for NUTS2, and 150,000-800-000 for NUTS 3.

In BIRO, we recognize that some extra levels may be worth to be included for the specific organizational levels of health systems, and added them to the NUTS classification. In setting up the BIRO database, each centre is asked to specify a relevant classification of geographical information that is either included as a reference to the place of residency of the patient (patient dataset), or the location of the centre (data source dataset).

In BIRO, a total maximum number of 8 nested levels is generally considered for the purpose of recording and mapping geographical information. Each country must supply a transcoding table, to link across all codes. If a variable does not exist in one country, the coding of the first non missing variable at the higher level is applied.

BIRO geographical levels include definitions (along with an example for Italy) shown in Table 3.4.7.

Each class can be linked to a particular class of shapefiles, depending upon the level available. It is possible that a standard shapefile is not available for the specific level of detail: open repositories offer libraries e.g. the ESRI “admin98”, which only provides NUTS 3 levels for some countries. Few countries may have maps available at the BIRO-2, BIRO-3 levels.

Mapping is carried out in BIRO by processing geographical information stored for the patient or specific clinical unit. To simplify the process, BIRO considers only two variables for the scope, i.e. one to be available for the patient, and another for the centre.

These variables do not share necessarily the same level of detail: patient references may be available in terms of postcode, while provinces may be used at the level of centre location.

The centre data descriptor must clearly indicate which of the 8 variables are used.

The statistical engine processes information stored in the database engine, producing aggregates by groups of patients and/or centre location.

The taxonomy is needed for two different reasons.

The first relates specifically to the statistical engine (local processing): the aggregate table resulting from the statistical analysis (see statistical object 4.8) is merged to the regional taxonomy table through the appropriate data source descriptor, and the variable corresponding to the level available for the polygon ID in the target shapefile (which of course must be same detail or coarser) is used for mapping. If necessary, the level must be rescaled (see Figure 3.4.5).

In the central engine (global processing) the main problem arises when maps from different regions/countries must be produced, with heterogeneous levels recorded by different registers.

Since only one target shapefile is chosen to map all regions at the same time, the level of detail of the polygon ID in the shapefile determines the target geographical level for all regions.

Each portion of the overall cumulative table must be extracted and merged to the relevant regional taxonomy table through the relevant regional data descriptor (see Figure 3.4.6).

Levels must be then rescaled to that used in target shapefile, so that extracted tables from different regions can be appended to a unique table that is used to produce an overall map.

In some cases, a decision can be made to change original options selected, to optimise mapping. If there are regions with very little detail (e.g. countries or NUTS 0), either they are dropped or only the coarser subdivision is used. In any case, aggregate tables must be all linked to the same shapefiles.

For more details on data representation in geographical mapping, see chapter on Data Definitions.

Software/Hardware specifications and performance

The statistical engine has been successfully developed without noticeable deviations from the original plan, successfully tested on both major OS Windows (Vista) and Linux (Fedora 10).

Hardware consisted of average Intel-based PCs /Notebooks, the least powerful with the following specifications: CPU speed 2.0GhZ, 1Gb RAM, hard disk capacity of 100Gb.

Installation of the software is identical regardless of the hardware, requiring R 2.8 or over, Latex, Java 6.0 and PostgreSQL plus various additional libraries/packages that are included in its distribution. Software is released using the GPL license by the BIRO Consortium (authors F.Carinci, L.Rossi for Serectrix).

In terms of performance, figures for the production of an annual local report from a test run on data from the Umbria register showed the following execution times on the same machine:

Centre	N Patients	N episodes	Elapsed Time
1	17,552	92,237	24' 25"
2	5,315	19,434	7' 01"
3	7,846	60,274	12' 20"
4	7,827	45,345	10' 51"
5	5,008	10,994	5' 22"

Since the application runs in parallel on different computers, the whole process relates to a total timeframe equal to the maximum, i.e. less than 25 minutes. Outputs occupy an average storage space of about 30Mb, including data to be transmitted to the central server.

To produce the overall report, the central engine implies the following timings for the different sets of 5 centres, progressively added in combination:

Centre	N Patients	N episodes	Elapsed Time
1	17,552	92,237	20' 12"
1+2	22,867	111,671	20' 54"
1+2+3	30,713	217,290	21' 33"
1+2+3+4	38,540	262,635	21' 56"
1+2+3+4+5	43,548	273,629	22' 27"

The progression suggests that, after an initial overhead required to load datasets in memory and to setup the environment, the system processes new centres quite rapidly. Performance increases with many centres included all at once.

To sum up, the whole statistical BIRO process for 5 diabetic clinics of average size takes about 50 minutes to be completed.

3.4.4 Discussion

The ultimate aim of the BIRO Information System (SEDIS) is to link policy-makers, practitioners and end users through secure Internet software specifically designed to provide reports from the analysis of a collaborative distributed data-warehouse.

The statistical engine represents a fundamental feature of SEDIS, as it delivers by design a range of outputs from the descriptive statistics more oriented at clinical diabetologists, to population-based standardized analysis that can be best suited for public health specialists.

The advantages offered by this component of BIRO are primarily due to the unique features of the overall design of the project, addressing the information-intensive management of chronic diseases.

In diabetes, definitions and practice guidelines change often, involving updates to the analytical software that must be re-run to get relevant up-to-date results.

BIRO allows to update its standardized information by linking statistical variables to a rich knowledge repository that uses evidence in context through a central “concept and data dictionary”. The dictionary is directly translated into database and statistical software, allowing to apply such definitions directly for the rapid production of new reports.

Through the generalisation of its data model, the same statistical results from the overall collaboration are saved into standard definitions, allowing to set appropriate terms of reference through which each connected region can apply the statistical system to benchmark average results against own data.

As a matter of fact, statistical modelling in BIRO allows to create an average population resulting from the whole network of centres in a very short time. Currently the same operation requires a significant effort to be realized, usually through ad hoc studies.

The interesting aspect of the statistical engine is that

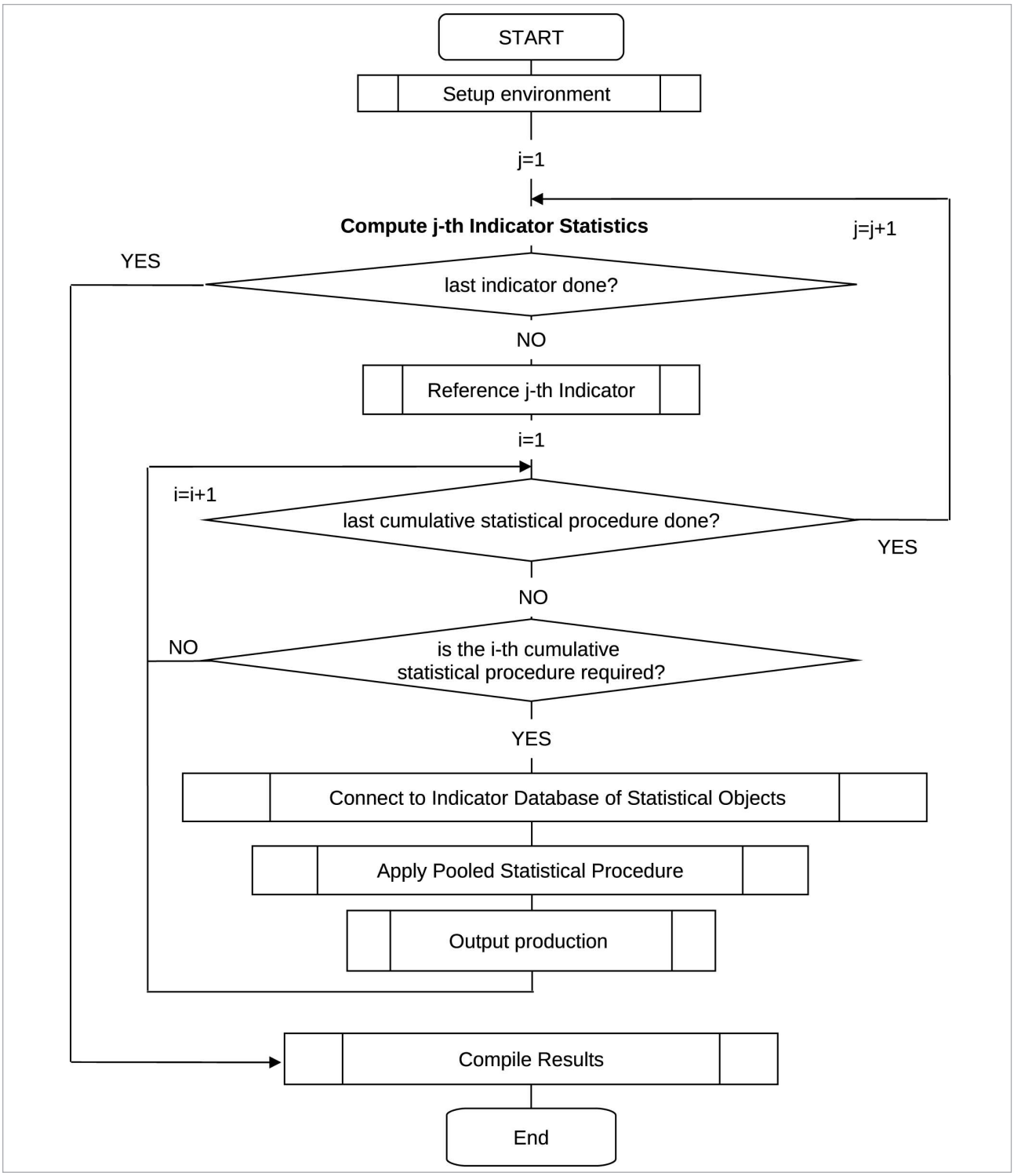


Figure 3.4.4. Central Engine Flow Chart

it both serves the European Union to provide updated data in a sustainable manner, as well as the local user to monitor in a rather inexpensive manner the clinical status of the population in own catchment area.

This overcomes the common problem of traditional epidemiological studies not returning enough information to the data collection unit. Such a controversial aspect frequently hampers the continuation of a collaboration and impedes the collection of high quality data through an active participation of users. BIRO has a shared infrastructure that may also help to improve validity and completeness.

The engine transfers the statistical ability to regions, areas, or individual clinical centres. It provides a platform for accurate benchmarking that currently does not exist at the point of health care provision.

BIRO allows answering very rapidly to questions e.g.: what is the average difference in glycated haemoglobin that a system can achieve within six months, with this therapy, in similar conditions? What is the average length of stay for a particular procedure? What outcomes can clinicians achieve for this particular population of patients? Why regions experience different variations? Why average outcomes are so different from my direct experience?

To make this information effective, results must be easy to interpret and use.

Both policy makers and physicians may gain particular advantage from browsing health reports through common formats, e.g. html and pdf files. Outputs may also be customized for own use. Tables and graphs can be imported in own presentations. Special reports can be produced for a class of physicians vs. a subset of regions/centres. A physician may inspect the average variation of glycated haemoglobin over time for a class of patients in a region, as opposed to the same results across different regions. Variation in clinical practice may be directly inspected.
The availability of a well-constructed and validated model represents an important step in the construction of a novel infrastructure that is capable of involving many more aspects and can be equally applied to different geographical areas and collaborative networks.

An important element of the BIRO framework is the development of open source software based upon powerful languages that have very little to envy from commercial counterparts. By the way, there are also limitations that are worth to be highlighted.

The statistical engine is based upon techniques for standardization and risk adjustment that do not allow to control for the potential bias existing in disease registers. As data is collected from automated sources, the inclusion of patients cannot guarantee about their level of representativeness.

In the statistical engine, no random selection process is put in place to get unbiased statistical estimates. The system relies upon collected data, on top of which it applies standard case-mix adjustment techniques.

There has been no development of Bayesian techniques to adjust for random variations that may occur particularly in clinical centres characterised by small sample sizes.

Nevertheless, BIRO is part of a progressive approach to disease registers that can enhance features of the current statistical engine.

Firstly, the BIRO project aims at involving more and more centres in the collaboration, and within the EUBIROD project recently started it already grew up to twenty-two centres, from the seven originally involved.

Secondly, the BIRO data specifications include specific items that take into account data quality, including the concept of “validated diabetic patient” that must be taken as an important parameter to monitor a clean composition of the population under study.

Finally, the BIRO system, with its flexible data model, encourages further use of data linkage locally to pool clinical data with different administrative data sources (hospitals, diabetic clinics, GPs, pharmaceutical expenditures, pathology tests, etc), progressively covering the overall diabetic population in an exhaustive way, as it has never been possible before.

Once this will be realised, sampling may be specifically used to monitor quality and precision of regional registers, which in the meantime may have become the gold standard in statistical information, from the perspectives of both sustainability and speed of use.

The range of statistical routines that will be included in the engine is planned to be expanded in the framework of the work package “Epidemiological analysis” of the EUBIROD project.

3.4.5 Conclusions

The statistical and central engines constitute core components of BIRO for the production of core outputs of the system.

Their application serves both the production of data for local stakeholders and the European Commission.

The application of the statistical engine in regional and individual clinical units can be used in different ways.

Through it, networks of professionals may self-evaluate more rapidly and effectively to implement *clinical governance* in the local health system.

Disease management may rely on accurate information to feed a virtuous cycle that can lead to improved health outcomes for the patients and improved information for the European Union through more accurate registers.

Each individual clinician, once inducted to using the software, can apply it independently and contribute to

the construction of a European network through the production and submission of aggregate data to the central server.

Through the engine, researchers can deploy sophisticated statistical models for ordinary use and deliver more accurate benchmarks through multivariate risk adjustment.

The development of BIRO statistical components offers an open product available at no charge that will allow disseminating capacity in the health sector for better information, in the interest of public health.

Table 3.4.7 Regional Geographical Taxonomy Table

Variable	Description	Class	*Example for Italy*
continent	name of continent	BIRO-0	European Union
country	name of the country	NUTS-0	Italy
macroarea	group of subnational areas	NUTS-1	Macro-regions
region	name of region	NUTS-2	Region
lha	local health authority	BIRO-1	ASL
province	province	NUTS-3	Province
dhu	district health unit	BIRO-2	Health District
postcode	name of subprovincial unit	BIRO-3	Commune

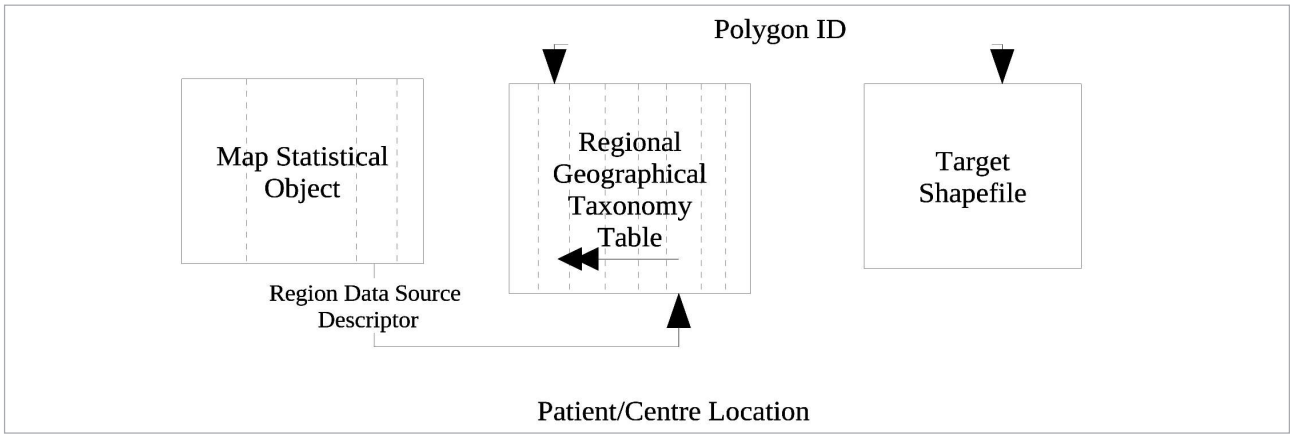


Figure 3.4.5. Linking geographical references to a target shapefile in BIRO statistical engine

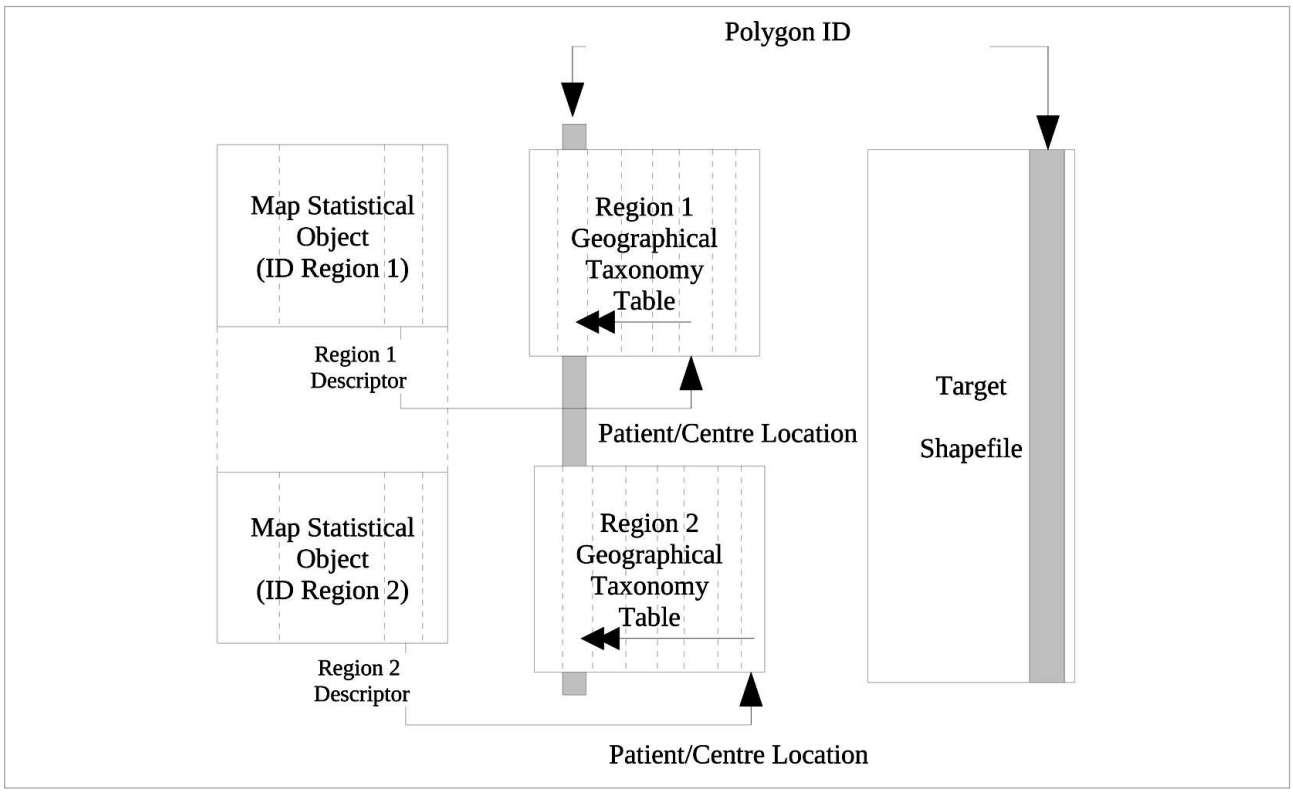


Figure 3.4.6. Linking geographical references to a target shapefile in BIRO central engine

References

1. The Tallinn Charter: Health Systems for Health and Wealth, WHO European Ministerial Conference on Health Systems: “Health Systems, Health and Wealth”, Tallinn, Estonia, 25-27 June 2008, available at: www.euro.who.int/document/e91438.pdf
2. AHRQ Quality Indicators, Guide to Prevention Quality Indicators: Hospital Admission for mbulatory Care Sensitive Conditions, Department of Health and Human Services, Agency for Healthcare Research and Quality, October 2001, Version 3.1 (March 12, 2007) http://qualityindicators.ahrq.gov/downloads/pqi/pqi_guide_v31.pdf
3. Carinci F and Romagnoli C on behalf of the DVSS Collaborators, A case of stewardship and systematic evaluation in decentralized governance: The Umbria Document for the Evaluation of Health Determinants and System Strategies (DVSS) in Greco D and the CCM Project Management Group, Palumbo F, Arcangeli L, Di Martino MP, La Falce M, Carinci F, Romagnoli C and the DVSS Collaborators, “Stewardship and governance in decentralized health systems: an Italian Case Study”, Italian Ministry of Labour, Health and Social Policy, Rome, 2008, available at: http://www.ccm-network.it/documenti_Ccm/pubblicazioni/Stewardship_Tallin_2008.pdf
4. Carinci F, Corrado D, Dettorre A, Pellegrini F, A multilevel approach to health systems analysis using RISS (Reporting-by-Intranet Stat System), 4th International Conference on the Scientific Basis of Health Services Research, Sydney, 22-25 September 2001.
5. M. Benedetti, F. Carinci, M. Federici, The Umbria Diabetes Register, Diabetes Research and Clinical Practice, Volume 74, S200-S204.
6. F.Carinci et al., RISS-H - Progetto Diabete - SDO di Pazienti Diabetici - Analisi stratificata per Ospedali, Regione Veneto 1999, available at http://www.fabcarinci.net/projects/riss
7. Carinci F, Rose W, Advantages of Distributed Statistical Processing in Health Information Systems: the H+MetaBase project, Technical Report, Monash Institute of Health Services Research, Melbourne, Australia, 2003.
8. SAS Language. Version 8.2. Cary, NC: SAS Institute Inc; 2001
9. Churches T, Carinci F, Open source at the interface between policy and academia: towards evidence-based information systems, 4th International Conference on the Scientific Basis of Health Services Research, Sydney, 22-25 September 2001.
10. Carinci F, Federici M, Massi Benedetti M, Diabetes registers and prevention strategies: towards an active use of health information, Diabetes Research and Clinical Practice, Volume 74, S215-S219.
11. Rice N, Leyland A, Multilevel models: applications to health data, J Health Serv Res Policy, 1996; 1(3), 154-164.
12. A. Whitehead, J. Whitehead, A general parametric approach to the meta-analysis of randomized clinical trials, Stat Med (1991) , 1665–1677.
13. Carinci F, Southern Population Health Information System (SPHIS). An integrated Population-based Data Warehouse, http://www.fabcarinci.net/projects/chsr/sphis.html
14. The R Project, Reference Index, available at: http://cran.r-project.org/doc/manuals/fullrefman.pdf
15. Wikipedia: Nomenclature of Territorial Units for Statistics, available at: http://en.wikipedia.org/wiki/Nomenclature_of_Territorial_Units_for_Statistics

The BIRO web portal

Peter Taverner, Svein Skeie

ABSTRACT

Introduction
A single point of access to the main products developed by the BIRO Consortium would allow developers to connect to both data contributors and users of the system. Providing links to the most essential aspects of quality of care and health status in diabetes through a user friendly interface can positively influence both policy and practice. An embedded mechanism for presentation and automated updating of selected indicator outputs produced by the statistical engine could allow browsing of current and relevant information.

Objectives
To build a dedicated web portal allowing users of a shared information system to navigate through the materials, methods and results obtained by the BIRO system.

Materials and methods
A web portal was developed using established open source standards e.g. the Apache webserver, Postgresql database, PHP, XML, PHP. The Drupal content management framework provides facilities for security, database-connectivity, content management, and menu systems that allows web administrators to maintain and update the web portal with minimal requirements of effort and IT skills. But it allows for programming of custom modules if they are required. Collaborative work among members of the BIRO Consortium allowed definition of a common design for the web portal, and production of its content according to own skills and expected contribution.

Results
The functionality of the web portal is provided at two technical levels. (i) A Web content management framework for easy maintenance of static content. (ii) Custom programmed modules for dynamic and automatically updated display of indicators and indicator metadata. These modules are driven by data and parameters derived from the Report Templates , primary data dictionaries and the statistical engine and supplied from outside the website. This allows displays to be maintained and adapted without re-programming.

Conclusions
The BIRO web portal was developed to both describe the development of the project and to provide a structurefor the publication of results on a continuous basis. The structure, mechanisms and ‘open source’ nature makeit an easily adaptable starting point, through which the Consortium can grow a more complete interface for a European Diabetes Register. As a general model, the web portal can also be seen as a cost effective solutionfor the publication of results that can be directly obtained in the framework of a collaborative information system.

3.5.1 Introduction

The BIRO project has developed and demonstrated a model that enables the building of multi-national shared information systems for sensitive clinical data. It has also done all the work necessary to implement the model for diabetes.

In the context of planning a sustainable shared information system, the 'Reports Template' work package proposed providing for distinct target audiences of governance,health care and research, and people with some reason to seek information on diabetes.

However, the scope of BIRO was technical, being focused on solutions, mechanisms and enablement. It has built all the functionality of a shared information system for diabetes but within the term of the project access to data has been modest. Thus, the biggest value of the indicators presented by the project lies not in the initial information they provide on the state of diabetes and diabetes care in Europe, but in what they illustrate of mechanisms and the examples they provide of pertinent reports generated by semi-automated data collection, statistical processing and display. Thus, for the web portal that presents the BIRO project it is people who have a vested interest in establishing a sustainable multi-national shared information system or that such a system be established that are the audience targeted. Texts have though been included that attempt to describe the project to any curious reader.

For subsequent projects like EUBIROD that will use the BIRO model to establish a sustainable shared information system the focus will be on the data and analyses they produce. Consequently they will have other and wider audiences and other agendas. Each such project will make their own choices of how they will relate to their audience and what services they will provide. They will adapt and enhance their own web portals accordingly.

The mechanisms for indicator presentation on the web portal have functionality that is important to the BIRO model in use: (i) Automatic updates, without which running costs of a frequently refreshed system would be high. (ii) Integration with Report Templates, Clinical reviews, data dictionaries and the statistical engine such that the choice of indicators to be presented and the form of presentation is managed through data and not by programming. (iii) Transparency. Providing convenient access to relevant metadata in order to enhance the understanding of each indicator and enable evaluation of the quality of the underlying data. This makes the Web portal application an integral component in the BIRO model and is therefore provided in the Transfer of Technology collection of components. Because of this requirement the application has been designed to be easily adaptable and has been built using only Open Source resources.

3.5.2 Objectives

Scope of the web portal is to provide texts and documents via a navigation menu, presenting all the essential aspects of a web site directly linked to a cost effective, fully automated, shared information system.

The web portal is designed to serve three main purposes:

- to document the work and achievements of the completed BIRO project.
- to provide and demonstrate mechanisms for inform-

Figure 3.5.1 Example of static text including diabetes information.

ative and effective dissemination of results generated by a sustainable shared information system based on the BIRO model.
- To provide open source software required to transfer technology at low cost across Europe

3.5.3 Materials and methods

Open source tools and resources were targeted to provide the basis for a web portal application that could be widely used and disseminated. This way, implementing and adapting the application in different setting would not require the purchase of any proprietary software.

The following technologies were chosen for implementation:

- Apache web server
- PostgreSQL/MySql relational database systems
- PHP (Web programming language)
- Drupal (content management framework)
- XML

Drupal (http://www.drupal.org) is an open source content management framework consisting of a small core that gives the developer a comprehensive interface for implementing custom made modules. The framework provides content-management, security, database-connectivity and menu-system. It also gives the ability for non-technical persons to maintain and update the portal, with minimal requirements of source coding.

3.5.4 Results

The Drupal content management framework has facilities and utilities for forming the appearance and layout of the web pages, managing static menus, inserting texts and linking to documents etc.

Drupal has been used to build the basic web application and to load static texts and documents e.g. pages to describe the problem of diabetes and its management (Figure 3.5.1). The same Drupal facilities enable people with limited web building skills to modify these aspects of the application to their own requirements.

The key functionalities ensured by the portal required the programming of two custom Drupal modules, whose ease of adaptability was ensured by parameters and data generated, registered, and maintained by other applications and processes external to the web application:

a. The custom Drupal module for Indicator presentation

This module provides two services:

- A menu that provides the reader with an overview of the indicators that can be viewed and the ability to browse or step through them (Figure 3.5.2). The menu is hierarchical and organised by the same theme chapters and order as expressed in the reports templates register. The module programmatically generates this menu based on the data provided by the reports template. In case of any change in the template, data are modified and saved, and menus are regenerated.
- A page on which the actual data for the indicator selected from the menu is displayed (Figure 3.5.3). The key to this part is that the module concerns itself solely with display of data supplied. What it will display, for a given indicator, is supplied as a collection of “display ready elements”. These can be of type text, table or chart/diagram. Specification of where on the page each element shall be shown and how

Figure 3.5.2 Example of indicator definitions.

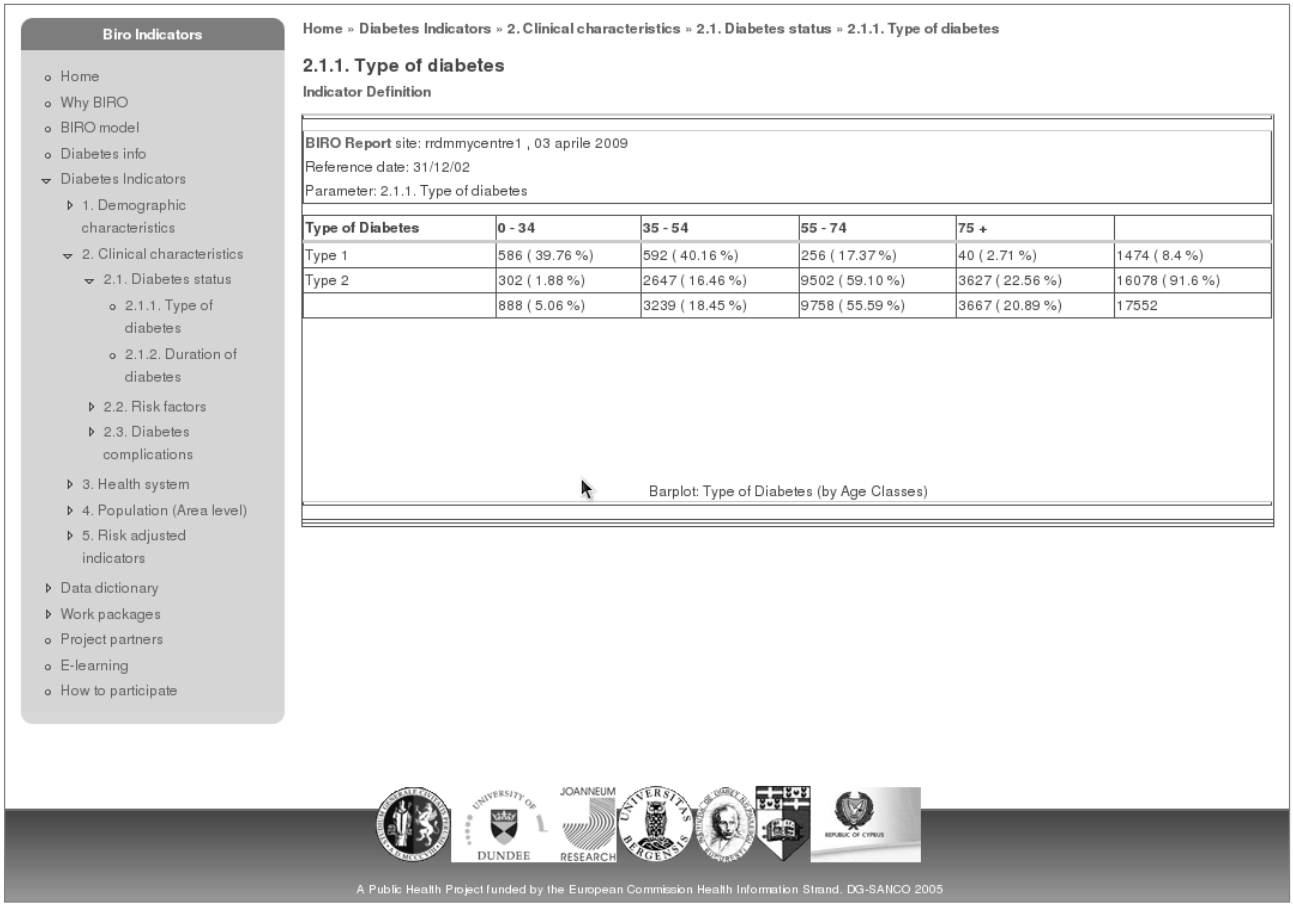


Figure 3.5.3 Example of indicator outputs.

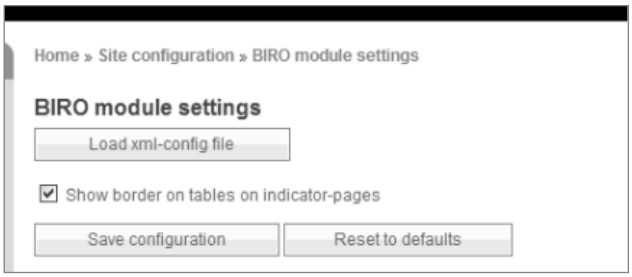


Figure 3.5.4 Configuration management screen.

big it shall be is also supplied as data. For positioning and sizing of elements the module operates with a simple grid. This makes it easy for the person designing the page layouts for the elements. For each element the position of the top left corner is given as a column and row coordinate, the width as a number of columns and the depth as number of rows.

When a new indicator is added to those already supported:

- it is named and the justification for its choice is documented with supporting references (see clinical review).
- the calculation is defined and the list of involved variables listed (see clinical review)
- the strata and form of presentation is defined (what type of chart, table or diagram etc). (see reports template)
- a program is written to instruct the statistical engine on how to access primary data, how to perform data processing and statistical analysis, and how to generate the output required as 'display-ready' elements.
- any other 'display-ready' elements required are created. For example: blocks of explanatory text. (see reports template)
- a file name is registered, for each display element, for where the content of that element will be stored. (see reports template)
- The layout of the elements that will make up the presentation of the indicator on the web page is specified. (reports template)

In case of changes in the collection of indicator base data, the web portal "Indicator presentation module" needs also to be re-configured. To do this, the layout specifications for each display element are collected from the reports template register and delivered to the web application. The module's configuration process is then invoked.

This will generate a new menu and prepare the module for actual display. The module will be then capable of showing each indicator presentation.
For each indicator it has a list of all the elements in the display. For each element it has the pathname of the file containing the content of the element and the specifications for its position and size on the page. This configuration process need be repeated only when the data it depends on from the reports template register are changed.

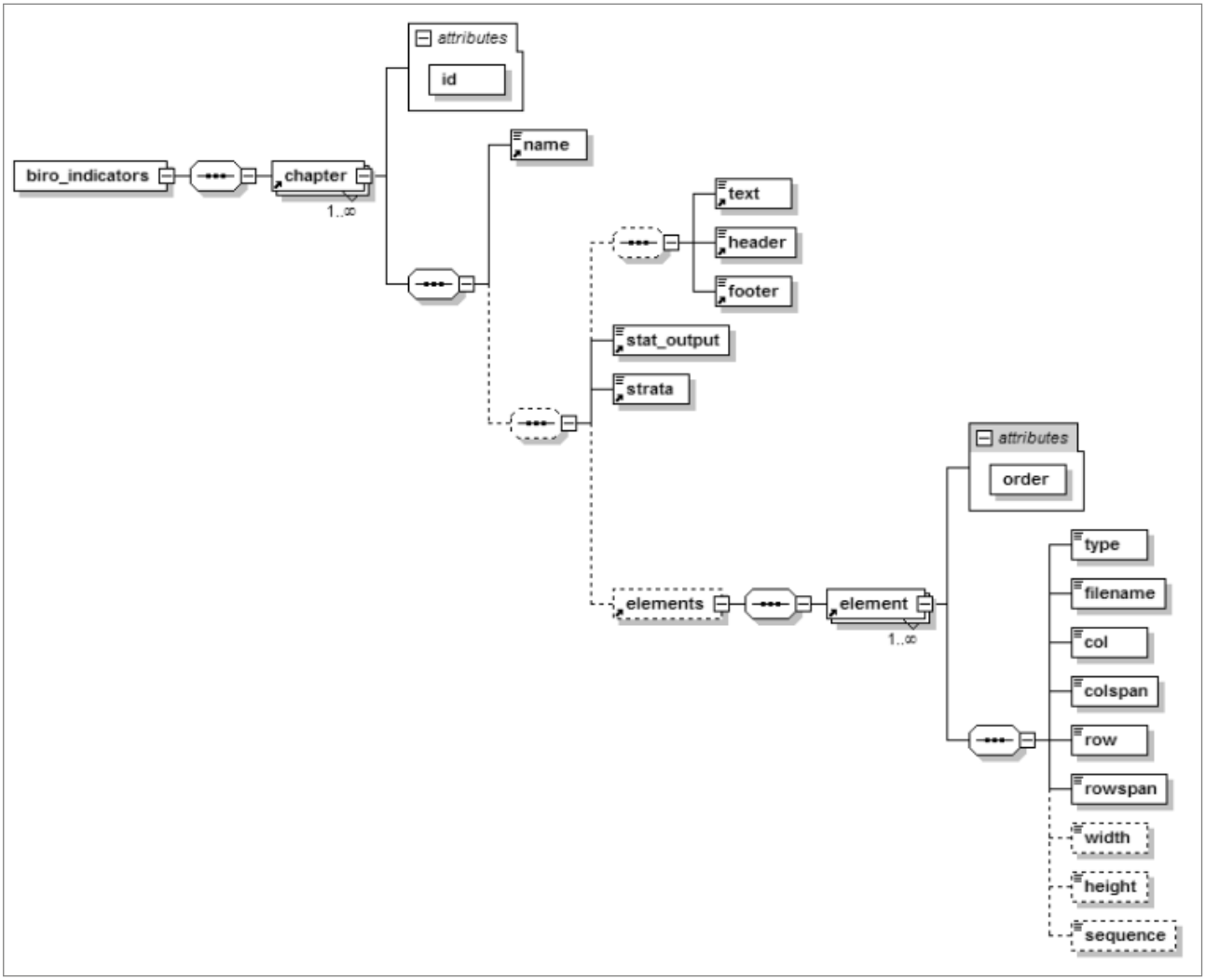


Figure 3.5.5 Schema defining the indicator configuration data document.

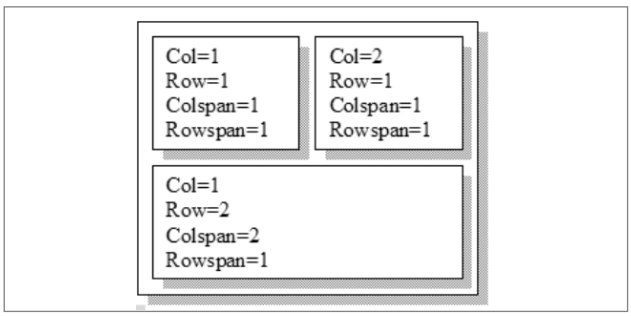


Figure 3.5.6 Example of element layout.

Updating the results presented on the web can be done whenever and as often as desired, simply by running the statistical engine. The definitions of the indicator and of what to produce as output are built into the program. A single program is run for each indicator to generate one or more display elements (tables or charts). Each element generated is written to its specified file, which is where the web application will read from when it needs it. Processing the data can be done when it is appropriate because of new data or defined time intervals. The display on the web will be automatically updated without involving a web application, since the portal just displays what is in the files, without knowing when or how it got there.

Technical details

When this module is installed it will create two working tables and generate an administration screen for managing the configuration process (Figure 3.5.4).

The configuration specifications for indicators are held in xml blocks in the Reports Template register, one for each indicator. When specifications are added or changed the configuration process for the module must be run. The first step is to collect all the indicator xml blocks into a single xml document and make it available to the web application.
A schema has been defined for this “indicator configuration data document” (Figure 3.5.5).

Currently building the xml data document and uploading it to the server is done manually.

The remaining steps are done programmatically and are invoked via the configuration management screen provided by the module. The xml data document is

Box 3.5.1. An example of an xml configuration file

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Sample XML file generated by XMLSpy v2008 sp1 (http://www.altova.com)-->
<biro_indicators xsi:noNamespaceSchemaLocation="indicators.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
       <chapter id="1">    <name>Demographic characteristics</name></chapter>
       <chapter id="1.1">
              <name>Age (Classes)</name>
              <text/><header/><footer/>
              <stat_output>histogram</stat_output>
              <strata>Gender</strata>
              <elements>
                     <element order="1">
                            <type>image</type>
                            <filename>1_1.png</filename>
                            <col>1</col>
                            <colspan>1</colspan>
                            <row>1</row>
                            <rowspan>1</rowspan>
                            <width>50%</width>
                            <height>50%</height>
                     </element>
                     <element order="2">
                            <type>table</type>
                            <filename>1_1.html</filename>
                            <col>1</col>
                            <colspan>1</colspan>
                            <row>2</row>
                            <rowspan>1</rowspan>
                     </element>
                     <element order="3">
                            <type>text</type>
                            <filename>desription1_1.html</filename>
                            <col>1</col>
                            <colspan>1</colspan>
                            <row>3</row>
                            <rowspan>1</rowspan>
                     </element>
              </elements></chapter>
       <chapter id="1.2">
              <name>Gender</name>
              <text/><header/><footer/>
              <stat_output>histogram</stat_output>
              <strata>Age</strata>
              <elements>
                     <element order="1">
                            <type>image</type>
                            <filename>1_2.png</filename>
                            <col>1</col>
                            <colspan>1</colspan>
                            <row>1</row>
                            <rowspan>1</rowspan>
                            <width>50%</width>
                            <height>50%</height>
                     </element>
                     <element order="2">
                            <type>table</type>
                            <filename>1_2.html</filename>
                            <col>1</col>
                            <colspan>1</colspan>
                            <row>2</row>
                            <rowspan>1</rowspan>
                            <sequence>2</sequence>
                     </element>
                     <element order="1">
                            <type>text</type>
                            <filename>desription1_2.html</filename>
                            <col>1</col>
                            <colspan>1</colspan>
                            <row>3</row>
                            <rowspan>1</rowspan>
                     </element>
</elements></chapter></biro_indicators>
```

Reference	Name	Parameter	Datatype
BIRO001	PAT_ID	Patient ID	String(12)
BIRO002	DS_ID	Data Source ID	String(10)
BIRO003	TYPE_DM	Type Of Diabetes	Enumerated
	Code: Value		
	1: Type 1		
	2: Type 2		
	3: Other Types of Diabetes		
BIRO004	SEX	Sex	Enumerated
BIRO005	DOB	Date of Birth	Date/Time
BIRO006	DT_DIAG	Date of Diagnosish	Integer
BIRO007	EPI_DATE	Episode Date	Date/Time
BIRO008	SMOK_STAT	Smoking Status	Enumerated
BIRO009	CIGS_DAY	Cigarettes per day	Integer
BRIO047	ALC_STAT	Alcohol Status	Enumerated
BIRO010	ALCOHOL	Alcohol Intake	Integer
BIRO011	WEIGHT	Weight	Real
BIRO012	HEIGHT	Height	Real
BIRO013	BMI	Body Mass Index	Real

Figure 3.5.7 Example of variable definitions.

validated against the schema. If valid, the module will traverse the xml and update the two tables in the database created for this. The indicator menus will then be rebuilt based on the new data. If the xml is not valid an error will be displayed and the configuration will not be changed.

The xml configuration block for each indicator specifies content of the display as a collection of elements and their layout as their position and size within a standard table frame. Elements can be of type image, text or table. The position and size of each element within the frame is specified by setting column, row, column-span and row-span. An element of type image can also have specified width and height as percent of its original size (Figure 3.5.6).

In the xml configuration block the content of each element is specified as the pathname of a file. The content of each file is “display ready” and requires only to be positioned and sized. Files for text and table elements contain html. Image elements contain a picture. It is this feature that enables programmatic updating of indicator displays as part of the statistical engine ‘refresh with new data’ process.

The statistical engine generates output for the web as “display ready” tables and charts and writes them to the a designated location using filenames specified in the xml configuration blocks. With each reprocessing of the data the content of these files is replaced. The web indicator module simply displays what is currently there.

The output to be generated by the statistical engine for each indicator is specified as part of the Report Templates.

To view the indicators the following information is required (Box 3.5.1):

Name	*Usage*	*Description*
Chapter ID	Mandatory	The chapter
Name	Mandatory	The name of the indicator
Text	Optional	Describing text
Header	Optional	Header-text for indicator
Footer	Optional	Footer-text for indicaor
Statistical	Optional	Type of output (histogram, line, etc)
Strata	Optional	
Sortorder	Mandatory	The order in which the indicators will show in menu

For each element specified for an indicator the following information is required:

Name	*Usage*	*Description*
Chapter ID	Mandatory	The chapter ID. References the chapter id in the corresponding indicator
Type	Mandatory	image, text or table
Filename	Mandatory	Contains the name of the file containing the data.
Row	Mandatory	
Column	Mandatory	
Rowspan	Optional	Default 1

Column-span	Optional	Default 1
Vieworder	Mandatory	Display-order of element.
Width	Optional	Only applies to images.
Height	Optional	Only applies to images.
Sequence	Optional	If type is table, this attribute can be used to identify which table to pick from file, if the file contains more than one table.

All the menu items in the Indicator menu link to the indicator module. The section of the indicator data to display is given to the module as a parameter. For example: "http://<hostname>/?q=biro/1/1", (where "biro" is the name of the Drupal indicator module), points to the module with the indicator "1.1" as a parameter.

The module queries the *biro_indicators* table for the specified indicator. If found it then queries the *biro_data* table for the elements to show. It generates the html code needed for laying out the elements according to the settings for each element and returns this to the Drupal engine.

The indicator menus (Figure 3.5.3) are rebuilt each time the configuration process is run.

b. Custom Drupal module for browsing data dictionaries

This module provides a generic mechanism for programmatically generating a browser within the web application from a structured file of data. Intervention from a web programmer is not required. The menu generated and the data displayed on selection of a menu item are determined entirely by the data file provided.

This offers an easy way of providing access to dictionaries of base data. It is particularly useful where the data are labile. The data can be maintained in some application external to the web and the module will programmatically update the browser.

The module has been used in BIRO to specifically address the problem of "transparency", i.e. giving the user access to the metadata underlying the result presentations, in order: (a) to provide a better understanding of presentations, and (b) to provide insight into the nature and quality of the data the result was based on.

The solution implemented in the BIRO web portal provides access to the definitions of the indicators (Figure 3.5.2). This includes the primary calculations, the variables involved and their definitions.

These data can be accessed by freely browsing each data dictionary. The dictionary for indicator definitions includes a list of the variables involved for each indicator. The reader can drill down to the variable definitions by clicking on a variable (Figure 3.5.7).

The display will then jump to that variable definition displayed in the variable dictionary. For greater reader convenience these metadata are also accessible directly from the displays of indicator results. When viewing an indicator result the reader can click a button on that page which will cause a jump to the Indicator definitions dictionary with that indicator already selected.

Technical details

The module requires the xml data documents to be uploaded to a designated area on the web server. The module can then use these to populate its relational tables in the web-server database and will then generate the menu required for browsing. The module provides an administration screen for this.

To freely browse the data select the menu item "Data dictionary" which will present two sub items: "Indicators" and "Variables".

Select "Variables" for a list of all the variables. Click on one of these and definition details will be shown. Select "Indicators" for a list of all the indicators. Click on one of these and details of its definition will be shown. This includes the variables involved. Click on a variable and the variable list is shown with that variable selected.

On the indicator display page there is a button labelled "Indicator definition" located right under the name of the indicator. When this is clicked the display jumps to the indicator definitions with that indicator selected.

Database tables for the Data dictionary module:

biro_datadict

Name	Datatype	Key	Description
Reference	varchar(255)	PK	Variable (Ex. BIRO001)
Field_name	varchar(255)		Variable-name (Ex PAT_ID)
Parameter	varchar(255)		Description of variable (Ex. Patient ID)
Datatype	varchar(255)		

biro_datadict_enum

Name	Datatype	Key	Description
Reference	varchar(255)	PK1	Variable (Ex. BIRO001)
Enum_code	Int	PK2	Ex. 1
Value	varchar(255)		Value of code (Ex Type 1)

biro_crossref

Name	Datatype	Key	Description
Chapter	varchar(255)	PK	Ex 1.1
Target	varchar(255)		Reference to indicator
Name	Text		Ex Age (Classes)

biro_crossref_stratum

Name	Datatype	Key	Description
Chapter	varchar(20)	PK1	Ex 1.1
Stratum	varchar(40)	PK2	Reference to variable, Ex DOB

biro_crossref_output

Name	Datatype	Key	Description
Chapter	varchar(20)	PK1	Ex. 1.1
Output	varchar(20)	PK2	Output-type, ex Histo-gram

3.5.5 Discussion

The web portal developed as part of the BIRO model represents the primary mechanism for dissemination of information and results.

In the case of the BIRO project, this scope concerns the state and quality of care of diabetes mellitus in the contributing population. For this type of information system data will be collected as an ongoing process and the state of the disease being studied will be described and monitored using a battery of carefully selected/designed indicators. These and other special reports, analyses and projections will be presented through the BIRO website[2].

Analysis of new data with subsequent updates of the indicator displays and reports must be performed regularly to keep the information presented current and interesting for the targeted audiences. A web application is the fastest and cheapest way of doing this.

The more extensive and frequent the data harvesting, the more complex and refined the data processing, the greater the number of indicators and reports presented, the greater the continuing burden of work. Unless all the steps involved can be to a high degree automated the cost of running such systems could become prohibitive. Automation has therefore been a high priority design goal in the BIRO model. This included also the design of the web portal.

While the number and design of indicators implemented on the web may not change often, it must be possible to specify design and content of each indicator display independently. There can be many indicators and if each required its own custom built page on the website both initial set up and subsequent changes would be tedious and costly. Furthermore, the web application would be very much less generic and less suitable as a 'transfer of technology' component.

Clinicians and epidemiologists choose or design indicators, and determine how they shall be possibly presented. It would be best if implementation on the web could be done by registration of specifications, without requiring the intervention of a web programmer.

Implementing a new indicator involves the statistical engine as well as the Web Portal. However, differently from the web portal, the statistical engine programming is required for each indicator. That programming must be done manually but the definitions of the calculations involved, the form of output required and other base data should be specified by the indicator designer and be available from the BIRO database.

A library of statistical engine programs has been already developed as part of the BIRO project, although can only run as batch script files.

Transparency is a vital aspect of quality assurance of systems that generate and present data. The website should provide this. Some of this will be documents that explain the essential features and mechanisms of the system and provide background information on the partners data sources.

However, much of the data essential to this are in 'dictionaries'[3]. Examples of these are: variable and indicator definitions and documentation from each partner of their compliance with the standard variable definitions. The reader should be able to browse such dictionaries. For user convenience, the web portal should also associate relevant collections of such metadata with each indicator display. When viewing an indicator display a click of a button should suffice to access those metadata that are most relevant to enhance understanding of that indicator and enable evaluation of the quality of the underlying data.

The web application needs to be generic and driven by externally provided parameters and data so that others implementing the BIRO model can adapt it to their own requirements and maintain it almost entirely through changes to externally provided data and document links.

The "metadata" solution implemented on the BIRO web portal is a somewhat restricted and rudimentary, but it is a good start to an important function. Subsequent projects, like EUBIROD, can be expected to develop enhanced versions allowing access to a greater range of metadata, and improving the convenience and presentation of the metadata associated with each indicator.

The web portal must be regarded as an integral functional component of the BIRO model, and associated deliverables.

3.5.6 Conclusions

The web portal developed by the BIRO workpackage 11:
• presents the BIRO project and the BIRO model
• has mechanisms for informative and low overhead presentation of frequently updated results
• is generic and technically suitable as a 'Transfer of technology' component.

As such, the WP has achieved its objectives.

The BIRO project is to be followed by the EUBIROD project to establish a sustainable shared information system for diabetes in Europe. EUBIROD will use both the generic BIRO model and the extensive work done in BIRO on applying that model to diabetes. But while EUBIROD has been enabled by BIRO, it is a distinct project with own objectives. Its focus will be on the data it collects and analyses and what uses are to made of them. As such, it will create its own web portal and can do this by adapting and enhancing the BIRO web portal component.

The BIRO web portal will be retained, not only because BIRO has been a successful project, but also because of the general value of the model.

Making methods available, and nurturing the concept and practice of open source software can reduce costs in the health care sector[4]. This way, plans whose thresholds were formerly too high can become immediately feasible. The proliferation of methods e.g. BIRO, in addition to their primary value as information systems, would also promote acceptance of international standards and identification of 'best practices'.

Providing support for the BIRO web portal could allow:

• providing better links to related projects,
• identifying new or enhanced open source software releases
• promoting and fostering the open source community

This could be done either by providing support for the BIRO web portal, so that it can provide this service or by establishing a new web portal for that sole purpose.

The web portal for the BIRO project can be viewed through the main page of the project[2], where more technical details are directly made accessible through the final WP11 report, available at the main website.

The entire source code and documentation can be directly downloaded from the reserved area of the BIRO project, made accessible through a special account that must be obtained by the BIRO project manager.

References

1. Raymond E. The cathedral & the bazaar, O'Reilly Media, 1999
2. The BIRO project website, available at: http://www.biro-project.eu
3. International Diabetes Institute, The Australian Diabetes, Obesity and Lifestyle (*AusDiab*) study, AUSDIAB Indexionary, available at: http://www.diabetes.com.au/AusDiab2000datadictionary
4. Churches T, Carinci F, Open source at the interface between policy and academia: towards evidence-based information systems, 4th International Conference on the Scientific Basis of Health Services Research, Sydney, 22-25 September 2001

Technology Transfer

Simion Pruna, Joseph Azzopardi, George Olympios, Vivie Traynor, Andriana Evripidou

ABSTRACT

Introduction
The BIRO project requires exporting data from local diabetes registers towards a common format that would allow using the same standardized software. Procedures will work on top of a BIRO database that is connected to a central server. The server receives tables from all members of the collaboration and produces European diabetes indicators on a routine basis.

Objectives
The aim of technology transfer is to describe the state of the art in diabetes medical records in three New Member States and to outline an initial evaluation of the BIRO software regarding possible obstacles to its practical implementation in Romania, Malta and Cyprus.

Materials and methods
New Member States involved in this activity did not possess an established, organized, integrated diabetes register that would allow a regular production of diabetes indicators. Through the conduction of a targeted workplan, the workpackage dedicated to technology transfer explored difficulties in the acquisition of BIRO material, the actual use of medical records in practice, and obstacles, risks and incentives to the dissemination of the BIRO system. Representatives from countries involved in the analysis discussed their capacity in participating to the definition of the system, through the collaborative production of documents for the collaboration, the ability to use data processing tools, and communication protocols e.g. access to the data server and installation of specialised software. A GUI tool was developed for the scope of allowing non technical people to use BIRO, and was evaluated in a dedicated workshop. Sample data from Romania, Malta and Cyprus were used to understand difficulties in the organization of databases and partners' ability to export data to the BIRO format to use the statistical system.

Results
The organization of diabetes databases is rapidly evolving in all countries examined. Recent developments show that large samples of patients/episodes can be successfully gathered in all cases, using tools that are already used and can converge towards more structured registers in a relatively short time. In all cases, the potential success relies on the personal determination and leadership of a small number of devoted actors. More involvement of policy makers and awareness of the public is required. The BIRO project demonstrated that the approach can be well understood by people with different background, but at the same time procedures involved require skilled technical experts who understand both information technology and the medical context in which these tools need to operate. The burden of maintenance (integration, storage and usage) is relatively a lesser problem, as it can be facilitated today by many tools available in the community, some of them being at no cost and yet very powerful and easy to use. Required data transformations to the BIRO format may not be that immediate, but the process is facilitated by a graphical user interface whose use facilitates a direct interaction between developers and users of the system. Current limitations and possible improvements have been formulated in a plenary session where notes have been taken on steps to improve the interface towards the definition of a comprehensive toolbox.

Conclusions
Barriers to cross-cultural communication still exist in all countries examined. Extensive implementation of the BIRO technology involves more integration of clinical and technological backgrounds. To become an effective component of clinical management, the BIRO Consortium must involve more all governance bodies (e.g. Ministry of Health, Professional Association of Diabetologists, etc.). A major focus on dissemination is required to expand the acquisition of the BIRO technology, allowing more users to come on board and test/evaluate the system in real life situations.

4.1.1 Introduction

Sharing data from multiple sources represents a major and growing challenge to healthcare services throughout Europe, particularly in new Member States.

The development of wide technology support and software tools for re-using and sharing data through database integration, represent a gold standard towards which many experts in the academia and government are attempting to make serious steps ahead.

Nevertheless, there are many difficulties hampering the evolution of effective solutions for health improvement e.g. the creation of a solid information infrastructure for diabetes management.

Data sharing involves more than just exchange of data: it requires solving important problems related to privacy, confidentiality and security[1-3].

Governments are asked to go beyond the obvious duty of guaranteeing an adequate protection of medical records. They must take decisions on balancing individual rights with public health planning. There are individuals who have the right to receive high quality of health care, yet they may not want to be enrolled in a register, or participate to a longitudinal study. There are institutions which may not allow to be precisely identified in a public performance report. Still, we keep saying that using diabetes indicators is paramount for the organization of health systems.

Technology to realise this may be a lesser problem, but still must be cost effective, and available to all to respect equity. The use of simple, agreed data descriptors (metadata) can facilitate the understanding of data to be collected, ensuring that medical records are of high quality through the application of shared coding procedures. The common goal of establishing a web portal may represent a visible outcome that users may substantially appreciate, as in the case of a collection of diabetes indicators that can be easily interpreted using an agreed vocabulary, continuously improved to ensure compatibility with current evidence. Introducing the whole process at the European level may be far from obvious.

The BIRO Consortium organized a special activity of "Technology Transfer" through which representatives from three New Member States (Cyprus, Malta and Romania), undertook a workplan allowing to explore barriers to the implementation of the system, starting from the use of new tools that could be easily deployed to all partners, e.g. open source technologies.

The scope of this activity can be summarised in two points:

- to evaluate the scalability and extensibility of the BIRO system to new Member States (Cyprus, Malta and Romania), by examining the current state of the art and barriers to its application
- to investigate the use of the BIRO system through on field testing on data from Cyprus, Malta and Romania in order to assess the validity and range of benefits obtained from the new system[4].

A central aspect tackled by this activity has been interoperability, i.e. the ability to exchange and use information across systems, organizations, and geographical boundaries. Software developed by the BIRO Consortium address this problem by developing tools that would assist in the extraction, analysis, and reporting of diabetes information[5-7].

Different categories of users have been envisaged by the system as users of indicators agreed through a targeted clinical review, including policy makers, health care professionals, and citizens.

A specific problem encountered in the BIRO system is that of data export from many different inputs towards a common format.

The BIRO Adaptor has been realised by the Consortium as specific software that can be used to such scope, delivering the common dataset in the form of the BIRO export (XML). Although the Adaptor requires to be configured and run properly, these steps have been included in the BIRO GUI according to specifications given by users.

Technology transfer dedicated specific attention to such tests.

The present chapter explores the state of the art in Cyprus, Malta and Romania, as an initial point for the evaluation of the potential use of BIRO.

For each country, details are presented in terms of (i) health care system; (ii) diabetes information systems. In a second section, the adaptation, implementation, setup and use of BIRO technology is presented as it was submitted to the three country cases.

In the third section, results of the application of BIRO software to local data from the three countries are briefly summarized.

The last section highlights major problems and obstacles that may limit the adoption of the BIRO system, and how they could be possibly overcome.

A short summary concludes the chapter.

4.1.2 State of the art in Cyprus

Health care system

Healthcare in Cyprus is provided by (a) public healthcare system, means tested, and (b) private healthcare system, privately paid for. All people with diabetes are entitled to free medical care in the public healthcare sector. This care is delivered at the five large hospitals and 32 rural health centres on the island. Care varies between the public and private sector and even between hospitals and rural centres, as there has not been an agreed common framework of care and management of diabetes. Diabetes care is largely

dependant on the individual doctor's knowledge and understanding of the disease.

Up to the time that Cyprus joined BIRO, no uniformed care existed largely because there were no national guidelines for the care of diabetes and no common framework within which people with diabetes would be looked after. The roles of other healthcare professionals in a multidisciplinary team i.e. diabetes specialist nurse, dietician and podiatrist were not recognised and the doctors would work alone in the care of diabetes. Structured education was not available and each doctor would provide his own teaching including dietary knowledge. This is still largely done in the private sector.

In 2005, as Cyprus joined BIRO, diabetes prevalence was known to be as high as 10.3% on the island. Since joining the BIRO program and with the tremendous assistance of the partners, Cyprus has come a long way in the care of diabetes.

This has been realised in a number of different ways. In April 2007 the first adult multidisciplinary run clinic on the island started operating at Larnaca Hospital. The team members are a consultant physician / diabetologist, a diabetes specialist nurse and a dietician. Other close associates include the ophthalmologist and the podiatrist (operating at present only in the private sector). Significant other associates e.g. nephrologists etc. are involved whenever necessary with the doctor referral system.

This clinic started operating as a direct result of the BIRO project and with the assistance of partners, developing a common framework within which people with diabetes can be looked after. The organization was built in accordance with the international guidelines on diabetes care and included a focus on all parameters suggested by the BIRO common dataset.

Aims of the clinic are:

- to provide evidence based practice and evidence based clinic guidelines
- to provide education, training and empowerment of people with diabetes, leading to self-management.
- to promote prevention or delaying of the onset of complications
- to promote prevention of diabetes in the community through education and screening

The Larnaca Clinic started running as a pilot study for the Ministry of Health. Three rural health centres of the Larnaca District (GPs) have been working in collaboration with the Larnaca clinic under the same common framework, including data collection based upon BIRO specifications.

The Ministry of Health has actively been supportive in the development of guidelines for the diagnosis and care of diabetes in the community. Further training for the care deliverers in diabetes has been organised and taking place. Healthcare providers in the field of diabetes have been travelling abroad, observing and working in diabetes centres along with colleagues, gaining valuable new knowledge and practical experience in the running of diabetes centres.

Twenty-five nurses undertook the Certificate in Diabetes Care course delivered by a leading UK University, effectively qualifying as Diabetes Specialist Nurses (DSNs). All these people are preparing to take up their places in effectively looking after people with diabetes in the community and the Diabetes Clinics within a multidisciplinary team approach.

Diabetes Information Systems

Since the project startup, BIRO contributed enormously to get Cyprus Ministry to develop diabetes care. This was favoured by the possibility to share and discuss the approach with individual partners and answering questions that would guide local experts through the process of starting up the system. Direct visit to gold standard centres allowed to see applications in practice and discuss with peers how to best manage the clinic, or to make audit rounds using information being collected routinely. In this case, networking amongst partners has proven to be one of the hidden, unwritten assets and benefits of the BIRO program.

With the start of the Larnaca clinic the beginning of a Diabetes Register has started.

The electronic database register for Diabetes was created in April 2007. It was developed in Microsoft Access by the Department of Information Technology Services (DITS). It includes all the data items of the BIRO common dataset, plus some additional ones. It has been installed at the Larnaca clinic and all existing data recorded up to the present in the diabetes clinic in Larnaca and the three rural health centres has now been transferred onto the electronic form. Data collected is initially recorded manually in all the centres.
All rules relating to data protection are adhered to and every person that is on the register has signed a consent form.

The expansion of the care program in diabetes and the adoption of the BIRO data collection program for the rest of the island will be taking place in stages over a period in the future. The three rural health centres in the Larnaca and Famagusta areas which have been collaborating with the Larnaca clinic have now got DSNs working part time, contributing vastly both at care and education of people with diabetes, and also in data collection. The Paphos General Hospital on the west of the island has now started to collect data for the program.

Action for diabetes in Cyprus continues along with BIRO developments (including EUBIROD), as formalised by the prospective workplan described in Figure 4.1.1.

4.1.3 State of the art in Malta

Health care system

All residents in Malta have access to a health system providing preventive, investigative, curative and reha-

Figure 4.1.1: BIRO Action plan in Cyprus

bilitative services. Health care is funded through local taxes, but is free at the point of delivery. A private health service is also available.

Persons with a low income (determined after a means test) are provided with free medications. A person who suffers from one or more of a specified list of chronic diseases (e.g. diabetes mellitus) is also entitled to receive free treatment irrespective of financial means.

The Government delivers primary health care mainly through eight Health Centres offering a full range of preventive, curative and rehabilitative services. The general practitioner and nursing services are supplemented by various specialized services that include antenatal and postnatal clinics, Well Baby clinics, Gynaecology clinics, diabetes clinics, ophthalmic clinics, psychiatric clinics, podology (Podiatric) clinics, Physiotherapy, and Speech therapy and Language Pathology clinic.

Community nursing and midwifery services are provided by the Malta Memorial District Nursing Association (MMDNA) on a contract basis.

The Government's Health Centre system works side by side with a thriving private sector and many residents opt for the services of private general practitioners and specialists who work in the primary care setting. Secondary care and tertiary care are provided from a number of public hospitals, the principal one being Mater Dei Hospital, which has around 800 beds. Mater Dei provides a full range of secondary and tertiary medical services, including transplant surgery and open heart surgery. Another 58 beds are available at Sir Paul Boffa Hospital, which has an oncology and dermatology unit, and 259 short/long stay beds are available at Gozo General Hospital. At Mount Carmel Hospital there are 563 psychiatric beds (short/long stay), while at Zammit Clapp Hospital there are 60 specialized geriatric beds. There are three private hospitals, St Philip's Hospital, with a capacity of 75 beds, in Santa Venera, Capua Palace Hospital, with 80 beds, in Sliema and St James Hospital with 13 beds in Zabbar. St Mark's Clinic with a capacity of 5 beds in Msida also offers private hospital services. National health promotion activities are coordinated by the Department of Health Promotion. This encompasses the Health Promotion Unit and the Nutrition Unit. Both Units liaise closely with the Department of Education and with the mass media. Several preventive programs are run on a national scale, such as the free immunization program, which covers a wide range of illnesses. Health Centres provide extensive preventive services, such as Well Baby clinics, Well Woman clinics, routine blood pressure and cholesterol check-ups, smoking cessation clinics, screening for diabetes, and ophthalmological check-ups. There are also specialized preventive activities that are hospital-based (such as thyroid function screening for neonates).

The main diabetes clinic in Malta is at Mater Dei Hospital. There are six other peripheral clinics. All work is coordinated through the central clinic at Mater Dei. Between them these clinics take care of 95% of all diabetic people in Malta and are therefore in an ideal situation for whole population studies on diabetes. Over 1,300 new cases and over 12,000 follow up cases are seen yearly at the main diabetes clinic at Mater Dei and over 12,000 follow up cases are seen yearly at the peripheral clinics. All newly referred patients are initially assessed and managed at the main clinic and referred where appropriate for follow up at one of the peripheral clinics. Patients are seen routinely for follow up visits approximately every few months or as often as necessary. The diabetes clinics in the peripheral centres are attended by peripheral health centre doctors; doctors or consultants from Mater Dei Diabetes Clinic visit the peripheral clinics

for consultations. The diabetes clinic has a fully developed diabetes computerized management system and was one of the pioneers in the development of the St Vincent DiabCare program. The computer database has 23771 patients registered. The Diabetes clinic works in close collaboration with the Medical School of the University of Malta through which research into diabetes is coordinated. The present research interests of the diabetes department are the epidemiology of diabetes, the use of computers for diabetes research and management and the study of nephropathy in families of people with diabetes.

All patients attending the diabetes clinic are periodically screened for diabetic retinopathy with a dilated fundal examination through ophthalmoscopy and/or retinal photography. Fluorescin angiography and Ultrasonography are available but are not routinely indicated as part of the examination of patients with diabetes. Routine screening is done either in the diabetes clinics by trained doctors or in the ophthalmic units at the central or peripheral clinics. Patients are referred to the ophthalmic units as required In particular, patients with sight threatening diabetic retinopathy: proliferative DR and advanced diabetic eye disease (Vitreous haemorrhage, retinal detachment and/or rubeosis iridis) are referred immediately and patients with preproliferative diabetic retinopathy or diabetic maculopathy are given an early referral. Patients with lesser degrees of retinopathy are referred accordingly. The government also offers a free service for screening for glaucoma. Laser photocoagulation and vitreous replacement are available in Malta.

A foot referral clinic and a peripheral vascular surgery unit are available in addition to the following podology services: (i) daily podology services are available at five peripheral health centres and at Gozo general hospital. The latter provide a service for all patients, including those with diabetes. Patients are seen by appointment but urgent cases are seen as required. (ii) Three podologists are available daily at the main diabetes clinic at Mater Dei hospital. Cases are referred by doctors at the diabetes clinic, by general practitioners, and from the Health Centres. Patients can call in directly to the clinic for advice. Over 400 new and over 5000 follow-up visits are made yearly at the podology clinic at Mater Dei Hospital. An orthotics and prosthetics service is available. in the diabetes clinics perform routine foot assessments and treat common foot lesions, provide education in preventative and therapeutic foot care, and refer patients to appropriate clinics where necessary.

A weekly joint Obstetrics - Diabetes clinic is held at Mater Dei hospital. About 75 new and 130 follow up yearly visits are held in this clinic

Diabetes mellitus is a relatively common condition in the Maltese Islands, affecting approximately 4-7 new paediatric patients per annum. These contribute a significant workload to the diabetic clinic and the paediatric wards (where any in-patient care is carried out) and utilise health care resources on a regular basis. Indeed, up to 10-15 cases with diabetic ketocidosis require admission per year, with children ranging from 10 months to 14 years of age. A Paediatric clinic is held at Mater Dei diabetes clinic every week. The children are managed by a paediatric diabetologist. Children requiring admission (often due to DKA) are accommodated in the paediatric wards and management the admitting paediatrician and patient's diabetologist.

All patients with diabetes need regular and periodic dietary advice. The present service needs to be improved at present only two dieticians are working at the diabetes clinic at Mater Dei on a part time basis.

Patients with Type 1 Diabetes are given individual education together with patients and relatives. Group education programs for Type 2 Diabetes patients are held at the diabetes clinic at Mater Dei and periodically at the peripheral centres. Staff from the Diabetes Clinic Team contributes regularly to the educational program of the Maltese Diabetes Association held monthly. Members of the Diabetes unit are involved regularly in various radio and television educational programs aimed for the general public.

Periodic educational programs are co organized with the University for undergraduate and post graduate students. Periodic educational programs are held in various government and non government schools.

All patients under the age of 36 are provided with 2 blood glucose strips per day. Pregnant patients are given 4 strips per day. All patients with diabetes are entitled to free medications for care of their condition.

All patients have easy access to laboratory and other investigative procedures as well as to regular ophthalmology, cardiology, nephrology, chiropody, dental and orthopaedic, psychology, social work and other services. Kidney transplant facilities are available.

Diabetes Information Systems

There is a unique diabetes information system active in Malta: data is either entered into the computer system directly by whoever is seeing the patient or more commonly written down manually and later entered into the system by a data entry operator. Entry is password protected.

Patients can be identified either by ID card Number, Passport Number or Surname and First Name. Any patient who has been registered on the hospital Patient Administration System can be accessed directly by the program. On initial entry the patient's Name, Surname, ID Card Number, Age and Sex are shown. Once the appropriate data has been entered, subsequent entries into the system will display in addition the Type of Diabetes, Waist circumference, Height, BMI and the number of years the patient has been diagnosed with Diabetes. This data will subsequently appear at the top of any screen opened during the patient's visit. Each screen allows entry of free text by the operator. When one enters data one has the option of creating a first clinic visit, a Follow up visit or a Yearly Review Visit.

Entry into the system allows access to a number of screens. They allow continuous follow up of the following characteristics:

1. Status of the patient. Visit Routine Parameters, Visit Type, Results of any Oral Glucose Tolerance Test performed, and the Clinical Status of the patient i.e. The Type of Diabetes if any, IGT, IFG, No diabetes, or Defaulter.

2. Clinic visit routine and drugs the patient is taking. This screen has four sub forms with the following data:
   a. Date of Visit, Height, Weight, Waist. circumference, BMI, BP, Fasting or Random. Blood Glucose, Hba1c, Urine analysis.
   b. Current Diabetes Medications and Dosage.
   c. Current Non Diabetes Medications and Dosage.
   d. Free text entry.

3. Patient Diabetic History. Year Diagnosed, Family History of Diabetes and Current Smoking and Alcohol Status, self monitoring blood glucose status, whether eyes and feet have been examined over previous year.

4. Symptoms related to Diabetes Complications. Peripheral neuropathy, intermittent claudication, Impotence, Postural hypotension, Angina, Autonomic Neuropathy, Current or Past Foot Ulcer or Gangrene.

5. Medical History. Any medical condition that the patient may have suffered from with the Year the condition was noted and Type of Management i.e. Medical, Surgical or Other. A sub form lists the current non diabetes medications the patient is taking.

6. Physical Examination. This screen allows the entry of physical examination data. A sub form data sheet lists the results of all previous physical examinations.

7. Complications. The complications of Diabetes and the Year noted.
   a. Neuropathy. Any abnormal sensation, the presence of Diabetic Mononeuropathy, polyneuropathy, Amyotrophy or autonomic neuropathy.
   b. Peripheral vascular disease. Diminished or Absent Foot Pulses, Ulceration or Gangrene, Bypass or Angioplasty, Above or Below Knee Amputation.
   c. Nephropathy. Microalbuminuria, gross albuminuria, the presence of renal failure and whether the patient has had renal dialysis or a transplant.
   d. Cerebrovascular disease. A history of Cerebral Stroke or Transient Ischaemic Attack and Year Noted.
   e. Coronary Artery Disease. Angina Pectoris, Myocardial Infarction, Angioplasty or Coronary Artery Bypass Surgery.
   f. Retinopathy. Visual symptoms, History of Glaucoma, Best Visual Acuity, the presence of and grading of Diabetic Retinopathy, Photocoagulation therapy, and the presence of Legal Blindness.

8. Dietary advice. Date of visit, the Diet Type prescribed

9. Laboratory Investigation results. Lipids, Creatinine, Haemoglobin, White blood, Count, Platelets, and Electrolytes, Thyroid Function Tests, Electrocardiogram, X-Ray findings and results of any other tests.

10. Education topics dealt with and Name of Educator.

The system is very stable and broadly used, allowing a straight export of data to all major database formats. The system is capable of exchanging data between the Maltese database and the BIRO common dataset.

4.1.4 State of the art in Romania

Health care system

Currently there is an estimated figure of over 500,000 diabetics living in Romania. Although the majority of individuals with diabetes have their first contact with a primary care doctor (family practitioner, internist, paediatrician, obstetrician, etc.) almost all diabetic patients are initially seen and regularly followed by a specialist (a diabetologist).

The Ministry of Public Health (MPH) coordinates, through Regional Health Authorities, and county offices, the public healthcare policy for all types of health units, public or private. MPH is responsible for public health, health policy, regulations, health programmes, and investments in public health establishments. However, some ministries have their own healthcare network. The national health programmes are administrated by a National Agency for Health Programmes, in the framework of the Ministry of Public Health.

As in many countries, diabetes management is a systematic approach to detecting and treating diabetes using practice guidelines to provide the following: (i) systematic approach for clinical decision making in the treatment of type 1 and type 2 diabetes and its complications; (ii) consistent set of scientifically based practice guidelines that can be adapted by each diabetes centre, according its resources. Across Romania there are 40 diabetes centers; and (iii) identify appropriate criteria for therapies during three treatment phases: startup, adjustment and maintenance.

The human resource crisis in low-resource country is merely acknowledged. There is discussion among professionals and decision makers of possible solutions involving long-term human-resource policies for training and retention of required health workers, who face normally excessive workload and lack of upgrading.

Ministry of Health has in its program a higher involvement of primary care specifically trained and skilled in comprehensive first contact and continuing care of type 2 diabetes.

Romania has adopted action plans specifically for diabetes that also include primary prevention approaches. Furthermore, the Ministry of Health has selected the prevention of type 2 diabetes as a focal theme for Romania in the first half of 2007. However, primary

care providers need a carefully developed set of practice guidelines that can be realistically implemented and shown to be efficacious and cost-effective. Also, primary care includes health promotion, disease prevention, health maintenance, patient education, diagnosis and treatment of acute and chronic illness in a variety of health care settings.

Current approaches to policy and management of medical assistance for diabetic patients in Romania is based on the recognition of the benefit of near-normal levels of glycemic control. It is well understood that a continuous relation exists between glycaemic control and the incidence and progression of microvascular complications, as well as the adverse effect of hypertension and smoking on microvascular outcomes. Few fundamental studies have produced a sharp attitude on the control of glycemic goals, improving long-term outcomes in the diabetic population: the UK Prospective Diabetes Study (UKPDS)[8] and controlled clinical trials such as the DCCT[9].

Nowadays, any improvement in glycemic control is seen as a means of preventing or slowing the progression of microvascular complications, with potential benefits in terms of reduction of the risk of macrovascular disease. Maintaining glycaemic levels as close to the non-diabetic range as possible has been demonstrated to have a powerful beneficial impact on diabetes specific complications, including retinopathy, nephropathy and neuropathy for both the type 1 diabetes and the type 2 diabetes.

National diabetes programs have undergone clinical trials and implementation studies, favouring results leading to changes in the original design of diabetes management in Romania. Annually (usually in May), experts in diabetes from Romania and around the world are being gathered in the National Diabetes Congress, normally taking place in renowned university centres. This Conference is regarded as a high level meeting point, where many aspects of diabetes care and research are jointly discussed and networked more directly across disciplines.

An important contribution in diabetes research in Romania is carried out by the diabetes university centres, among which the Institute of Diabetes, Nutrition and Metabolic Diseases “Prof. N. C. Paulescu”, the Romanian partner of the BIRO project, plays an important role.

Paulescu is a public body offering multidisciplinary clinical care for the patients of all ages and with all type of diabetes (60000 cases only in Bucharest).

The institute performs basic and clinical research activities in the field of metabolic diseases. Major areas of interest are: neuro-electrophysiology, nutrition, genetics, epidemiology and immunology related to metabolic diseases.
As a university and research institute in diabetes, it is a very highly rated institution, with a very strong scientific basis, providing leadership in the collation of diabetes information.

The annual activity of the Paulescu Institute consists of medical assistance for diabetic patients in Bucharest (78800 diabetic patients), graduate and post-raduate training in Diabetes, Nutrition and Metabolic Diseases; and research in diabetes and metabolism, immunology, genetics, epidemiology, electrophysiology, education and psychological support.

Diabetes information systems

In 1997, a set of programs written in Epi-Info were used for the first time to create a Diabcare database in Bucharest and 10 other counties of Romania. The Vincent Diabcare data collection service was initially organized through an e-mail network[10] linking university teaching with diabetes care centres (Bucharest, Brasov, Braila, Buzau, Craiova, Constanta, Galati, Suceava, etc.) through the same interface for data collection.

After this initial experience, three other successful diabetes information systems were created to exchange data across clinical care units. New interfaces were designed for health care providers, to directly interact with patients as a normal component of health care provision. The main problem of these applications was represented by the use of routine procedures that were not always maintained properly by clinical units.

The first successful medical record system was realised under the banner of the EU project “Black Sea Tele Diab” (BSTD), based upon standard e.g. the Good European Health Record (GEHR) and the CEN ENV 13606 on "Electronic Communications" and EU research projects e.g. Inco-Copernicus[11]. The project BSTD was coordinated by Sheffield University, Hallamshire Hospital (UK), with software quality led by The Hull University (UK). The system was developed using a modular design and an object-oriented method. The GEHR contains the set of concepts dealing with co-operation between healthcare providers around the care of a patient. Several concepts dealing with care plans and clinical pathways were also defined. The CEN ENV 13606 defines in a generic way EHR-system components, their inter faces and behaviour to provide a useful formalism for reconciling and re-using detailed data specifications across different use cases[12-14]. The CEN ENV 13606 consists of four parts: "extended architecture", "domain term list", "distribution rules", and "messages for the exchange of information".

The meta-data model used by BSTD complied with WHO-Europe Diabcare dataset (Vincent initiative). The system was written in C++, using Microsoft Access as a DBMS and including the following functions: Patient Records (EHCR), Clinical Protocols, Reports and Statistics, Graphs, Data Communication and System Administration[15].

The main objective of the project was similar to BIRO, i.e. enabling a quality improvement cycle of diabetes services through better monitoring of clinical care in Romania and countries of the Black Sea area.

The Consortium created a diabetes health record system based on the GEHR model where diabetes

was considered as a prototype for development of a model for chronic disease management. The “patient records” function allowed the registration of demographic information for a new patient, the recording of a new Data Sheet, corrections and the ability to search for and view a sheet.

Like any HL7's approach, the GEHR approach of BSTD deployed an object model defined according to the standards of the unified modelling language (UML), consisting of two parts: the concrete GEHR Object Model delivering the EHCR information container, and the GEHR meta-models called Archetypes, to express the clinical content.

Such architecture provided a common data structure for electronic health care records (EHCRs), taking into account ethical, legal, security and educational requirements. A wide range of data types including quantitative results (measurements), semi-quantitative results (low, moderate, high) and qualitative results (yes/no) were also considered.

The function “system administration allowed recording unique individual identifiers; tracking providers using the system; managing passwords and access rights for users; managing units and ranges used for measurements; managing the interface of the system; exporting data in CSV format; following clinical protocols.

The BSTD system was released as free software and underwent formal clinical evaluation in diabetes centres from Romania, Ukraine and Moldavia. It was finally implemented in clinical practice in few diabetes centres, including Paulescu and the Diabetes Ambulatory Centre and Diabetes Centre Hospital “Malaxa” in Bucharest.

Experience gained during development of the BSTD has been extended to develop the first Romanian Diabetes Register adopting the EHCR: SincroDiab. This software tool allowed central longitudinal management of diabetes episodes, shared across a range of providers.

Main objectives of Sincrodiab were: (i) To ensure quality and integrity of medical records at the clinical setting; (ii) to avoid redundancy and duplication of records; (iii) to develop an operational database including all episodes and procedures for diabetic patients. Same function in BSTD were also developed for SinchroDiab.

As an electronic healthcare record, SincroDiab represented a means of documenting the care process for an individual to manage information concerning the delivery of health care.
The software was written to deal with sensitive, personal medical data and complied with the major architectural features of the CEN standards: GEHR and prENV 13606 - Health informatics - Electronic healthcare record communication, and the codes of conduct of data protection legislation in Europe e.g. the Data Protection Directive[3].

The system was implemented successfully in clinical routine as free software.
In 2006, a new product called “Hipocrate” was introduced by the Romanian Ministry of Health, followed by another product called Rodiab. Different approaches were introduced during the last years in Romania, raising a problem of consistency in the exchange of health records, that may be resolved by a new standard. Currently, the Romanian framework appear still too heterogeneous to allow a rapid introduction o f common systems and interoperable services. The result until now has been poor use of health data. Health information systems in diabetes tend to be fragmented, inaccurate, cumbersome, untimely, and isolated. There might be substantial barriers for an extensive implementation of a common approach e.g. the one advocated by BIRO.

The vast majority of software development tools used in the diabetes sector today do not inherently support data exchange mechanisms. In generally software developers in the field in Romania typically have limited experience implementing data exchange protocols that would allow a rapid estimation of agreed indicators.

Furthermore, there are implicit business incentives for limiting the interoperability of different health information systems, leading to a diversity of attitudes and mix of skills that must be faced in the implementation of new protocols.

4.1.5. Adaptation, implementation, set-up and use of BIRO technology

Technology Workshop

Across three years, the main focus of technology development in the BIRO consortium has been the production of source code allowing to use local data for the purpose of producing rapid statistical reports according to shared criteria.

The BIRO Technology Transfer Workshop was held on 17th of January 2009 at the BIRO final technology meeting in Bergen, Norway.

The approach was very practical, oriented to simulate actions that could be repeated on complete datasets after the installation of the software at each site. The test provided important suggestions on how to improve data processing procedures and user-friendly functionalities.

Installation of the software

To use BIRO, it requires a single location to install and host the software - Java, DBMs (Access, SQL variety: MSSql, MySQL,Postgres), ODBC drivers, etc..

A short user guide was circulated by the Coordinating Centre to explain how to proceed to software setup. The test took place on notebooks owned by partner institutions, mainly running on MS Windows. Required software, e.g. the Java Runtime Environment, the Postgres database, the statistical software R and the

typesetting framework MikTek, were downloaded and properly installed where needed. Some adjustments were needed to define environment variables and install a BIRO local database for each test dataset. Finally, software developed in-house operated by the "BIROBox GUI" (database and statistical engine, comunication software) was also installed.

This session was completed by each partner without particular difficulties.

Data export and the BIRObox

All partners contributed with test data and configured own database accordingly, to be prepared for the export of a standardized Postgres BIRO database.

At this point two possible options were explained. It was either possible to use the Adaptor through a character based interface, using a configuration file that would provide indication of the drivers and other specifications.

To configure the Adaptor, a user needs to specify the SQL query that retrieves all the required data and exports it to one table. Commands are grouped into scripts. Programming proficiency is necessary to undertake these steps and to write commands, which may represent a major drawback for the usability of the first version of BIRO.

To facilitate high level users in data export towards the BIRO format, a graphical user interface (BIRObox GUI) was produced, allowing direct field-mapping configuration of the BIRO Adaptor.

The result in both cases is porting local data towards an XML standard BIRO file, based upon a predefined schema. All software has been written in the platform independent programming language Java, amenable to be installed on both Windows and Linux operating systems.

All participating appreciated the construction of the BIROBox as an instrument that would integrate all BIRO software tools and improve the usability and extendibility of the BIRO system. For this reason, the whole Consortium agreed to focus on its use on real databases in Bergen.

Users recognized that the Box is highly strategic, as it will represent the way BIRO will be directly presented to the user.

Participants experienced five different functionalities of the BIRO box:

1. *User interface for mapping the file presented/created by the partners (csv file etc) to BIRO fields and their relative measurements/values.*
On this front the final product demonstrated to work sufficiently well. It was acknowledged that it still needed fine tuning, as results were obtained only after frequent breaks due to errors and a series of trial and errors. New additions of field values seem to be hard coded rather than read from a table. The end product contains records with one patient id and his respective visits, where each file carries the necessary tags to identify region and date of upload as well as the period of time covered.

2. *User interface to export data and produce xml files.*
Steps involved in exporting a subset of data from a database to XML are: connect to the database; specify the SQL to run to retrieve the data; specify the location of the flat file (XML); and data export.
It was suggested that the run on the csv file is done in a way that records not passing criteria requested are stored to an error log file, rather than having a program break and restart at each and every problem encountered. Records may than be filtered by the location controller or be discarded if necessary. The idea that the results are saved in an xml file was questioned and discussed in detail and finally agreed as a possible solution.

3. *Create BIRO database. This step was successful for those steps with no errors in the creation of XML.*

4. *Statistical Engine.* This step was not carried out during the workshop.

5. *Reports. Same as previous.*

On installing the Box, the following folders are also created:

• folder to hold all xml files: one file per patient to include all his visits and relevant data for that visit. Subsequent downloads of data will overwrite previous and a copy may be taken to another subfolder.
• folder to host local (country/region) database aggregated data. Sub folders will hold subsequent runs.
• folder to host local (country/region) statistics results. Sub folders will hold subsequent results.
• folder to host results that will be accessible by Biro central. Possibly shared/accessible by BIRO central.

The workshop highlighted that the Box may increase BIRO usability particularly through mapping from the original items, that can be effectively visualised: each mapping can display a help, with a description of fields being selected, which is helpful for users.

However, the user needs to know in advance how to merge local files with the export, which implies an in depth knowledge of both the original and the BIRO common dataset.

This implies that to some extent the user should also be medically proficient, to understand classifications and data coding. The motivation of users in the optimal use of this mapping tool remains an issue that must be resolved by identifying the ideal professionals to be targeted by BIRO for its continued use.

Fields mapping has been tested on a variety of data that are of common use in diabetes. Usability of the Box has been tested directly on databases, looking at different measurements and ideal cut offs for classifi-

cation for continuous variables, e.g. Systolic blood-pressure (low:< 90 mmHg; normal: 90 mmHg - 140 mmHg; high: > 140 mmHg), or other situations in using counts with integer values, boolean variables for true/false e.g. retinopathy, e.g.). Inconsistencies have been noted in the configuration of fields mapping, compared to the BIRO data dictionary, which will be resolved in an improved version.

Further tests are required, based on different mapping scenarios, to better understand the needs of the prospective users.

Customised approaches

The nature of diabetes data is multidimensional and thus it can can become extremely complex. The BIROBox may not be necessarily adequate to fit all needs, particularly those of advanced users that may want to process many tables/databases at the same time, which would require sophisticated procedures to map/merge records before exporting them to the BIRO format.

In these cases, the organization of local data may require using powerful third party tools independently from BIRO, made available either as a commercial solution (e.g. Business Objects) or open source software (e.g. Pentaho).

There may be inevitable problems in the extensive adaptation of the BIROBox to customised needs, as it may not necessarily fit heterogeneous databases where clinical and administrative data can be simultaneously included to cover a wide spectrum of procedures related to health care management.

More collaboration between users and software developers is required to fill the gaps and sort out the majority of issues.

As a final resource, it is always possible for users to deploy own solutions complying with BIRO specifications. This would not require using the Box, and it would allow still contributing to the EU information system. This would still fit the scenario of a sustainable EU diabetes register, where different range of users may coexist and cooperate at the same time.

4.1.6. Country tests

Results of BIRO application on data from Cyprus

The high level of compliance of the Cyprus data with the BIRO Common Dataset is a very special condition not to be found easily in European diabetes registers. In fact, the Cyprus dataset was developed precisely on the basis of BIRO definitions, and most fields can be directly used to compute BIRO indicators.

The most challenging task during the BIROBox trial on Cyprus data was the set up of the connection with the local data source. The test dataset being provided, as well as the whole diabetes register in Cyprus, was developed using Microsoft Access.

Since no JDBC Driver for Microsoft Access was available at the moment of the Bergen meeting, the Microsoft Access dataset was directly exported as a CSV file (comma separated value), directly used as source for the BIROBox. An alternative connection may be implemented using ODBC, so to establish a JDBC-ODBC bridge that would be implemented to link the BIROBox to the ODBC source.

After the correction of some errors in the format of the exported CSV file (e.g. missing dataset header, unexpected delimiter, etc), the BIROBox successfully read the data source.

The Adaptor configuration phase, especially the file mapping between the local dataset and the BIRO dataset, was carried out very easily and rather quickly.

The most problematic phase of the test application, i.e. the set up and connection with the data source, was successfully passed.

Results of BIRO application on data from Malta

The National Diabetes Register, updated daily, uses a tailor made application in VB6 with an SQL2000 database backend.

The diabetes dataset brought to Bergen consisted of a very small set of 100 records from the Diabetes Clinic Mater Dei Hospital in Malta. The dataset showed to include all fields necessary for the BIRO project.

Certain fields required to be narrowed down to fit the BIRO specifications, as in the following cases:

BIRO Field:	Retinopathy.
BIRO Coding:	0 (no), 1(yes) 2(Referral) accepted.
Malta Coding:	Captures values for left and right separately.
Solution:	If any of the fields (L/R) is filled, marked as positive.

BIRO Field:	Diabetes.
BIRO Coding:	2(no), 1(yes) Accepted.
Malta Coding:	1,2,3 or unknown.
Solution:	Need to specify if other values (3/Unknown) skipped/ignored or gathered with one value or the other.

BIRO Field:	Oral Therapy
BIRO Coding:	Five different values accepted
Malta Coding:	Needs to be configured differently or grouped.

BIRO Field:	Smoker
BIRO Coding:	0 (no), 1(yes) 2(x smoker)
Malta Coding:	0 and 1.

BIRO Field:	Injections.
BIRO Coding:	Number per day.
Malta Coding:	Needs to be configured or calculated.

BIRO Field:	Self monitoring.
Malta Coding:	Field only added to database recently.

Apart from problems outlined above, the phase of setup and creation of the BIRO database was successfully passed. A test run of the statistical engine denoted that many extreme situations occur with a limited sample, indicating bugs in the R code that can be overcome with appropriate checks, and data inconsistencies that may occur by default in the creation of diabetes indicators from small samples.
The particular application seems to suggest that minimal sample size requirements may be needed for the practical use of BIRO.

Results of BIRO application on data from Romania

The institute “Paulescu” in Bucharest presented a MySQL sample database with a single table taken from diabcare data with 4,627 patients collected from few countries in Black Sea region. Data from multiple data sources had to be concatenated and be converted to MySQL database format.

In this case the Adaptor was used in its character based version, successfully connecting original data to the database. Once written, a local configuration file was run in a completely automated way, outputting the XML file, which was then exported to the BIRO central database.

4.1.7 Discussion

The activity of technology transfer allowed to describe practical problems encountered during the installation and use of BIRO software by real users.

Computer applications can be of great scientific importance, particularly in diabetes, as the data collected aims to provide insight on the quality of health care, and improve efficiency through the analysis of organizational problems experienced by medical units.

However, the provision of reliable and comparable data needs to take into account various difficulties.

Most diabetes electronic record applications rely on databases as the persistence mechanism. Standards are fundamental to data sharing, integration, secure gathering and information exchange.

There is a general lack of interoperability across different data formats, architectures, and naming conventions in diabetes that result into heterogeneous sources at the European level. Bringing together data with different patterns, structures and syntax, and allowing users to access information located in different places can be a challenge. In this respect, the BIRO system sets the basis for a new form of collaboration.

Various tools must be available to implement the solution and allow users to interact on the whole cycle, from the collection of the individual parameter to the interpretation of diabetes indicators.

Mapping functionalities may be paramount to make fuller use of local data that in many cases is still of dubious quality and based upon heterogeneous definitions (even within the same database).

More training is required to fully exploit the potential of medical records, everywhere.

Technology transfer in BIRO was key to the purpose of fitting the system to the immediate needs of its users. This process must be repeated again in order to progressively advance the agenda for a wide implementation.

Practical decisions have been derived for the initial release of the software.

All visits from a zero would be filtered for upload each time the upload is undertaken, and the upload is planned to be made every 6 months. For instance all visits from the beginning of the year, say January 2007, would be filtered for upload in the file of June 07, Dec 07, Jun 08 and Dec 08 etc for the whole cycle, until a new zero date is agreed.

Once the data/view is created one can save the filtered records as a CSV file. One can always tweak the data after in Excel and save as a CSV file through it. The file may also be modified by the respective partners, directly to replace and remove patient ID, without undermining the integrity of the data and respecting confidentiality. Each upload will be overwriting the previous. However, a copy for backup purposes may always be taken.

The following barriers in the local adaptation of BIRO have been identified:

- clinicians are little aware of EU standards in data classification
- software does not sufficiently comply with BIRO and/or produce exports
- communication is scarce between policy makers and researchers
- systems implemented are not properly evaluated
- software products are not interoperable
- financial support for improved data quality is scarce
- multidisciplinary training across medical/IT disciplines is not encourage
- strong procedures to ensure privacy, confidentiality and security are required.

BIRO offers strong stimuli for an attractive research and training, with its base on high quality research and innovation.

Possible incentives for its uptake include:

- showcases and pilots to demonstrate BIRO innovation, using means e.g. web services for public involvement
- raising awareness about the power of structured reports as "intelligence services" for the organization of modern health systems
- public-private partnerships to understand and implement efficient BIRO IT solutions
- presentation of examples of usage of BIRO as the best practice for the evaluation and better planning of diabetes care.

4.1.8 Conclusions

The BIRO project provides policy makers and managerial levels with a platform for the use of diabetes information at Regional/National/European level.

Learning from direct users and understanding their needs is the essential foundation for an effective use of a system addressing a diverse and broad audience.

Technology transfer has successfully presented a new stream of open source tools developed by the BIRO collaboration, showing: (i) increased opportunities for innovation, quickly integrated into flexible tools; (ii) a range of practical problems, that can be rapidly solved on site through test rounds; and (iii) range of technological problems that must be addressed separately and solved through the support of a community of programmers.

The fundamental precondition for the efficient and cost-effective data sharing proposed by BIRO is the continuous training in the field of diabetes information.

The BIRO project confirms that practical applications in health care, as in any real situation, to be successful and reliable, must pass a range of serious tests of feasibility, to be continuously carried out at any stage of their implementation.

References

1. BIRO Preliminary Privacy Impact Assessment, available at: http://biro-project.eu/documents/downloads/Preliminary %20PIA%20Report.pdf
2. BIRO Privacy Impact Assessment, Results from Step 1, available at: http://biro-project.eu/documents/downloads/PIA_Deliverable_ 150907.pdf
3. Di Iorio C et al., Privacy Impact Assessment in the Design of Trasnational Public Health Information System: the B.I.R.O. Project, submitted
4. BIRO Technology Transfer deliverable, available at: www.biro-project.eu
5. Massi Benedetti M, Carinci F, Orsini Federici M, The Umbria Diabetes Register, Diab Res Clin Pract 2006; 74:S200-S204.
6. Carinci F, Orsini Federici M, Massi Benedetti M, Diabetes registers and prevention strategies: towards an active use of health information, Diab Res Clin Pract 2006; 74:S215-S219.
7. Transnational diabetes guidelines from the International Diabetes Federation (European Region) on behalf of the St Vincent Declaration Initiative of IDF (Europe) / WHO/ Europe) http://www.staff.ncl.ac.uk/philip.home/guidelines
8. UK Prospective Diabetes Study (UKPDS) Group (1998) Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 352: 837-853.
9. DCCT Research Group (1995). The association between glycemic exposure and long-term diabetic complications in the Diabetes Control and Complications Trial. Diabetes 44:968-983
10. Pruna S., Herinean Radu and Pruna. A. Set up an Internet infrastructure, interconnecting clinicians and scientists in countries of the Black Sea area to promote WHO/EURO care quality programs. Fifth Conference of the European Society for Engineering and Medicine, ESEM 99, Barcelona May 30-June 2nd, 1999,293-294.
11. Pruna S., Dixon R., Harris N. Black Sea TeleDiab: Diabetes Computer System with Communication Technology for Black Sea Region. IEEE Transactions on Information Technology in Biomedicine, Vol. 3, No. 3, September 1998, pp. 193-196.
12. Dixon R, Grubb P, Lloyd D. EHCR Support Action Deliverable 3.5: "Final Recommendations to CEN for future work". Oct 2000, available at http://www.chime.ucl.ac.uk/HealthI/EHCRSupA/documents.htm.
13. ENV 13606-1 - Electronic healthcare record communication - Part 1: Extended architecture. CEN/ TC 251 Health Informatics Technical Committee.
14. ISO 8601 standard describing formats for representing times, dates, and durations, available at: http://www.cl.cam.ac.uk/~mgk25/iso-time.html.
15. Pruna S., Georgescu M., Stanciu E., Dixon R.M., Harris N.D. The Black Sea Tele-Diab System: Development-Implementation-Clinical Evaluation Medical Infobahn for Europe, Proceedings of MIE2000 and GMDS200, 656-660. (PMID: 11187635 [PubMed - indexed for MEDLINE]).

The role of BIRO in European Diabetes Information

Joseph Azzopardi, Fabrizio Carinci

ABSTRACT

Introduction
Information on process and outcome indicators in diabetes is heterogeneous and unreliable across Europe. The lack of comparable indicators makes decision making difficult and proper control of the disease hard to organize. Governments need to identify strong solutions to push the entire health sector towards tight control of diabetes as a relevant element of a regular evaluation of health systems performance.

Objectives
An evaluation process has been included in the BIRO project to provide appropriate input for corrective actions, particularly with regards to the feasibility of data collection and use of ad hoc statistical software.

Materials and methods
Independent experts in the fields of clinical practice, information technology and epidemiology/statistics were asked to form an independent panel in order to provide comments on project deliverables for selected major work-packages. A structured questionnaire was distributed to each expert. Written feedback was taken into account and a discussion undertaken at plenary meetings to agree on corrective actions.

Results
Evaluators provided valuable input for the improvement of products delivered by the Consortium. Practical problems involved fine tuning of parameter definitions and making the data comparable over different countries. Comments focused on the composition of the common dataset, the data dictionary, and outputs described in the report template. Limitations were noticed in data export, which was perceived as dependent on the skills of designers and not sufficiently defined in the metadata. Clinical and epidemiological were invited at the 2nd Annual Meeting to examine variables included in the common dataset and report templates. The discussion led to substantial improvements in the outputs of the project.

Conclusions
Complete information on diabetes requires a significant investment straight into the clinical setting, where the diabetic patient is seen regularly, and there is opportunity to monitor the status and progression of the disease. These aspects were acknowledged by evaluators, recognizing the importance of well structured data definitions and the capability of the system to deliver timely information at the European level. The BIRO project, with its sequel EUBIROD, sets the ground for timely intervention through appropriate application of concepts examined in BIRO, focusing on the direct participation of healthcare professionals.

4.2.1 Introduction

Health care providers need information systems to monitor standards and improve patient care. The diabetes problem in particular, cannot be dealt with effectively as long as comparable data on the disease are not widely available. This information is critical in identifying disparities in diabetes health outcomes and important in implementing effective programs to improve standards of care.

The statistics on diabetes are overwhelming. In the European Union alone, the number of people with diabetes is about 50 million; and rising. Half of these people do not know they have the disease. Diabetes is the cause of serious complications and often leads to premature death. Standards of care vary widely across regions, with a considerable number of people receiving sub optimal care. The disease is expensive to treat and complications are even more costly to treat than the disease itself.

A review of the existing data shows an objective difficulty in collecting and analyzing information at the European level. There are very large discrepancies between reports, but one common message: diabetes is an increasingly threatening disease, which grows at an alarming rate.

The EUCID[1] project has identified core indicators for a new standard and clear basic targets for the collection of complete data across the EU.

Some evidence has been already reported but it must be properly integrated.

A major risk indicator i.e. body mass index increases with age and identifies about 50% of the general population being overweight in most countries, of which about half is obese. Current social trends may only worsen conditions, particularly for disadvantaged strata: it is time to act strongly with proper prevention strategies.

The incidence of Type 1 Diabetes Mellitus seems to be less urgent, but it is highly variable across Europe, flagging attention towards particular geographical areas and high risk strata. Regardless of the type of diabetes, data suggest that the rise of diabetes is increasingly affecting the younger generation with an impact of changing lifestyle behaviours.

Prevalence is rapidly on the increase, but by much, and which countries are doing worse, it is difficult to ascertain with current data.

Figures available are too variable between sources, and unstable across the years. The three fold difference found between EUCID (median=3%) and the IDF[2] (median=8.7%) is difficult to explain, but it may be due to the delivery of inconsistent figures from different surveys conducted by Member States.

High quality prevalence data, based upon standardized criteria, are crucially needed. Until then, forecasts for prevalence up to 2025 may be only taken as a possible scenario.

Mortality rates are extremely important, but once again they can be affected by very different sources, which possibly use different coding styles. EUCID concluded that, on average, a town of Europe of 100,000 inhabitants would lose 30 citizens per year because of diabetes.

As far as secondary indicators are concerned, it is worth reminding that impaired glucose tolerance has a prevalence at least as high as that of diabetes. A percentage of 10% of those subjects will develop diabetes in one year.

Clinical management can be followed up through a long list of process indicators, showing that, despite the large variations between countries and the possible data problems related, in many cases the majority of countries deviate, sometimes to a large extent, from the desired 100%.

Overall, process and outcome indicators in diabetes highlight that health systems in Europe are not optimally organized to deliver the results expected for a proper control of the disease.

At the very least, currently they cannot be made directly accountable for results that may be partial and inconsistent, but at the same time seem to indicate a deviation from optimal care that must be taken rather seriously.

Governments need to identify strong solutions to push the entire health sector towards tight control of diabetes as a relevant element of a regular evaluation of health systems performance.

An evaluation process has been included in the BIRO project every year, to provide appropriate input for corrective actions and updated plans for software development.

4.2.2 Materials and methods

In order to strengthen the BIRO framework in relation to its future use and perspectives, the project has been evaluated on various aspects that are relevant to diabetes care, including:

- current knowledge of the disease
- benchmarking and decision making at population level
- ability to monitor short term health outcomes and tracking of major risk factors
- quality of life

A total of four independent experts with clinical, epidemiological, medical informatics and legal background were asked to provide comments on deliverables from main workpackages.

The Consortium submitted to all evaluators the same limited number of general questions, inviting comments on particular aspects of project development.

Questions included the following aspects:

1. Related to the project tasks
   • tasks fulfilled according to the technical annex to the contract
   • justification of any deviations

2. Scientific quality
   • understanding and use of relevant literature
   • recognition of current developments in Europe
   • appropriateness of study design

3. Innovativeness
   • specific and general level of innovation

4. Policy relevance
   • utility for decision makers
   • consideration of relevant key stakeholders

5. Report structure
   • direct interpretation
   • appropriateness of the structure
   • presence of gaps and redundancies
   • clarity of presentation

6. Other
   • general suggestions for improvement

Two out of the four independent evaluators participated the 2nd Annual Meeting, where a reserved time was made available to directly explain their views on the project to the Consortium.

4.2.3 Results

In general, the feedback received from evaluators was very positive and provided valuable input for the improvement of all products by the Consortium.

The clinical expert believed that the innovative aspect of the project was that all partners had access to a regional database. To the best of the expert's knowledge, this work seemed completely unprecedented for routinely collected data during the process of care. He noticed that the team was keen to identify solutions in the real world and not just defining indicators to leave the collection to uncontrolled individuals in the field. However, it was acknowledged that practical problems involved fine tuning definitions and making the data comparable over different countries. By all means, the definition of indicators in this context was seen as being really innovative. The common dataset represented a first attempt to build up an infrastructure, whose optimization requires lots of discussions, as witnessed by the debate held at the 2nd Annual Meeting held in Cyprus.

On this occasion, the epidemiological expert remarked that the scope of the project was novel as it would prove the concept, first in Europe, that it is possible to coordinate collected data related to diabetes on an international scale. The magnitude of the effort is likely to reflect accurately the distribution of diabetes and its complications in Europe. Several errors were noted in the report template, and corrected accordingly. Notably, it was emphasized that the recurrent use of the term “patients” should be disencouraged, and replaced by “people with diabetes”.

The IT expert, focusing on the data dictionary, noticed the lack of an overall UML class data model for the data that BIRO is collecting and generating, coupled by the fact that data export content is constrained to the skill and knowledge of designers of the data source extraction routines, rather than by metadata. Many issues were also raised in relation to the meta-data description. It was also noted that the review of the literature for privacy impact assessment was quite difficult to follow for a non-legal person, although its contents are very relevant for the development of technology. Thus, it would be advisable to summarize the legal background and discussion, specifically for those with different backgrounds. On the other hand, the methodology for the selection of the best architecture was seen as quite appropriate for the scope, and innovative.

The legal expert in charge of evaluating privacy impact assessment was contacted close to the end of the project, so it was not possible to receive his review before the deadline for the preparation of the present volume. Comments will be included in the final report of the evaluation.

4.2.4 Discussion

Much as we know, the need for more information on diabetes is obvious. Representative longitudinal data on quality of care and on morbidity and mortality are lacking in a number of countries. There is, in fact, a considerable amount of data that is often not easily available, fragmented, or poorly presented. This information can be much better utilized with better collaboration and sharing of information.

Proper exchange and analysis of data is hampered by the following factors:

- insufficient use of information systems by clinicians and policy makers
- poor collaboration between regional and European sources
- limited application of sophisticated statistical programs in European health reports
- unavailability of appropriate software in the public domain
- problems with privacy concerns and secure data transmission

Mechanisms have to be put in place to collect and analyze comparable epidemiological evidence. Coordination and collaboration among the European Member States will be critical to this effort. It also requires a standard dataset and appropriate information technology for the analysis of data and dissemination of results.

The B.I.R.O project has demonstrated that a sustainable system is possible. It can produce routine diabetes reports in a relatively straightforward way, based upon

the exchange of aggregated data, securely transmitted from diabetes registers to European institutions.

The system is able to produce routine summary diabetes reports, including population based, standardized performance indicators advocated by other projects funded by the EU under the Public Health Program (DG-SANCO).

B.I.R.O can help making comparable data on diabetes becoming widely available. It goes beyond the simple collection of information. It tracks the quality of diabetes care by providing periodic, customized outcome reports from different existing systems.

Comments received from evaluators, very positive in general, remark that well structured definitions, and a set priority on a few fundamental parameters, are necessary to revolutionize the way information is collected at the European level.

As a matter of fact, it would be much more important to strive for completeness and quality of few columns in a database, rather than struggling to deliver complex systems that will never be used sufficiently.

Many epidemiological measurements may be highly desirable, but they will never be fully attainable in the clinical setting on a routine basis: the patient likes to be looked after, independently from the computer.

Evaluators, based upon endless discussions on single items in the common dataset, helped to disentangle the desirable from the feasible, with a much higher dose of realism than Consortium participants alone.

During the Cyprus session it was said that even "diabetes type" may be pretentious to report. Maybe that could be perceived as too little to achieve, but feasibility still remains an issue of BIRO implementation that needs to be fully appreciated when the system is up and running in multiple centres.

On the other hand, an external observer has noticed that once BIRO will be installed and capable of demonstrating that it could operate properly, there is nothing that would stop it from becoming not only the European Diabetes Register, but a possible global system to track diabetes worldwide.

After all, a BIRO system working for the European Commission, as it is open source, may be adapted for other uses e.g. the production of automated WHO reports and the IDF Atlas[2].

4.2.5 Future perspectives

It is necessary to translate existing political commitments, declarations and conclusions into tangible policy actions.

Diabetes is a complex health problem, requiring multidisciplinary approaches that range from health promotion and prevention to screening, diagnosis, treatment and care.

Diabetes, like other chronic diseases, needs a long-term vision and a short-term strategic approach - measuring outcomes continuously and not just meeting targets at set dates.

At the European level, collaborative projects funded by the European Commission provide a fundamental foundation for improvement:

- EUCID has defined a set of realistic and agreed measurement targets
- BIRO proposes a way to implement a system devolving the responsibility for the automatic collection of standardized information to European regions

BIRO exploits work autonomously undertaken by regions, where a "region" is not purely intended as an administrative entity, but as a network collecting health information according to a homogeneous and well defined set of standardized rules.

This definition may can eventually identify a geographical region, or even a country (typically a smaller State e.g. Malta, Cyprus etc), but in a broad sense, a "BIRO region" can be even a cluster of clinicians joining a disease management program or an epidemiological study, who can all contribute information for the EU.

The coalition of BIRO and EUCID Consortia represents a strategic extension of both frameworks. The project "European Best Information through Regional Outcomes in Diabetes" (EUBIROD)[3] started on 1st September 2009 and will last 36 months. If successful, it will allow to automatically populate a EU Report from 20 countries, and beyond, through an extension in Kuwait favored by the generous contribution of the Dasman Institute, through the use of BIRO.

EUBIROD may foster training and involvement of health professionals to improve coding and registration through the creation of a BIRO Academy.

To progress in this direction, it is important that everybody recognizes the importance of collating complete information on diabetes straight into the clinical setting. It is there that the diabetic patient is met regularly, and a practical opportunity exists to monitor the status and progression of the disease.

It is the direct responsibility of National Governments to give the highest priority to public health measurement in clinical practice, sooner rather than later, following the example of Cyprus, which alone represent an important outcome of the BIRO Consortium.

The case demonstrates that diabetes registers and a regular use of routine databases, linked to constant epidemiological knowledge from high quality studies, may offer cost-effective support to decision making, and better practice immediately.
Evidence shows[4] without informed action, the consequences of not acting on information, would be very serious for European health systems.

BIRO Software

Valentina Baglioni

This appendix reports main technical aspects and pointers for a road test of the BIRO system.

Requirements

Hardware

Average PC Desktop/Notebook with at least 1Gb RAM, >200Gb Hard Disk and ADSL connection for use with Central Server.

Operating System

The BIRO system has been successfully tested on Linux Fedora 10 and Microsoft Windows/Vista.

Software

Running the BIRO system requires the following software to be installed in sequence:

- the Java Runtime Enviroment or Java Development Kit[1] (ver. 1.6.0_06 or higher)
- the statistical software R[2] (ver. 2.8.0 or higher)
- the typesetting software LaTex[3] or its Windows equivalent Miktek[4] (ver. 2.7.0 or higher)
- BIRO software installation zip.

Software is freely downloadable from the restricted area of the official BIRO website (www.biro-project.eu), directly accessible to all partners of the BIRO network. For more information please contact the BIRO Coordinating Centre:

Department of Internal Medicine,
University of Perugia
Via E. dal Pozzo,
I-06126 Perugia - Italy
Ph:+39.075.5727627
Fax: +39.075.5727627
email: biroeu@unipg.it

User Datasets

BIRO users should have a total of six well-structured datasets available to properly run the software: the “merge table”, the “activity table”, the “population table”, the “diabetic population table”, the “site header and profile” information, the “geographical table”. All possible [data fields] and their expected formats are listed in the BIRO deliverables D3.1 Common Dataset and D4.3 Dictionary/XML Updates, which are available at the results page of BIRO web site (http://www.biro-project.eu/results.htm).

The *merge table* has the following structure:

patient_ID , episode_date , [data field] , [data field] , [data field] , ...

The columns patient_ID and episode_date represent primary keys of the table; each row of the merge table represents a specific clinical episode of a specific patient.

The table complies with the BIRO common dataset, whose structure is represented in summary in Table A.1. SQL queries on the local database may be required to obtain the merge table as a result of linking separate tables in a diabetes register.

The merge table can be part of a database or simply a CSV (comma separated value) text file.
For the database, BIRO works with any DBMS but with few being practically tested (PostgreSQL, MySQL, MS SQL), For CSV, only the following separators are allowed: “,”, “;”, “|”.

Some notes on the practicalities:

- only mandatory columns are: patient ID, episode date and type of diabetes
- there is no need to include all BIRO fields in each merge table: the import may even be carried out from a subset of fields routinely collected
- in the merge table, local field names may still appear in the columns: label export to BIRO names can be performed by the BIRO software at a later stage
- each field must be compliant with the BIRO format, e.g. dates for BIRO date fields, enumerated field for matching BIRO field, etc.
- only specific measurement units can be used within BIRO
- date formats and enumerated values may be adapted to the BIRO format at a later stage

The *activity table* shall contain information about the enrolment of patients to the centre, i.e. dates of entry and exit from the centre and related reasons (birth, diagnosis, transfer toward/from another centre, death, lost to follow-up).

The activity table shall have the following structure:

patient_ID, start_date, start_reason, end_date, end_reason

The columns Patient ID, start_date represent primary keys of the table. Two different records with the same starting date related to the same patient are not allowed. The same patient may appear in more than one record because is possible for patients to have one continuous or several disjoint periods of activity based upon their diagnosis dates, place of residency or follow-up status.

Details about requested fields are reported in table A.2.

The activity table is crucial for proper estimation of indicators, allowing the statistical engine to include only those patients who are active at a certain point in time. At the moment the activity table is not mandatory, but is highly recommended. As a minimum requirement, at least a start and end date shall be provided. If not possible, the statistical engine will consider the date of diagnosis as the starting date.

Table A.1: BIRO Common Dataset Summary

Ref.	BIRO Name	Description	Data Type	Enumerated Values/ Measurement Units
BIRO001	PAT_ID	Patient ID	String(12)	
BIRO002	DS_ID	Data Source ID	String(10)	
BIRO003	TYPE_DM	Type Of Diabetes	Enumerated	1 = Type 1, 2 = Type 2 3 = Other
BIRO004	SEX	Sex	Enumerated	1 = Male, 2 = Female
BIRO005	DOB	Date of Birth	Date/Time	
BIRO006	DT_DIAG	Date of Diagnosis	Date/Time	
BIRO007	EPI_DATE	Episode Date	Date/Time	
BIRO008	SMOK_STAT	Smoking Status	Enumerated	1 = Current Smoker 2 = Non-Smoker 3 = Ex-smoker
BIRO009	CIGS_DAY	Cigarettes per day	Integer	
BIRO010	ALCOHOL	Alcohol Intake	Integer	g/week; g/day
BIRO011	WEIGHT	Weight	Real	Kg
BIRO012	HEIGHT	Height	Real	m; cm
BIRO013	BMI	Body Mass Index	Real	
BIRO014	SBP	Systolic Blood Pressure	Integer	mmHg
BIRO015	DBP	Diastolic Blood Pressure	Integer	mmHg
BIRO016	HBA1C	HbA1c	Real	%
BIRO017	CREAT	Creatinine	Integer	µmol/l; mg/dl
BIRO018	MA_TEST	Microalbumin	Enumerated	1 = MA Normal, 2 = MA Abnormal 0 = No MA Test Recorded
BIRO019	CHOL	Total Cholesterol	Integer	mmol/L ; mg/dl
BIRO020	HDL	HDL	Integer	mmol/L; mg/dl
BIRO021	TG	Triglycerides	Integer	mmol/L; mg/dl
BIRO022	EYE_EXAM	Eye Examination	Enumerated	1 = Yes, 0 = No
BIRO023	RETINA	Retinopathy Status	Enumerated	1 = No Retinopathy, 2 = Background Retin., 3 = Referable Retin.
BIRO024	MACULA	Maculopathy Status	Enumerated	1 = No Maculopathy 2 = Referable Macul.
BIRO025	FOOT_EXAM	Foot Examination	Enumerated	1 = Yes, 0 = No
BIRO026	PULSES	Foot Pulses	Enumerated	1 = Present, 0 = Absent
BIRO027	FTSENS	Foot Sensation	Enumerated	1 = Normal, 0= Abn.
BIRO028	ESRF	End Stage Renal Therapy	Enumerated	1 = Yes, 0 = No
BIRO029	DIALYSIS	Renal Dialysis	Enumerated	1 = Yes, 0 = No
BIRO030	TRANSPLANT	Renal Transplant	Enumerated	1 = Yes, 0 = No
BIRO031	STROKE	Stroke	Enumerated	1 = Yes, 0 = No
BIRO032	ULCER	Active Foot Ulcer	Enumerated	1 = Yes, 0 = No
BIRO033	MI	Myocardial Infarction	Enumerated	1 = Yes, 0 = No
BIRO034	LASER	Laser	Enumerated	1 = Yes, 0 = No
BIRO035	HYPERTENSION	Hypertension	Enumerated	1 = Yes, 0 = No
BIRO036	BLIND	Blindness	Enumerated	1 = Yes, 0 = No
BIRO037	AMPUT	Amputation	Enumerated	1 = Yes, 0 = No
BIRO038	HYPERT_MED	Antihypertensive Medication	Enumerated	1 = Yes, 0 = No
BIRO039	DRUG_THERAPY	Hypoglycemic Drug Therapy	Enumerated	1 = Insulin Only, 2 = Tablet Only 3 = Insulin+Tablets, 4 = None (Diet)
BIRO040	ORAL_THERAPY	Oral Drug Therapy	Enumerated	1 = Sulphon., 2 = Biguan. 3 = Glucos. Inhib., 4 = Glitazones 5 = Glinides
BIRO041	PUMP_THERAPY	Pump Therapy	Enumerated	1 = Yes, 0 = No
BIRO042	NASAL_THERAPY	Nasal Therapy	Enumerated	1 = Yes, 0 = No
BIRO043	INJECTIONS	Average Injections	Integer	
BIRO044	SELF_MON	Self Monitoring	Enumerated	1 = Urine, 2 = Blood Gl., 3 = Both
BIRO045	EDUCATION	Diabetes Specific Education	Enumerated	1 = Yes, 0 = No
BIRO046	LIPID_THERAPY	Lipid Lowering Therapy	Enumerated	1 = Yes, 0 = No
BIRO047	ANTIPLATELET_ THERAPY	Anti-platelet Therapy	Enumerated	1 = Yes, 0 = No
BIRO048	DMP_ENROL	Patient Enrolment in DMP for Diabetes	Enumerated	1 = Yes, 0 = No

Similarly to the merge table, it is not necessary to map date fields and enumerated fields to a predefined format, as it could be done using BIRO.

The *population table* shall include information about the total population and mortality in the catchment area of the centre/region. In particular, the number of persons (dead or alive) should be stratified on the basis of years, age bands and gender. An example of population table is reported in Table A.3

Age bandings based upon EUCID criteria have been approved by the Consortium as listed in deliverable D4.1 (Data Dictionary) are reported in Table A.4.

The population table should be provided in form of csv file.
The *diabetic population table* refers to diabetic patients within the catchment area. The expected structure is the one shown in Table A.5.

The diabetic population table should be provided as a single csv file. It is not mandatory, but highly recommended. If not present, the statistical engine will attempt to reconstruct it directly from the merge table.

Site header and profile information must be provided directly by partners through a specific form including contact details of centre/regional referents as (Table A.6). Other information is collected through the centre profile (see Table A.7).

Geographical references are collected according to the structure described in Chapter 3.1. An example of its use for Italy is also provided in Chapter 3.4. Multiple records are required to map each NUTS level to all other sub-levels (see Table A.8).

Running BIRO

As the main point of entry to the usage of BIRO is represented by the BIROBox, here we will focus on its setup, launch, and usage.

Setup

The BIROBox setup file is included with the main BIRO software bundle. The self extracting zip file also creates the BIRO System directory structure shown in TableA.8[5].

The setup environment will contain all the source code and libraries and documentation related to the BIRO System.

Table A.2: Activity Data fields

Reference	BIRO Name	Parameter	Data Type	Enumerated Values
BIRO001	PAT_ID	Patient ID	String(12)	
BIRO049	START_DATE	Start date	Date/Time	
BIRO050	START_REASON	Start reason	Enumerated	1 = birth 2 = diagnosis 3 = transferred in
BIRO051	END_DATE	End date	Date/Time	
BIRO052	END_REASON	End reason	Enumerated	1 = death 2 = transferred out 3 = lost to follow up

Table A.3: Example of population table

year	ageband	popM	popF	morM	morF
1997	1	19356	18289	8	14
1997	2	18623	17240	3	0
1997	3	18641	17562	5	1
1997	4	19819	18511	4	2

Table A.4: Age Bandings

Band ID	Lower range	Upper Range
1	0	14
2	15	24
3	25	34
4	35	44
5	45	54
6	55	64
7	65	74
8	75	84
9	85	none

Table A.5: Example of diabetic population table

year	ageband	typedm	diabM	diabF
1997	1	1	100	90
1997	2	1	201	300
1997	3	1	343	250
1997	4	1	432	300

Table A.6: Site header fields

Ref.	Field Name	Parameter	Data Type	Enumerated Codes
BIRO101	DS_COUNTRY	Country of Origin	String(25)	
BIRO102	DS_TYPE	Data Source Type	Enumerated	1 = GP 2 = Hospital Clinic (Internal Medicine) 3 = Hospital Clinic (Diabetes) 4 = Regional Shared-data Register 5 = Regional Primary Care Project 6 = Disease Management Programme 7 = Hospital Discharge Information 8 = Insurance Programme 9 = Retinal Screening Programme 10 = Diabetes Specialist Nurse Clinic 11 = Complete National Data 12 = Sample National Data 13 = Sample Regional Data
BIRO103	DS_NAME	Data Source Name	String(25)	
BIRO104	DS_DENOM	Data Source Denominator	Integer	
BIRO105	DS_AREA	Geographical Area	Integer	
BIRO106	DS_WEBSITE	Website Address	String(50)	
BIRO107	DS_ADDRESS_1	Mailing Address Field 1	String(25)	
BIRO108	DS_ADDRESS_2	Mailing Address Field 2	String(25)	
BIRO109	DS_ADDRESS_3	Mailing Address Field 3	String(25)	
BIRO110	DS_ADDRESS_4	Mailing Address Field 4	String(25)	
BIRO111	DS_POST_CODE	Post Code of Data Source	String(25)	
BIRO112	DS_C_CONTACT	Clinical Contact	String(25)	
BIRO113	DS_C_EMAIL	Clinical Contact Email Address	String(50)	
BIRO114	DS_T_CONTACT	Technical Contact	String(25)	
BIRO115	DS_T_EMAIL	Technical Contact Email Address	String(50)	

Table A.7: Site Profile Fields

Ref.	Field Name	Parameter	Data Type	Codes
BIRO116	DS_BEDS	Hospital Beds	Integer	
BIRO117	DS_PHYSICIANS	Physicians	Integer	
BIRO118	DS_DIABETOLOGISTS	Diabetes Specialist Consultants	Integer	
BIRO119	DS_DOCTORS	Doctors	Integer	
BIRO120	DS_DSN	Specialist Diabetes Nurses	Integer	
BIRO122	DS_DMP_PHYSICIANS	Physicians Offering DMP's for Diabetes	Integer	

Table A.8: Geographical References

Ref.	Field Name	Parameter	Data Type	Codes
BIRO200	Continent	Continent (BIRO Custom Level)	string	
BIRO201	Country	Country (NUTS Level 0)	string	
BIRO202	MacroRegion	Sub-National Area (NUTS Level 1)	string	
BIRO203	Region	Region (NUTS Level 2)	string	
BIRO204	HealthAuthority	Local Health Authority (BIRO Custom Level)	string	
BIRO205	Province	Province (NUTS-3)	string	
BIRO206	DistrictUnit	District Health Unit (BIRO Custom Level)	string	
BIRO207	PostCode	Post Code (BIRO Custom Level)	string	

Some special folders needs to be highlighted as they will contain the outputs of the BIROBox:

- *BIRO/software/_de_/data*
- default folder hosting the BIRO Export XML files produced by Adaptor
- *BIRO/software/_se_/output/data*
- folder hosting local statistical objects produced by the Statistical Engine
- BIRO/software/_se_/output/report
- folder hosting local statistical report produced by the Statistical Engine

Running the BIROBox

The BIROBox can be started simply by double clicking on the file "runBIROBox.bat" located inside the BIRO folder.

Clicking on the "Configure and Run Adaptor" button allows the user to manage the Adaptor configurations and run the Adaptor on top of the selected configuration. Multiple configurations are allowed for the Adaptor because the user may have multiple data sources. The user can create as many configurations as he wish. Each configuration can be deleted, copied or edited.

Configuring the BIROAdaptor

The first step for configuring the Adaptor (Figure A.1) is setting the connection to the local data source, a database or a CSV file. If a CSV file, the user must specify the filename and the separator format. It is absolutely necessary that the first row of the CSV source file contains column headings.

In case of a database, the user must specify the connection and login details: driver type, host and port, database name, user name and password. If none of the drivers listed is suitable for the local DBMS, the user can then add a custom DBMS driver by clicking on the "+" button (see Figure A.2).
In order to create a new DBMS driver the user must specify:

- the DBMS name (e.g. "PostgreSQL")
- the JDBC driver class name (e.g. "org.postgresql.Driver")
- the URL pattern with flags for host and port, database name, database username (optional), database password (optional) instead of real values (e.g. "jdbc:postgresql://<hostAndPort>/<databaseName>")
- the absolute path of the jar file containing the JDBC driver for the local DBMS (e.g. "C:\myFolder\postgresql-8.2-504.jdbc3.jar")

The creation of a new driver requires some IT proficiency, so it should be done with the assistance of an IT expert. When clicking on the "next" button, the connection is tested: ff something goes wrong and is not possible to establish a connection to the data source, then the user must stop configuring the Adaptor and attempt to repeat the operation with different options.

In the second step, the user must configure the mergetable (Figure A.3). If the data source is a CSV file or the mergetable is already present in the local database, the user has to select the first option and write the mergetable name into the text field. By clicking the button on the right the user may inspect the local database and obtain the complete list of the tables. He/she may choose to import the whole mergetable, or just a subset of records, through the definition of the start date and end date for episodes. If the mergetable is not present in the local database, the user can create it by writing the appropriate SQL query in the text area.

In the third step, the user has to configure the activity table (Figure A.4) by specifying its name or the query to obtain it.

In the data source configuration panel (Figure A.5) the user has to input static information about the centre (data souce ID, clinical and technical contacts...) and about the catchment area (total population, number of diabetologists, nurses, doctors...).

Mapping local fields to BIRO fields is probably the most complex aspect of configuring the Adaptor (Figure A.6). All BIRO fields are listed on the left side of the panel. For each of them, the user must specify if the field can be extracted from the local source, the name of the local column and the local data format. Mandatory fields are highlighted with a bold red font. Fields to be extracted are highlighted in bold. There

Table A.9. BIRO system directory structure

```
BIRO
    software
        lib
            html
            latex
            maps
            db
            R
                source
                    biro
                    packages
                        linux
                        pdf
                        vignette
                        win
        _de_
            data
                <centre_id>
            source
                gui
                adaptor
                postgres
                xml
        _se_
            output
                <datetime>
            data
                <year>
                <datetime>
                episode.csv
        source
            R
                backup
                formats
                include
                main
                scripts
                BIRO_se_run.r
        output
            data
                <year>
                <datetime>
            reports
                <datetime>
        _cs_
            source
                <datetime>
        _ce_
            data
                <year>
            source
                R
            output
                data
                    <datetime>
                reports
                    <datetime>
    web
        html
    docs
        guides
        deliverables
            wp7
                d7.1
                d7.2
    misc
        video
        audio
<centre_id>
    patient.csv
        pop.csv
        mortality.csv
<year>
    <centre_id>
<year>
    <centre_id>
<year>
<centre_id>
<year>
<year>
<centre_id>
    <local_comp.csv>
    <cum_comp.csv>
<year>
<year>
<local_comp.csv>
    graphs
    tables
    html
    images
    pdf
    report.aux,.tex,.toc,.pdf,.html
<centre_id>
<centre_id>
    <local_comp.csv>
    graphs
    tables
    html
    images
    pdf
    report.aux,.tex,.toc,.pdf,.html
<statobjects>
graphs
tables
html
images
pdf
report.aux,.tex,.toc,.pdf,.html
```

Box 3.3.4: BIROBox – Example of BIRO Export XML file

```
<?xml version="1.0" encoding="cp1252"?>
<!-- BIRO Export File -->
<ECDataExport xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <Patient>
    <Profile>
      <ProfileFieldName>DOB</ProfileFieldName>
      <ProfileFieldValue>1935-07-19</ProfileFieldValue>
    </Profile>
    <Profile>
      <ProfileFieldName>PAT_ID</ProfileFieldName>
      <ProfileFieldValue>1165</ProfileFieldValue>
    </Profile>
    <Profile>
      <ProfileFieldName>SEX</ProfileFieldName>
      <ProfileFieldValue>1</ProfileFieldValue>
    </Profile>
    <Profile>
      <ProfileFieldName>TYPE_DM</ProfileFieldName>
      <ProfileFieldValue>2</ProfileFieldValue>
    </Profile>
    <EpisodeData>
      <EpisodeDate>2003-07-01</EpisodeDate>
      <Data>
        <EpisodeFieldName>HBA1C</EpisodeFieldName>
        <EpisodeFieldValue>73.0</EpisodeFieldValue>
      </Data>
    </EpisodeData>
  </Patient>
</ECDataExport>
```

are three types of fields, each one requiring a different mapping.

- Date fields: mapping is done by choosing the date format in use in the local data source
- Enumerated fields: for each enumerated value the user has to write the correspondent value in local data source. Several choices are possible: null value, any string, null or any string, regular expression, custom text.

- Numeric fields: the user must choose unit of measurement adopted in the local data source. Simple fields, like patient ID and BMI, don't require any mapping.

The last step of Adaptor Configuration is the output file setting (Figure A.7): the user can select where to save the ZIP file produced by Adaptor. As previously described, the default folder is:
/*BIRO/software/_de_/data*

By going back to the Adaptor configuration manager panel and clicking on the Run Adaptor Button, the process is started and a progress status window is showed. At the end of the process, an XML file per patient is created (Box 3.3.4) and all XML files are compressed in a single big ZIP file.

The configuration of the BIRO local database is necessary to run the BIRODatabaseManager or the BIRO Statistical Engine (Figure A.8). The BIRO database configuration is similar to the data source connection configuration: the same url (DBMS Driver, database name, host and port) and login details (username and password) are requested to the user.

The Statistical engine requires few data to be configured (Figure A.9):
- centre ID
- current year
- start date and end date (time interval for data analysis)
- reference date
- population file and diabetic population file (CSV)

The latter are fundamental files containing the total population and the diabetic population in the catchment area for the correct calculation of population based indicators. In particular, the total population should be stratified by year, age band, gender (see Table 3.3.1), while the diabetic population requires strata by year, ageband, type of diabetes and gender (see Table 3.3.2).

Every time the user runs the StatisticalEngine, a statistical report is produced in html and pdf format, stored in folders named with the current timestamp.

The Communication Software panel (Figure A.10) shows the complete list of statistical objects created by the Statistical Engine to be transferred to the central engine. For each statistical object, the creation and last sending date are duly specified.

When the user select one of the statistical objects in the list and clicks the "send" button, the Communication Software creates a compressed folder, sent it to the central server where it will be decompressed and permanently stored in the central database.

The Central Server currently does not have any special GUI tool available and is only operated by the BIRO development team.

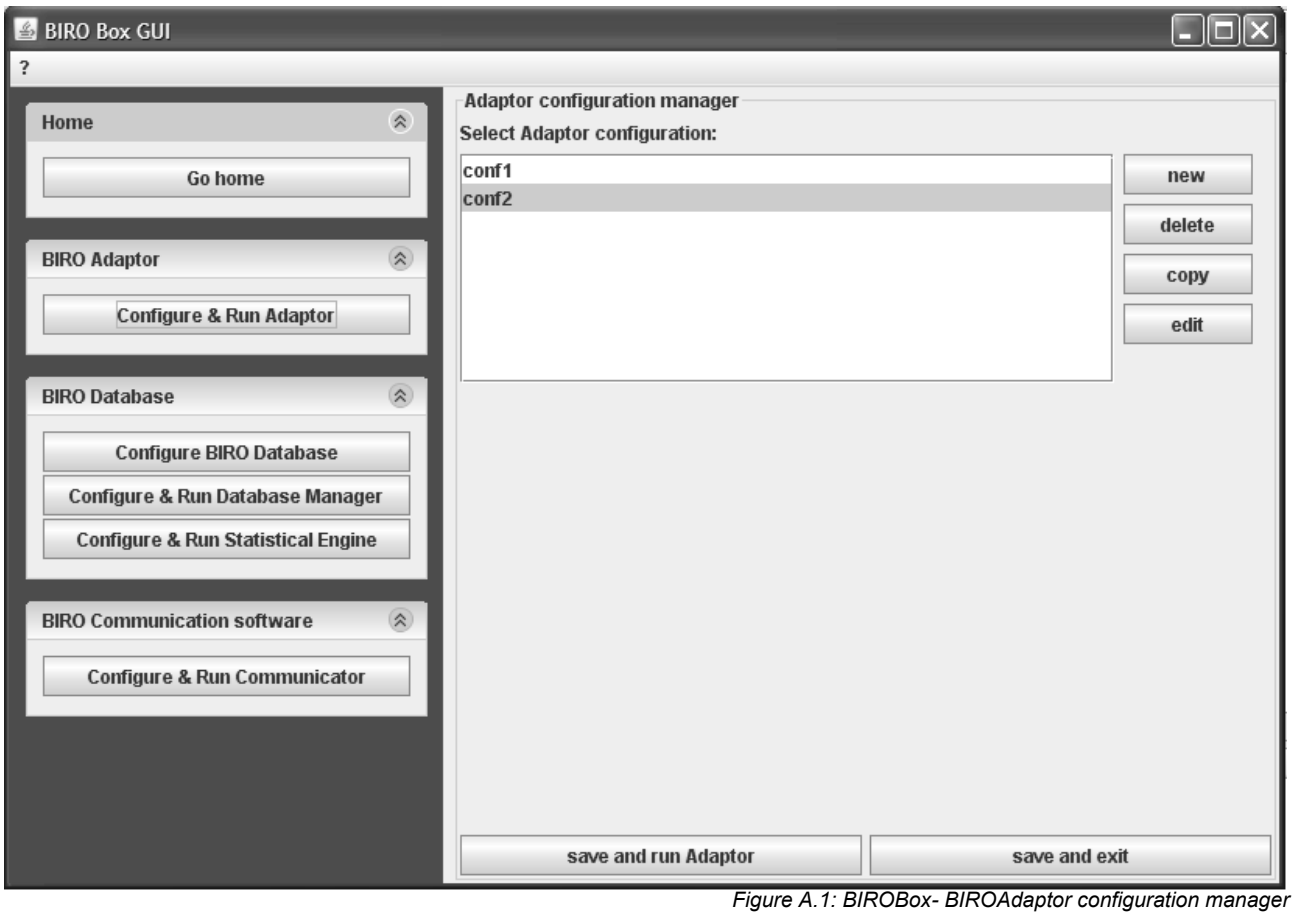


Figure A.1: BIROBox- BIROAdaptor configuration manager

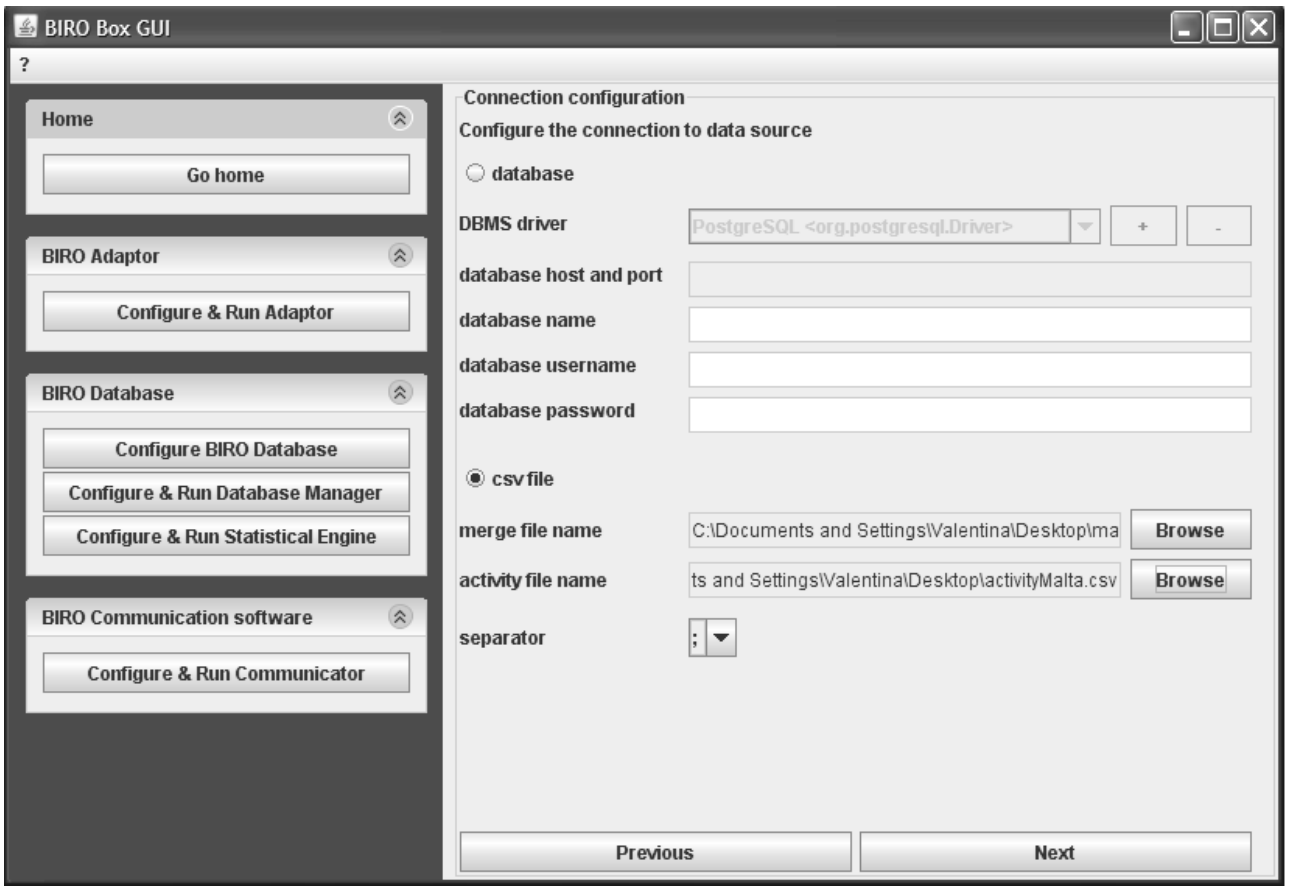


Figure A.2: BIROBox- BIROAdaptor connection configuration

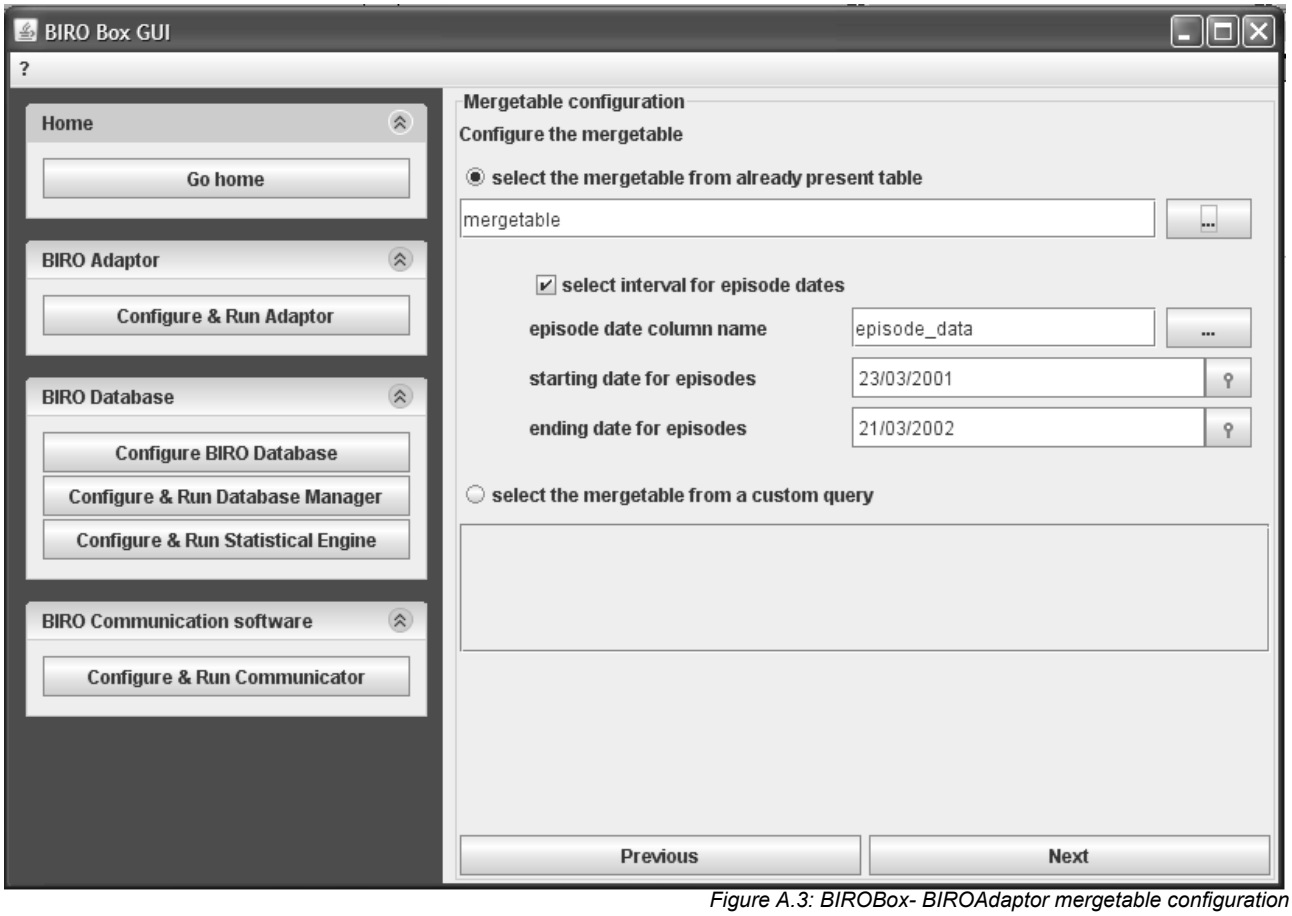


Figure A.3: BIROBox- BIROAdaptor mergetable configuration

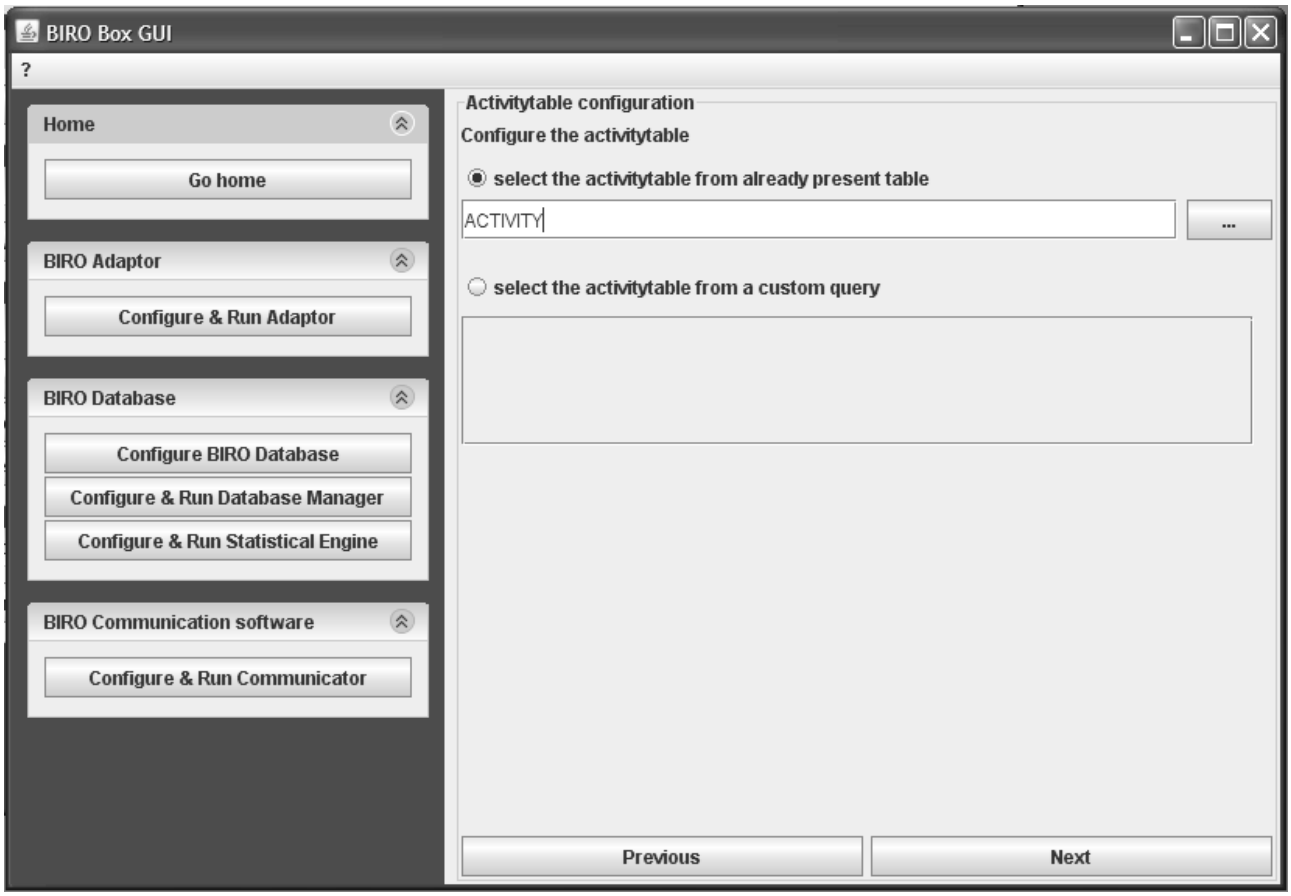


Figure A.4: BIROBox- BIROAdaptor activityTable configuration

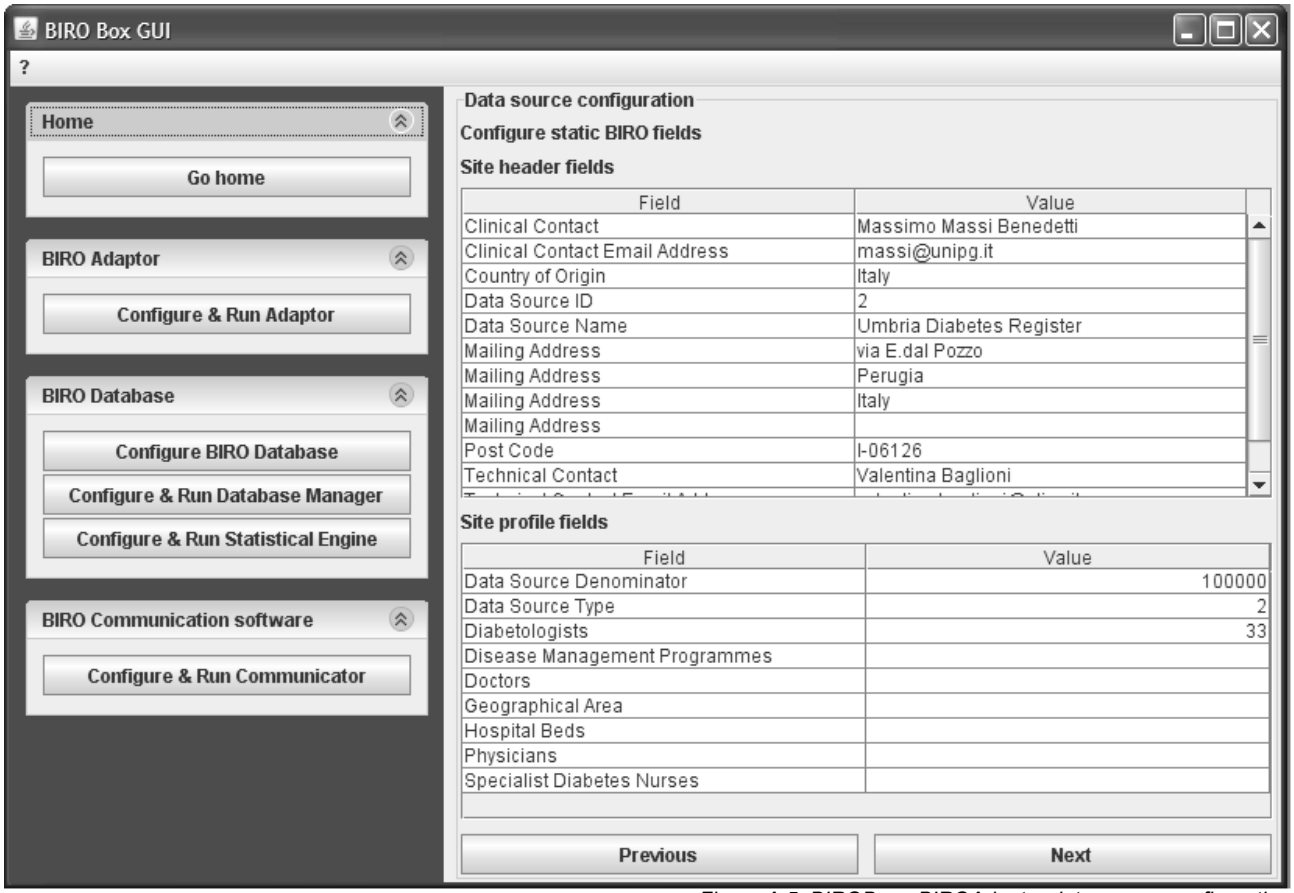


Figure A.5: BIROBox - BIROAdaptor data source configuration

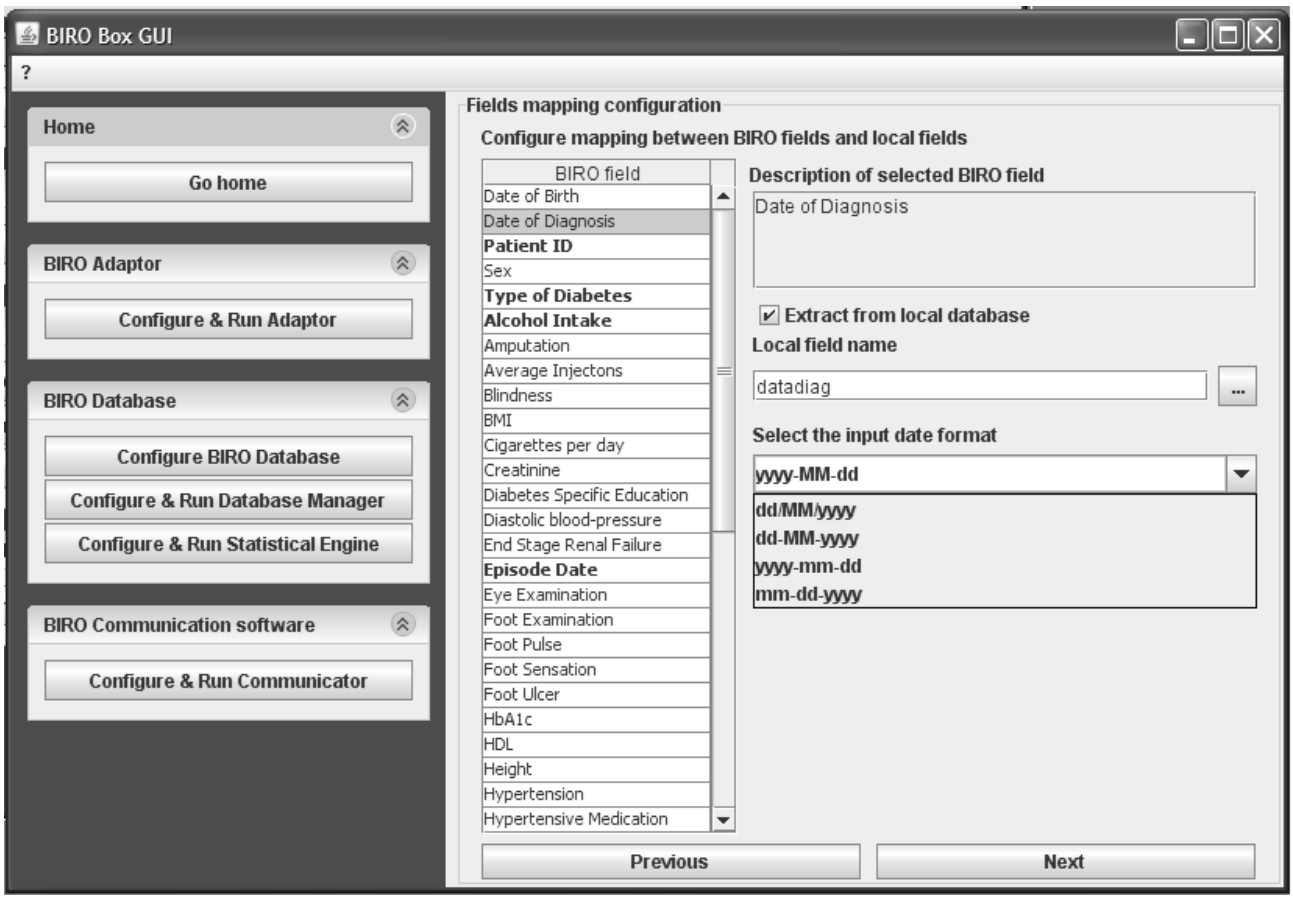


Figure A.6: BIROBox -BIROAdaptor fields mapping configuration

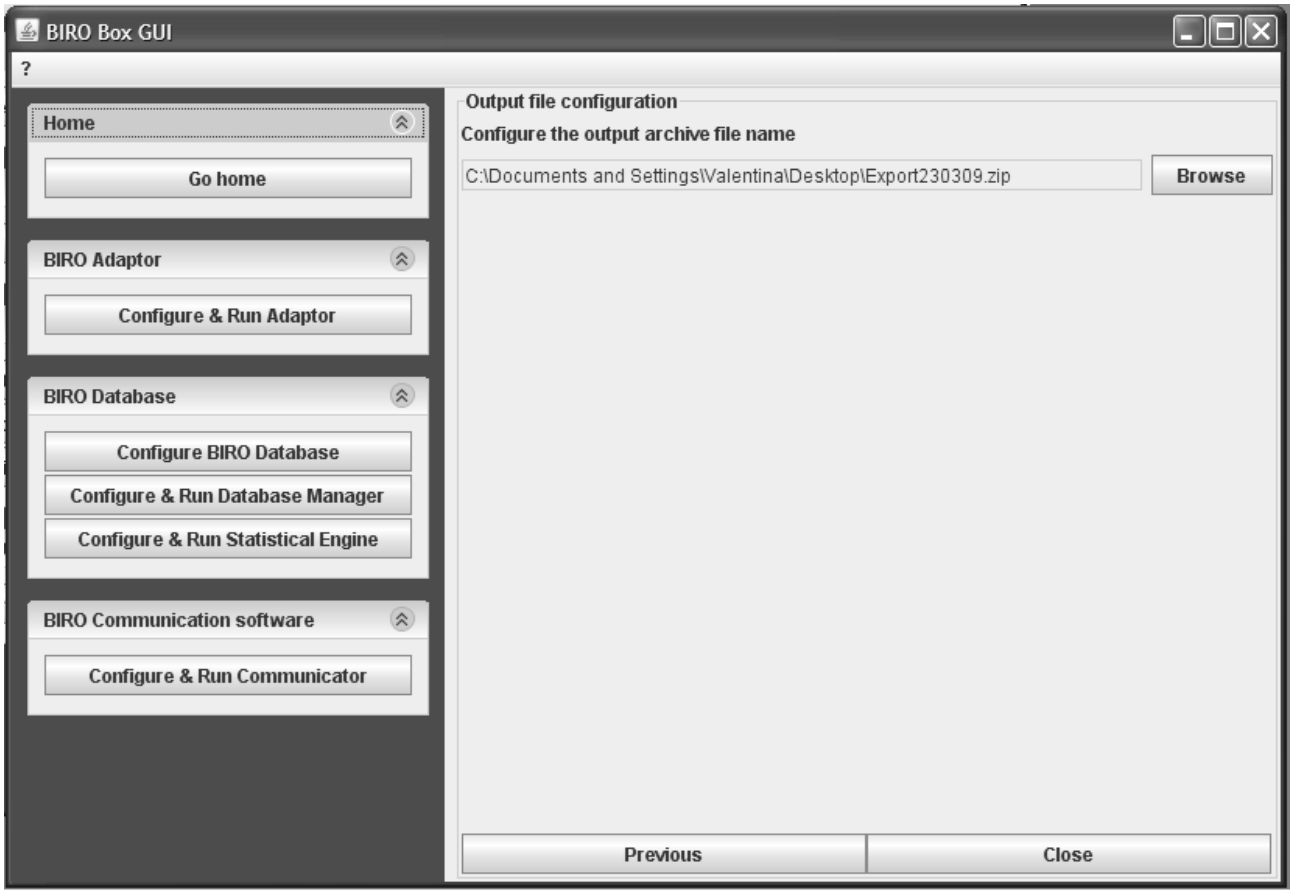


Figure A.7: BIROBox - BIROAdaptor output file configuration

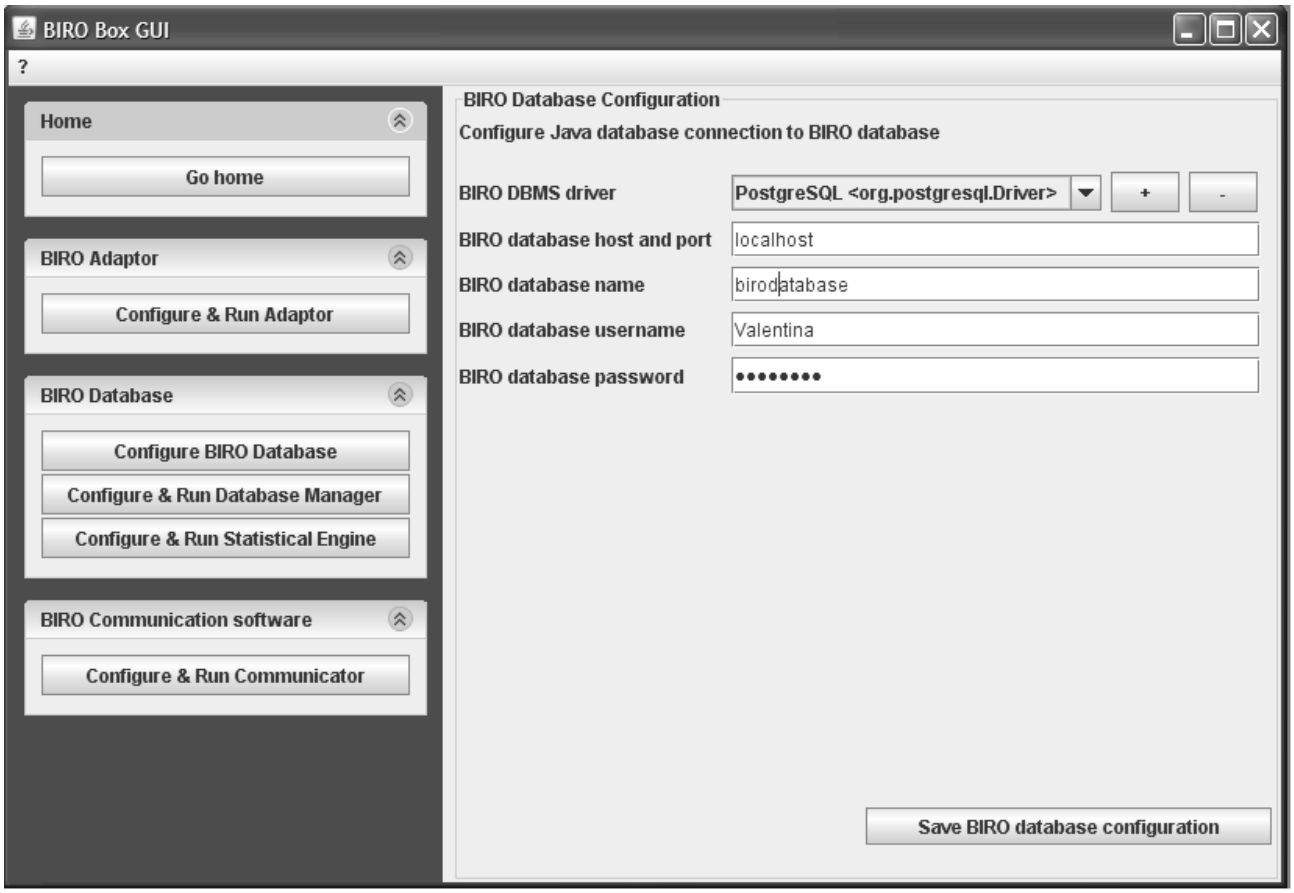


Figure A.8: BIROBox- BIRO database configuration

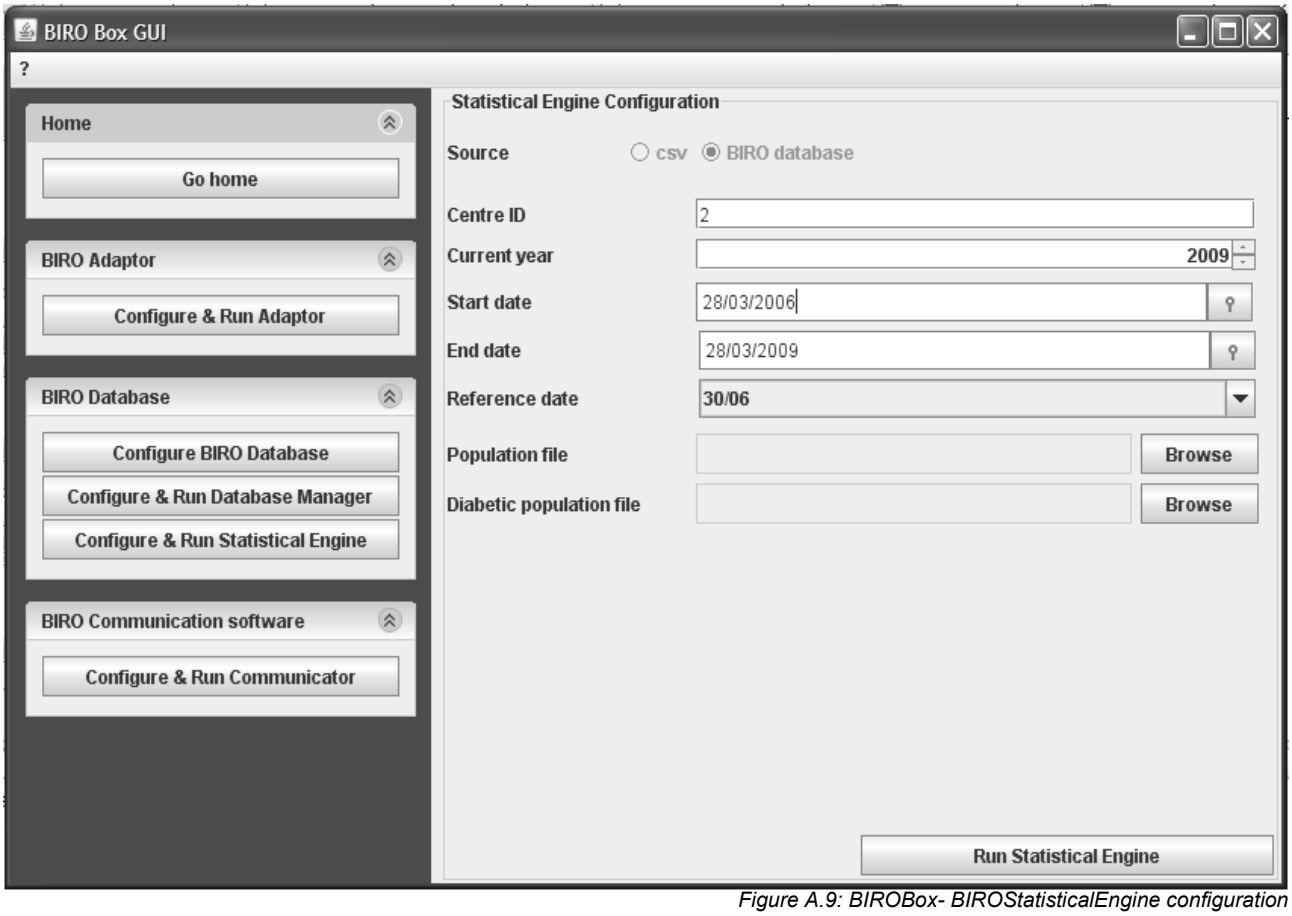


Figure A.9: BIROBox- BIROStatisticalEngine configuration

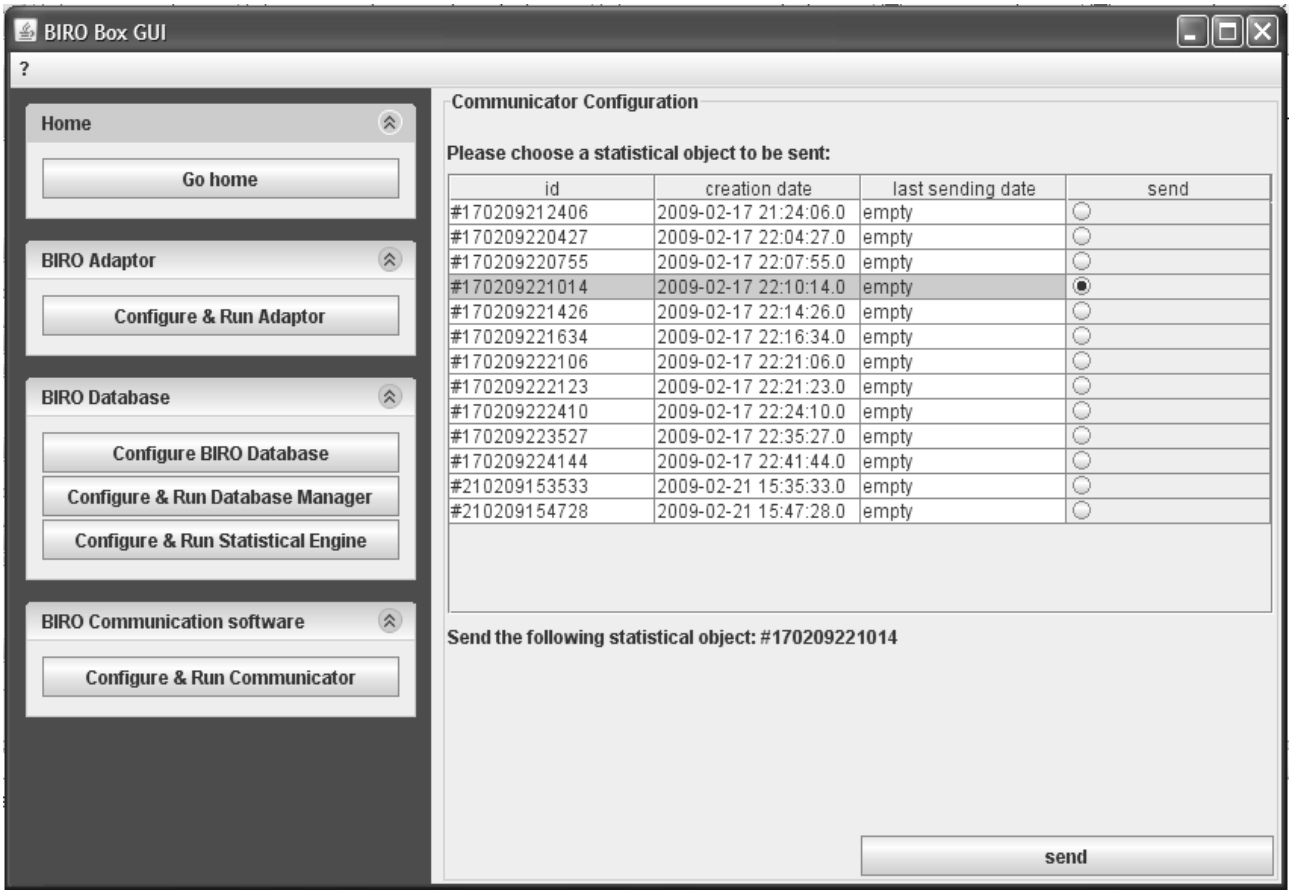


Figure A.10: BIROBox- BIROCommunicationSoftware configuration

References

1. Java Runtime Enviroment, available at: http://sun.com/
2. R software for statistical computing and graphics, available at: http://r-project.org/
3. LATEX typesetting system, available at: http://latex-project.org/
4. Miktek typesetting software, available at: http://miktex.org/
5. deliverable D8.1 Statistical Engine, available at: http://biro-project.eu/results

www.biro-project.eu

The Biro Consortium

JOANNEUM RESEARCH
Austria

Ministry of Health
Cyprus

Paulescu Institute
Romania

University of Bergen
Norway

UNIVERSITY OF DUNDEE
Scotland, UK

University of Malta
Malta

University of Perugia
Italy

Major Subcontractors

Serectrix
Supporting Decisions
Italy

Telemedica Consulting
Romania

Major Sponsoring Institutions

Directorate-General for Health & Consumers

Regione Umbria

NOKLUS

Supported by

FONDAZIONE CASSA RISPARMIO PERUGIA

Printed in Perugia, Italy, 2009