Standardised Diabetes Data Definitions

Scott Cunningham

University of Dundee

Introduction

- Background
- Methodology
- Core Dataset
- Data Sources
- Data Dictionaries
- XML Schema
- EUBIROD Development
- Summary

Background

- Standardised data essential
- Agreed data definitions allow consistent analyses
- Data items based on Core Indicators from Clinical Review
- Data Dictionary explains ‘data about the data’

Methodology

- Dataset comparison
  - DiabCare
  - Scottish Diabetes Core Dataset
  - Umbria AMD Data File
  - EUCID
- Candidate Data Items Identified
- Scottish and FQSD Definitions compared
- Define BIRO shared dataset and definitions
- Assign ‘Validity’ weighting
  - High – Consistent across all datasets
  - Medium – Inconsistency in 1 dataset
  - Low – Inconsistencies across >1 datasets

Methodology

- International System of Units
- Mapping criteria defined
- Consistent data for analysis
- Partners describe ‘metadata’ about local data collection and standards

Core Dataset

- Dataset items recorded as a “Parameter”
- Parameters have a unique reference
- Clear definition
- Associated data type
- Unit of measurement (e.g.kg/m^2)
- May have an upper or lower range

Core Dataset

- Basic Patient Information
  - e.g. Type of Diabetes, Date of Birth, Year of Diagnosis
- Risk Factors
  - e.g. Cigarettes / Day
- Clinical Measurements
  - e.g. Weight, Height, SBP, DBP, HbA1c, Creatinine
- Examinations
  - e.g. Eye Examinations
- Outcomes
  - e.g. End Stage Renal Failure

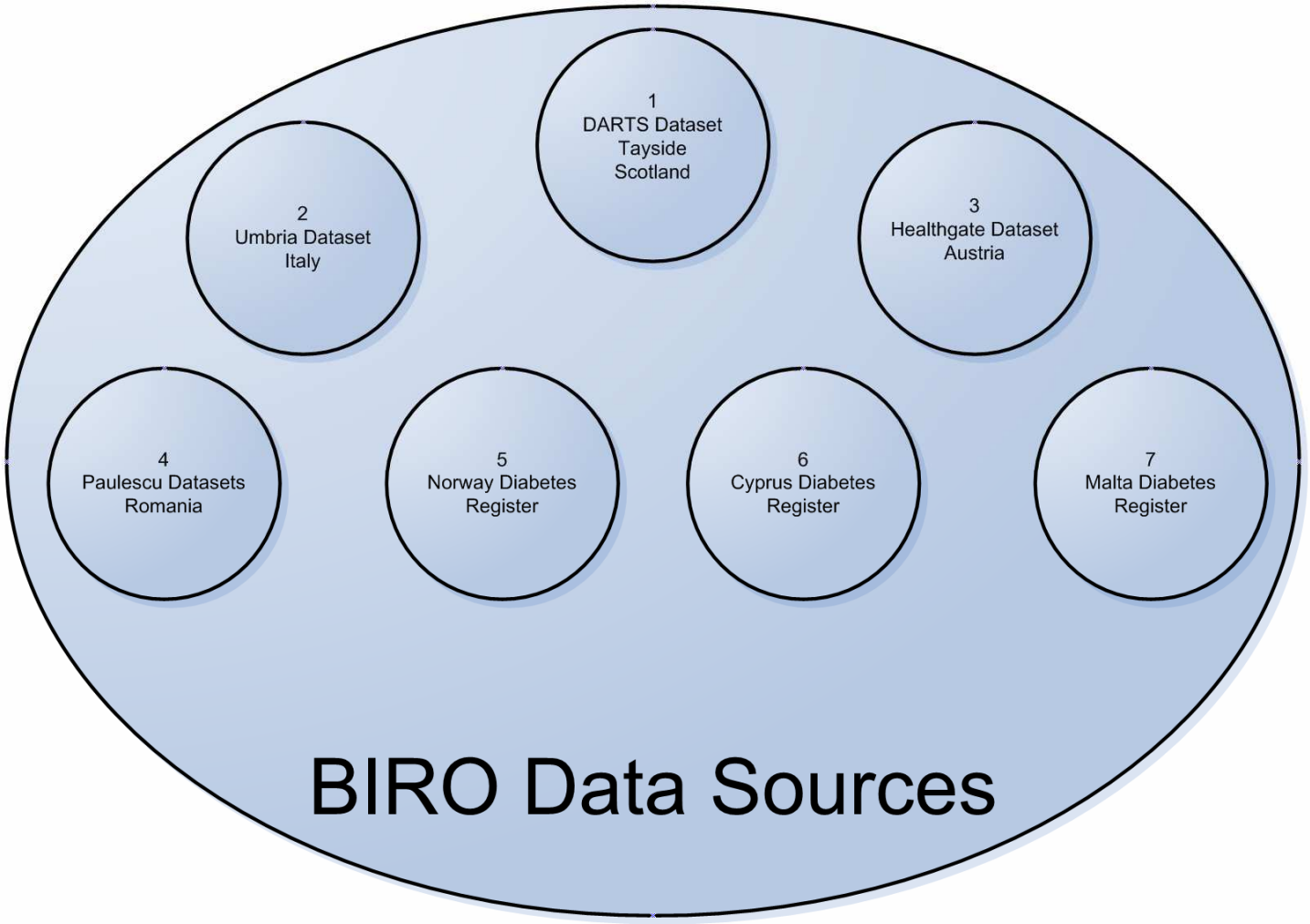
Core Dataset

Reference	Field Name	Parameter	Data Type	Enumerated Codes
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 Types of Diabetes
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	
BRIO047	ALC_STAT	Alcohol Status	Enumerated	1 = Current Drinker 2 = Non-Drinker 3 = Ex-Drinker
BIRO010	ALCOHOL	Alcohol Intake	Integer	
BIRO011	WEIGHT	Weight	Real	
BIRO012	HEIGHT	Height	Real	
BIRO013	BMI	Body Mass Index	Real	
BIRO014	SBP	Systolic Blood Pressure	Integer	
BIRO015	DBP	Diastolic Blood Pressure	Integer	
BIRO016	HBA1C	HbA1c	Real	
BIRO017	CREAT	Creatinine	Integer	
BIRO018	MA_TEST	Microalbumin	Enumerated	1 = MA Test Normal 2 = MA Test Abnormal 0 = No MA Test Recorded
BIRO019	CHOL	Total Cholesterol	Integer	

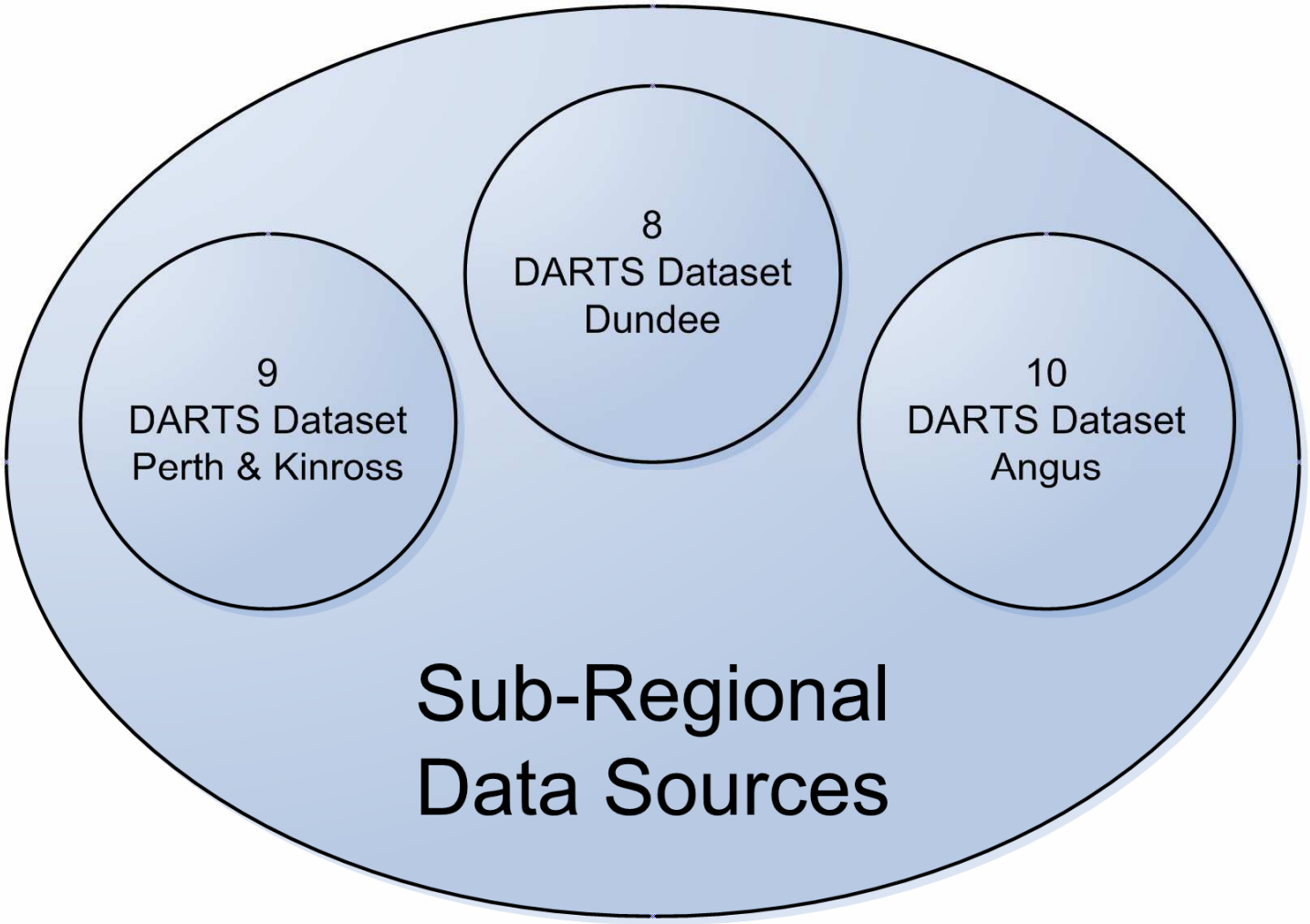
Core Dataset

- Need to describe clinical sites
  - Country
  - Data Source Type
    - e.g. GP, Hospital Clinic (Diabetes), Regional Register
  - Name of Clinic
  - Clinic Denominator (Population)
  - Geographical Area
  - Clinical and technical contacts

Data Sources



Data Sources



Data Dictionaries

- Data Standardisation
- Metadata
  - Consistency
  - Completeness
  - Quality
  - Additional comments
- Can be displayed alongside outputs
  - Explain discrepancies
  - Provide commentary on data comparisons

XML Schema

- Technical infrastructure to support collection of data
  - Standardised
  - Clearly defined
  - Consistent across all sites
  - Supports technical documentation
  - Contains all ‘metadata’

EUBIROD Development

- Reanalyse current datasets
  - Existing ‘BIRO’ datasets
  - Datasets for new partners
- Requirements
  - Data schema
    - Items captured
    - Units of measurement
  - Local Definitions
- Cross-reference and update documentation

Summary

- Data dictionaries and data standards can improve
  - Quality
  - Relevance
  - Consistency
  - Comparability
- A European minimum common dataset for Diabetes has been created based on existing systems
- A data dictionary has been developed to capture local knowledge and ‘metadata’
- XML schemas define the currency for electronic data capture