B.I.R.O.

Best Information through Regional Outcomes

A TAXONOMY OF STATISTICAL OBJECTS

FOR DISTRIBUTED DATA ANALYSIS:

THE STATISTICAL AND CENTRAL ENGINE OF THE BIRO SYSTEM

Fabrizio Carinci
Serectrix s.n.c.

Rome Meeting
20 April 2008

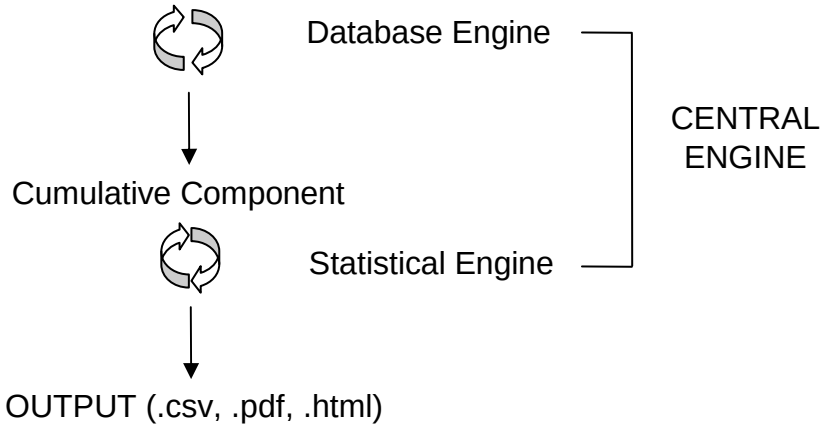
Software procedural flow

Local DB
Step 1. Local (client) data processing and statistical analysis
BIRO Adaptor
XML Export
Local Database Engine
<#rec> … individual data => Statistical Engine => Local Component (.csv)

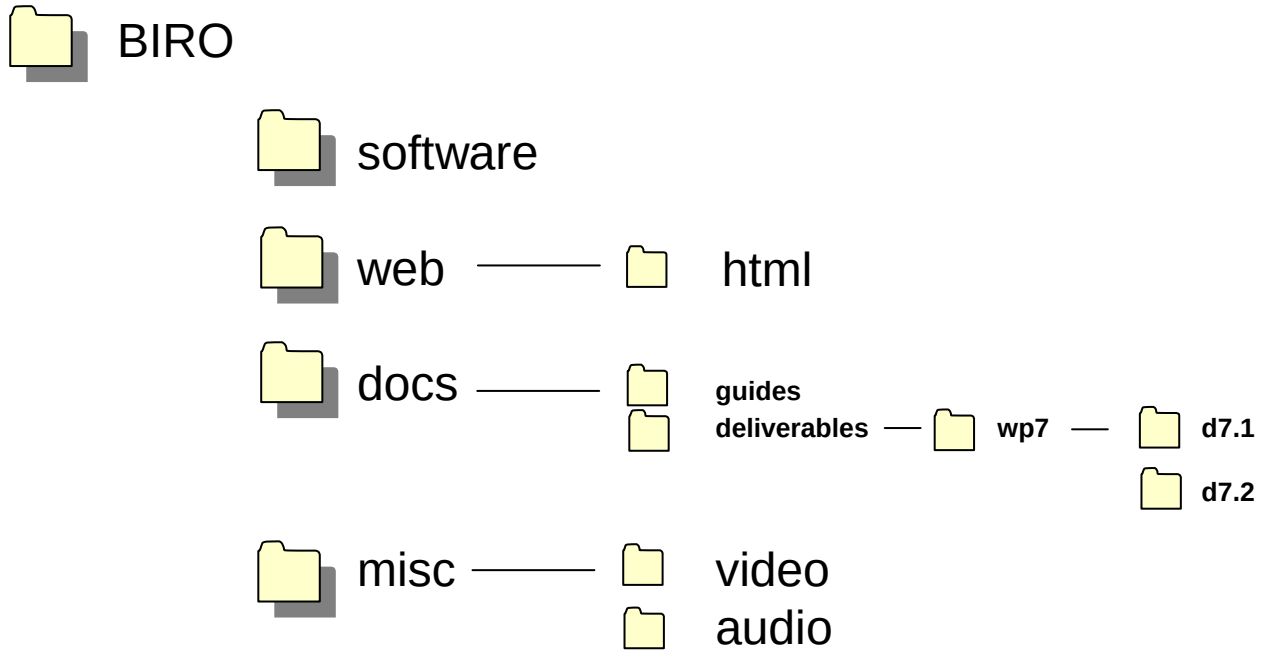
Step 2. Data transmission
Communication Software

Step 3. Global (server) statistical analysis

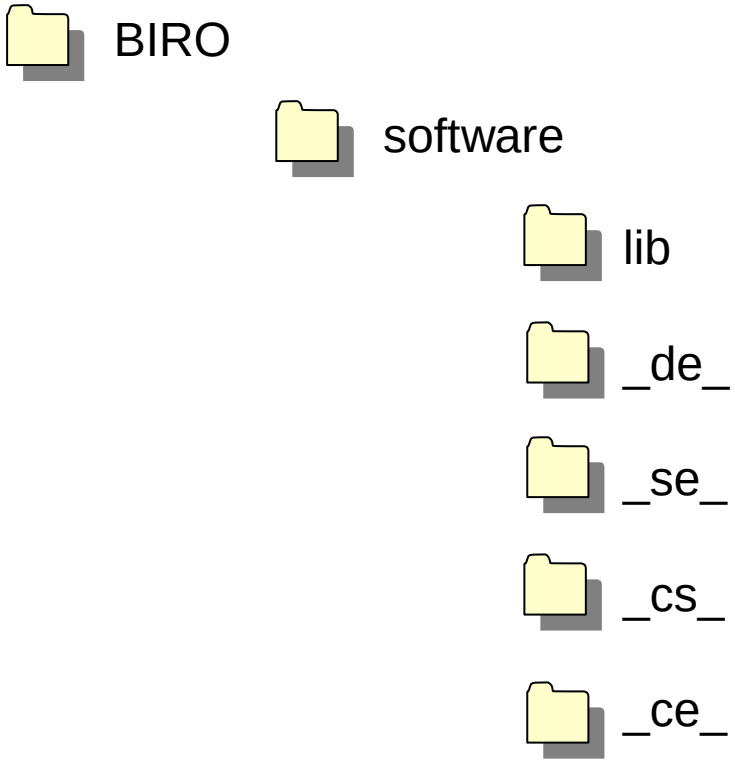
BIRO Server



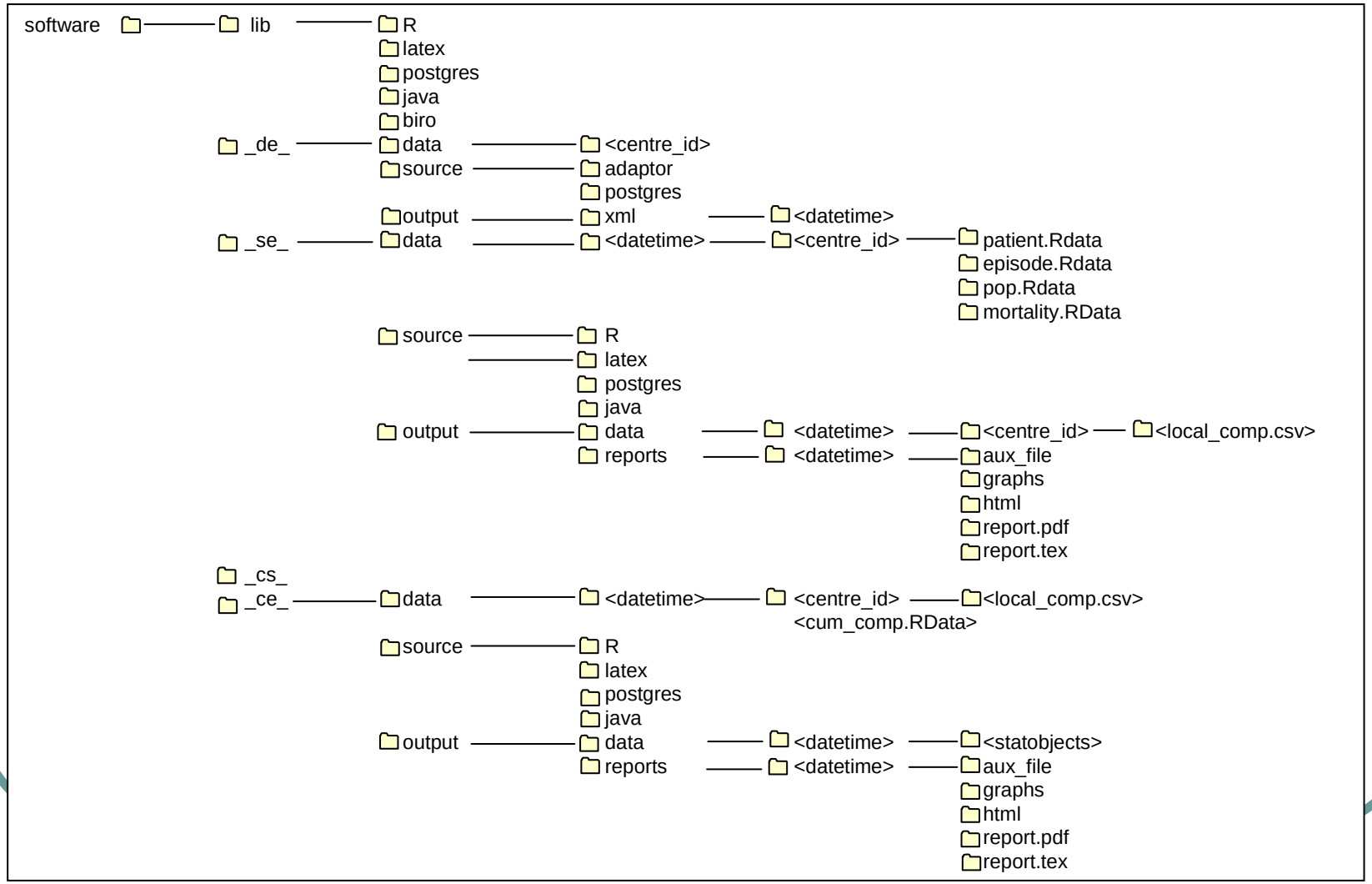
Biro System Directory Structure: root



Biro System Directory Structure: software (1)



Biro System Directory Structure: software (2)



Statistical Object

- 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

(working definition)

Statistical Objects: Meta-data

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)

Statistical Objects Dictionary

SECTION 1. FREQUENCY TABLES

Univariate Frequency Distribution, Outliers, Contingency Table

SECTION 2. MEASURES OF LOCATION

Arithmetic Mean, Percentile, Range

SECTION 3. MEASURES OF DISPERSION

Variance, Interquartile Distance

SECTION 4. GRAPHICAL ELEMENTS

Bar plot, Histogram, Partial boxplot, Overall boxplot, Line plots, XY Plots, Webplot Maps,Forest plot

SECTION 5. REGRESSION

Linear regression,Logistic regression,Meta-analysis

SECTION 6. STANDARDIZATION

Standardized rate, O-E

Arithmetic Mean

Code	2.1
Statistical Object	Arithmetic Mean
Description	Weighted average of a single characteristic, with weights equal to the number of observations for each specific value of the target variable
Variables	CONTINUOUS
Properties	The mean of the overall sample is equal to the weighted mean of the arithmetic means from all local repositories
Local Component	Data vector composed of two quantities: sum of the values of the target variable; total number of observations DATA: **<2.1.a>id, date, stratum, sum_x, n**
Cumulative Component	Sum of the sum of values from each local object DATA: **<2.1.a> id, date, stratum, sum_x, n**
Output	Single value of the overall arithmetic mean: cumulative object, divided by the sum of the total number of observations from each local object DATA: **<2.1.a>mean** Single value of the arithmetic mean by centre: cumulative object, divided by the sum of the total number of observations from each local object, for each centre, for each stratum DATA: **<2.1.b>id, date, stratum, mean**