



B.I.R.O.

Best Information through Regional Outcomes

A Public Health Project funded by the European Commission, DG-SANCO 2005

WORK PACKAGE 4

D4.3: DICTIONARY / XML UPDATE

DOCUMENT v1.0

July 2009



Clinical Technology Centre
University of Dundee
Ninewells Hospital and Medical School
Dundee
DD1 9SY
Scotland

Contents

1. SUMMARY	3
2. REFERENCES	4
3. DOCUMENT CHANGE HISTORY	5
4. METHODOLOGY	6
5. CLINICAL REVIEW DATA ITEM UPDATE	7
DATASET ADDITIONS	7
DATASET MODIFICATIONS	9
CONVERSION FACTORS	11
6. REVIEW OF DATA DICTIONARY	14
PATIENT ACTIVITY STATUS	14
GEOGRAPHICAL CLASSIFICATION	17
DATA MAPPINGS	18
7. REVIEW OF XML SCHEMA	19
ACTIVITY STATUS	19
GEOGRAPHICAL CLASSIFICATION	20
XML SCHEMA AMENDMENTS	21
8. PRESENTATION AND OUTPUT	22
9. DATA SOURCE DEFINITIONS	23
10. ONLINE QUESTIONNAIRE	25
LOGIN AND DATA SOURCE SELECTION	25
SITE HEADER	26
SITE PROFILE	26
FIELD EXPORT PROFILES	27
SUMMARY OF DATA ENTRY	28
11. DYNAMIC DOCUMENTATION	29
12. DEFERRED UPDATES	30
DATASET UPDATES	30
DATA SCHEMA UPDATES	30
UML CLASS DIAGRAMS	31
APPENDIX 1: REVISED DATASET – CLINICAL PARAMETERS	33
APPENDIX 2: REVISED DATASET – CLINICAL SITE PARAMETERS	37
APPENDIX 3: REVISED DATASET – GEOGRAPHICAL CLASSIFICATIONS	39

1. Summary

The purpose of BIRO Work Package 4 is to create an electronic directory inclusive of concept and data dictionaries for diabetes care and prevention, thereby allowing dynamically links to the clinical knowledge-base and to the systematic evaluation of health systems outcomes. The electronic dictionary and XML schema tie in closely with the indicator development in work package 2, the minimum common dataset defined in Work Package 3, and the data dictionary documentation as defined in deliverable 4.1.

This deliverable constitutes an update to Work Packages 3 and 4 in response to evaluator feedback and the revised Clinical Review deliverable.

2. References

BIRO Project Website: www.biro-project.eu

BIRO WP2 Clinical Review Indicator Development Results

BIRO WP3 Common Dataset

BIRO WP4 D4.1 Data Dictionary

BIRO WP4 D4.2 XML Metadata Dictionary

BIRO WP7 Reports Template

Chemistry Conversion: <http://www.vin.com/scripts/labquest/converthtml.pl>

Conventional Units to SI Units: http://www.globalrph.com/conv_si.htm

ESRI Shapefiles: <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>
http://en.wikipedia.org/wiki/Shape_files

EUBIROD Project Website: www.eubirod.eu

Nomenclature of Territorial Units for Statistics: <http://en.wikipedia.org/wiki/NUTS>

3. Document Change History

Version	Date	Author	Reason for Update
0.1	April 2009	Scott Cunningham	Initial Template
0.2	April 2009	Scott Cunningham	Update based on partner feedback + additional diagrams
1.0	July 2009	Scott Cunningham	Final update for submission to European Commission

4. Methodology

BIRO Work Package 2 has created a list of Core Indicator Candidates based on the published scientific literature. Within this the scientific justification for each of the corresponding data items is included. BIRO Work Package 3: Common Dataset identifies the data items required to fulfil the requirements of the core indicator list. This has been defined on the basis of the compatibility of each parameter across each dataset. Deliverable 4.1 defined how BIRO data will be held, mapped and represented within the Data Dictionary. Deliverable 4.2 defined the data that must be captured about each data item to provide an in-depth profile for retrospective data analyses. This data can then be displayed alongside clinical outputs on the BIRO website as defined in Work Package 7, the Reporting Template.

The objectives of BIRO Work Package 4, Deliverable 4.3 are to provide an update on the documentation created earlier in this Work Package and from Work Package 3, based on partner feedback, evaluator feedback and from subsequent Work Packages. These updates include additional data items based on the revised Clinical Review and through an expansion on the data capture requirements through the development of the Statistical Engine.

The update process consisted of the following

- Review of Data Item Requirements
- Review of Data Dictionary
- Review of XML Schema
- Presentation and Output

In addition, a new area is covered in this documentation detailing the approach for 'Metadata Capture'. This includes the following:

- Data Source Definitions
- Online Questionnaire
- Dynamic documentation (creation of XML files)

5. Clinical Review Data Item Update

The revised Work Package 2 documentation published in January 2009 provided an update to the core indicator list to be supported. This resulted in a small number of dataset modifications as detailed below:

Dataset Additions

Parameter:	Lipid-Lowering Therapy	
BIRO Ref:	BIRO053	
Field Name:	LIPID_THERAPY	
Data Type:	Enumerated(0, 1)	
Definition:	1 = Yes	Date of record of treatment using lipid lowering drugs is valid
	0 = No	Date of record of treatment using lipid lowering drugs is NULL or contains invalid date
Mandatory:	No	
Validity:	Medium	
Data Mapping:	Lipid-lowering therapy not recorded on DiabCare Basic Information sheet. For NHS Scotland data, lipid-lowering medication will be extracted using prescribed drug British National Formulary (BNF) Code	

Parameter:	Anti-Platelet Therapy	
BIRO Ref:	BIRO054	
Field Name:	ANTIPLATELET_THERAPY	
Data Type:	Enumerated(0, 1)	
Definition:	1 = Yes	Date of record of treatment using anti-platelet drugs is valid
	0 = No	Date of record of treatment using anti-platelet drugs is NULL or contains invalid date
Mandatory:	No	
Validity:	Low	
Data Mapping:	Anti-platelet therapy not recorded on DiabCare Basic Information sheet or Umbria Dataset For NHS Scotland data, anti-platelet medication will be extracted using prescribed drug British National Formulary (BNF) Code	

Parameter:	Patient Enrolment in Disease Management Programme (DMP) for Diabetes	
BIRO Ref:	BIRO048	
Field Name:	ENROL_DMP	
Data Type:	Enumerated(0, 1)	
Definition:	1 = Yes	Date of record of patient enrolment in structured Diabetes Disease Management Programme
	0 = No	Date of record of patient enrolment in structured Diabetes Disease Management Programme is NULL or contains invalid date
Mandatory:	No	
Validity:	Low	
Data Mapping:	Enrolment in Disease Management Programme not recorded on DiabCare Basic Information sheet or Umbria Dataset	

Parameter:	Physicians Offering Disease Management Programmes (DMP's) for Diabetes	
BIRO Ref:	BIRO122	
Field Name:	DS_DMP_PHYSICIANS	
Data Type:	Integer	
Definition:	The number of Physicians offering and recruiting for structured Diabetes Disease Management Programmes	
Mandatory:	Yes	

Dataset Modifications

The following data items have been amended to become mandatory data items as they are required for a large proportion of the statistical modelling provided by the local BIRO Engine. For 'Sex' an additional enumerated value has been included to capture patients whose gender has not been recorded.

Parameter:	Sex	
BIRO Ref:	BIRO004	
Field Name:	SEX	
Data Type:	Enumerated (1, 2)	
Definition:	1 = Male	Male Phenotype at birth
	2 = Female	Female Phenotype at birth
	0 = Not Recorded	Gender not recorded
Mandatory:	Yes	
Validity:	High	

Parameter:	Date of Birth	
BIRO Ref:	BIRO005	
Field Name:	DOB	
Data Type:	Date/Time	
Definition:	Date of birth of subject (ccyy-01-01)	
Upper Range:	Current Date	
Mandatory:	Yes	
Validity:	High	
Data Mapping:	Only year of birth is necessary in order to maintain anonymity. As such all dates of birth should be stored as 'ccyy-01-01' so that only the year of birth is known. This will also allow the full data of birth to be stored without changing the dataset in future.	

The recording of 'Geographical Area' has been amended to incorporate a more appropriate Unit of measurement. This has been changed from m² to km².

Parameter:	Geographical Area
BIRO Ref:	BIRO105
Field Name:	DS_AREA
Data Type:	Integer
Definition:	Area of coverage for data source
Units:	km ²
Mandatory:	Yes

The following data item has been removed due to a change in the methods of analysis of Disease Management Programmes. The initial indicators analysed only the number of programmes run within the geographical location, while the updated criteria require information on Patient Enrolment and Physicians involved in recruitment as detailed above.

Parameter:	Disease Management Programmes
BIRO Ref:	BIRO121
Field Name:	DS_PROGS
Data Type:	Integer
Definition:	Number of disease management programmes in data source geographical area. Availability of a DMP influences the level of structured and evidence based treatment.
Units:	Disease Management Programmes
Mandatory:	Yes

Appendix A shows the updated short list of BIRO Clinical data items, with Appendix B showing the updated Data Source data items. New data items are marked in green with items removed shown in red.

Conversion Factors

Following the publication of the BIRO dataset (WP3), further validation has been performed on the conversion factors to map data from conventional to standardised units of measurement. As a result, the following corrections are necessary based on the reference materials detailed earlier in this document.

Parameter:	Creatinine
BIRO Ref:	BIRO017
Field Name:	CREAT
Data Type:	Integer
Definition:	Value in $\mu\text{mol/l}$
Units:	$\mu\text{mol/l}$
Mandatory:	No
Validity:	High
	SDCD defines creatinine in $\mu\text{mol/l}$. FQSD defines creatinine in mg/dl – Values can be mapped.
Data Mapping:	To convert mg/dl to $\mu\text{mol/l}$, multiply by 88.4

Parameter:	Total Cholesterol
BIRO Ref:	BIRO019
Field Name:	CHOL
Data Type:	Real(nn.nn)
Definition:	Serum total cholesterol can be either fasted or unfasted
Units:	mmol/L
Mandatory:	No
Validity:	High
	FQSD also allows total cholesterol in mg/dl.
Data Mapping:	To convert mg/dl to mmol/L, multiply by 0.0259

Parameter:	HDL
BIRO Ref:	BIRO020
Field Name:	HDL
Data Type:	Real(nn.nn)
Definition:	Serum HDL cholesterol can be either fasted or unfasted
Units:	mmol/L
Mandatory:	No
Validity:	High
Data Mapping:	To convert mg/dl to mmol/L, multiply by 0.0259

Parameter:	LDL
BIRO Ref:	BIRO046
Field Name:	LDL
Data Type:	Real(nn.nn)
Definition:	Serum LDL cholesterol can be either fasted or unfasted
Units:	mmol/L
Mandatory:	No
Validity:	High
Data Mapping:	To convert mg/dl to mmol/L, multiply by 0.0259

Parameter:	Triglycerides
BIRO Ref:	BIRO021
Field Name:	TG
Data Type:	Real(nn.nn)
Definition:	Value in mmol/L
Units:	mmol/L
Mandatory:	No
Validity:	High
Data Mapping:	To convert mg/dl to mmol/L, multiply by 0.0259

6. Review of Data Dictionary

In addition to the changes listed above, further updates have been implemented in the Data Dictionary to allow the capture of further key elements for various analyses. The new requirements are discussed in detail below.

Patient Activity Status

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.

It was agreed at the project meeting in January 2009 that it would be useful to be able to track a patients' periods of 'activity' if the contributing data source is able to supply this information. It would also allow for more sophisticated statistical analyses.

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 Reasons are:

- Death
- Transferred Out
- Lost to Follow-up

It is possible for a patient to have one continuous or several disjointed periods of clinical activity based on their diagnosis dates, location of residence or follow up status. The reference information for these new data items are as follows:

Parameter:	Activity Start Date
BIRO Ref:	BIRO049
Field Name:	AD_START_DATE
Data Type:	Date/Time
Definition:	Date of commencement of current period of patient activity
Mandatory:	No

Parameter:	Activity Start Reason	
BIRO Ref:	BIRO050	
Field Name:	AD_START_REASON	
Data Type:	Enumerated(1, 2, 3)	
Definition:	1 = Birth	Patient born with diabetes on start date
	2 = Diabetes Diagnosis	Patient diagnosed with diabetes on start date
	3 = Transferred In	Patient transferred in with diabetes diagnosis
Mandatory:	No	

Parameter:	Activity End Date
BIRO Ref:	BIRO051
Field Name:	AD_END_DATE
Data Type:	Date/Time
Definition:	Date of completion of current period of patient activity
Mandatory:	No

Parameter:	Activity End Reason	
BIRO Ref:	BIRO052	
Field Name:	AD_END_REASON	
Data Type:	Enumerated(1, 2, 3)	
Definition:	1 = Death	Patient with diabetes died on end date
	2 = Transferred Out	Patient with diabetes transferred out on end date
	3 = Lost to Follow-up	Patient with diabetes lost to follow
Mandatory:	No	

Geographical Classification

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

Two international standards were considered for this purpose. These were “ESRI Shapefiles” and “Nomenclature of Territorial Units for Statistics (NUTS)” classifications. After careful consideration of the likelihood of each BIRO partner to be able to obtain the required information for these schemes, at the present time 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 and will be documented later in this deliverable. The reference information for these new data items are as follows:

Parameter:	Level 0 Classification
BIRO Ref:	BIRO200
Field Name:	GC_CONTINENT
Data Type:	String(10)
Definition:	Continent (BIRO Custom Level)
Mandatory:	Yes

Parameter:	Level 1 Classification
BIRO Ref:	BIRO201
Field Name:	GC_COUNTRY
Data Type:	String(10)
Definition:	Country (NUTS Level 0)
Mandatory:	Yes

Parameter:	Level 2 Classification
BIRO Ref:	BIRO202
Field Name:	GC_MACRO_REGION
Data Type:	String(10)
Definition:	Sub-National Area (NUTS Level 1)
Mandatory:	No

Parameter:	Level 3 Classification
BIRO Ref:	BIRO203
Field Name:	GC_REGION
Data Type:	String(10)
Definition:	Region (NUTS Level 2)
Mandatory:	No

Parameter:	Level 4 Classification
BIRO Ref:	BIRO204
Field Name:	GC_HEALTH_AUTHORITY
Data Type:	String(10)
Definition:	Local Health Authority (BIRO Custom Level)
Mandatory:	No

Parameter:	Level 5 Classification
BIRO Ref:	BIRO205
Field Name:	GC_PROVINCE
Data Type:	String(10)
Definition:	Province (NUTS-3)
Mandatory:	No

Parameter:	Level 6 Classification
BIRO Ref:	BIRO206
Field Name:	GC_DISTRICT_UNIT
Data Type:	String(10)
Definition:	District Health Unit (BIRO Custom Level)
Mandatory:	No

Parameter:	Level 7 Classification
BIRO Ref:	BIRO207
Field Name:	GC_POST_CODE
Data Type:	String(10)
Definition:	Post Code (BIRO Custom Level)
Mandatory:	No

Data Mappings

Feedback from the evaluators report on D4.1 and D4.2 asked for clarity on the responsibility of data mapping from source databases to the agreed BIRO definitions. All partners are responsible for mapping their 'source' data to the agreed BIRO dataset. This includes the conversion of units of measurement, and local data definitions to those defined by BIRO. In order to assist in this process, the BIRO Adaptor provides the facility to configure source and destination units and mapping criteria based on the conversion factors detailed in BIRO Work Packages 3 and 4..

7. Review of XML Schema

As a result of the updates detailed above the two key changes required for the XML Schema are the additional data capture requirements for 'Activity Status' and 'Geographical Classification'. These are documented below.

Activity Status

Patient Activity status is discussed in section 6. The XML below shows how this can be represented based on the ECDataExport.xsd schema.

```
<ActivityData>  
  <AD_START_DATE>2005-01-01</AD_START_DATE>  
  <AD_START_REASON>Diagnosis</AD_START_REASON>  
  <AD_END_DATE>2009-01-01</AD_END_DATE>  
  <AD_END_REASON>Death</AD_END_REASON>  
</ActivityData>
```

This data can be repeated to allow several periods of patient activity.

AD_START_DATE: The date of the period of care

AD_START_REASON: Reason for inclusion in diabetes care systems

AD_END_DATE: The date of the end of this period of care

AD_END_REASON: Reason for exclusion for diabetes care systems

Geographical Classification

A method of recording geographical areas was also discussed in section 6 of this document. This allows for the import of potentially large NUTS data files, but can also be used to identify the location of the individual patients. The parameters captured are as follows:

```
<GeoClassification>
  <GC_CONTINENT>EU</GC_CONTINENT>
  <GC_COUNTRY>IT</GC_COUNTRY>
  <GC_MACRO_REGION>ITE</GC_MACRO_REGION>
  <GC_REGION>ITE2</GC_REGION>
  <GC_HEALTH_AUTHORITY>101</GC_HEALTH_AUTHORITY>
  <GC_PROVINCE>ITE21</GC_PROVINCE>
  <GC_DISTRICT_UNIT>10101</GC_DISTRICT_UNIT>
  <GC_POST_CODE>54011</GC_POST_CODE>
</GeoClassification>
```

GC_CONTINENT: At present, this will always be EU, until the project expands beyond Europe

GC_COUNTRY: The country of the data source being described (NUTS-0)

GC_MACRO_REGION: A NUTS-1 sub-national area

GC_REGION: A NUTS-2 region

GC_HEALTH_AUTHORITY: A BIRO modification to allow the capture of health board information

GC_PROVINCE: A NUTS-3 compliant province

GC_DISTRICT_UNIT: A BIRO modification to allow sub-health-board analyses

GC_POST_CODE: The lowest level of geographical data

For patients, the 'ECDataExport.xsd' schema has been modified to capture this data for the individual. For the purpose of capturing the geographical areas associated with a data source, a new schema, 'ECGeoClassifications.xsd' has been created. This contains the fields listed above alongside the relevant data source ID. This data can then be loaded and interpreted by the local engine.

XML Schema Amendments

Since the completion of BIRO Deliverable 4.2 the following <"SiteProfile"> elements have had their data types converted from 'xsd:positiveInteger' to 'xsd:integer' in order to allow '0' values:

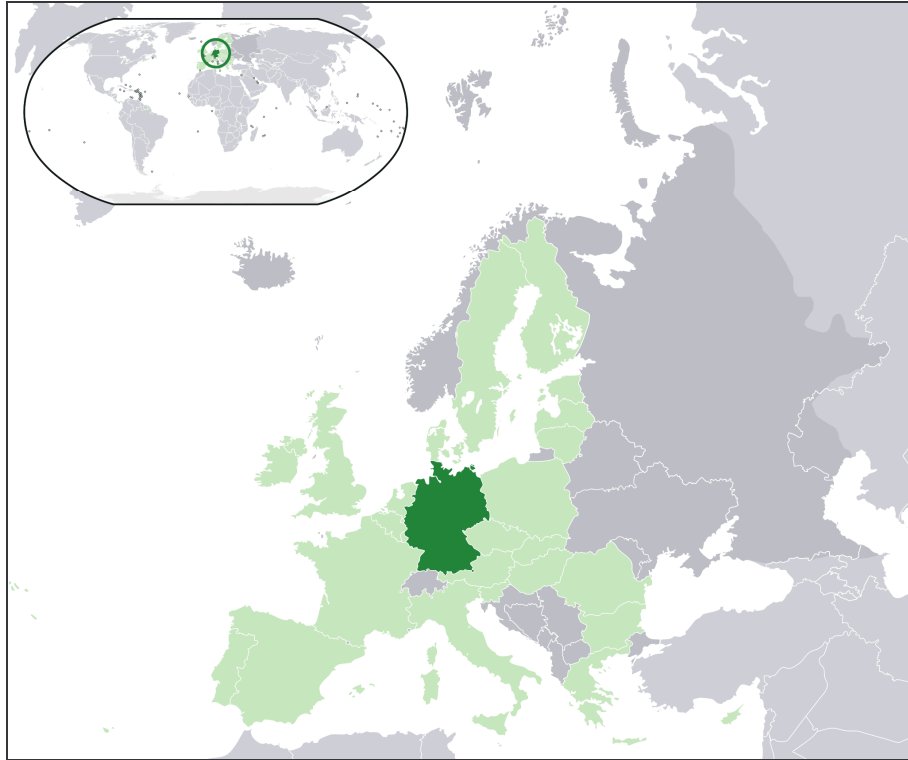
DS_BEDS
DS_PHYSICIANS
DS_DIABETOLOGISTS
DS_DOCTORS
DS_DSN
DS_PROGS

This is appropriate as it is perfectly feasible for non-clinical data sources to maintain responsibility for <1 of each of these parameters.

8. Presentation and Output

The main updates available in relation to the presentation of BIRO data relates to the 'Geocode' information that 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.

The example below shows how information of this type can be used to highlight a country (in this example, Germany) on a map of Europe.



The various levels of granularity can then allow the user to 'drill-down' to the sub-regional level:

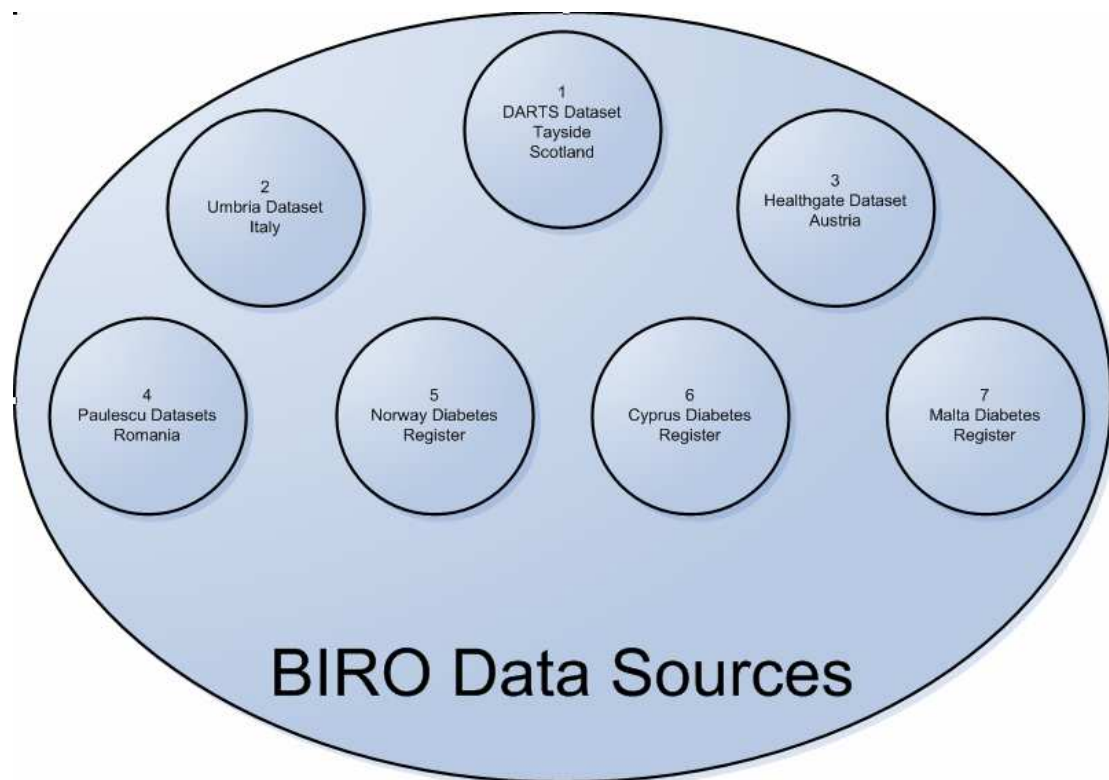


It is believed that this level of data presentation and display will lead to a greater level of user satisfaction and interest when browsing the final BIRO website.

9. Data Source Definitions

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:

1. DARTS Dataset, Tayside, Scotland
2. Umbria Dataset, Italy
3. Healthgate Dataset, Austria
4. Paulescu Datasets, Romania
5. Norway Diabetes Register
6. Cyprus Diabetes Register
7. Malta Diabetes Register

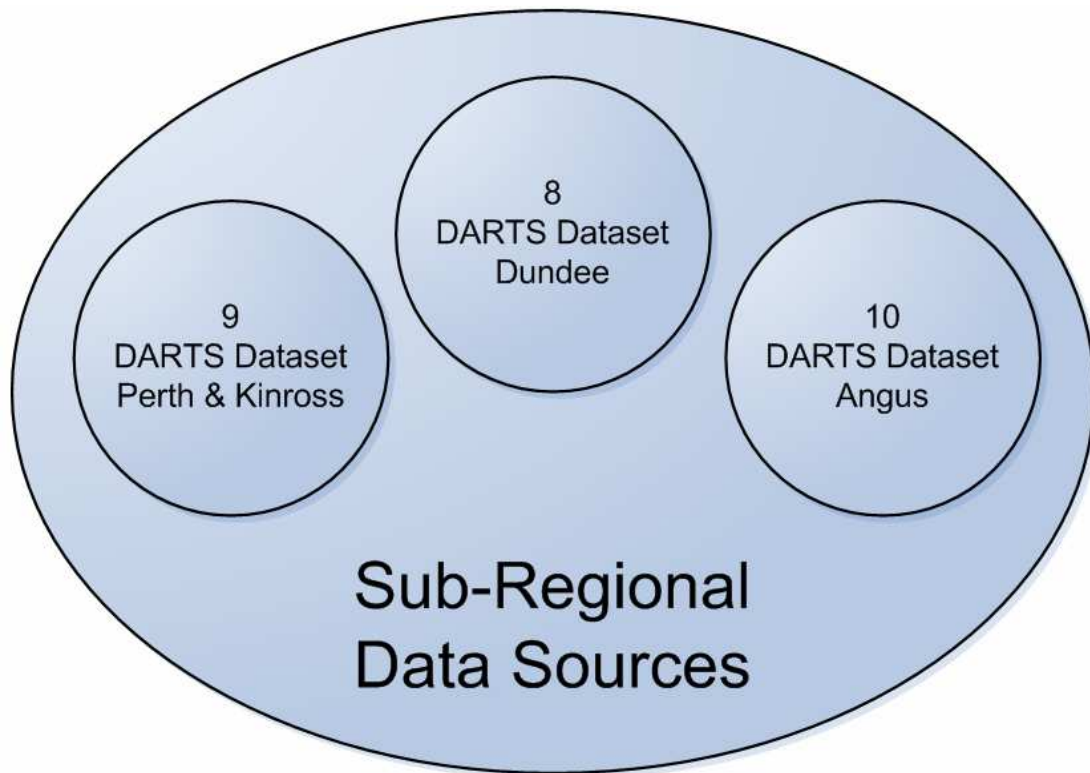


These top-level domains can be further split to provide a more detailed level of granularity. For example, if BIRO wished to analyse data from within the DARTS dataset, contributing data could be split by Tayside sub-regions and marked appropriately:

8. DARTS Dataset – Dundee
9. DARTS Dataset – Perth & Kinross
10. DARTS Dataset – Angus

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

Work Package 4: D4.3 Dictionary / XML Update



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:

- 11. Healthgate Clinical System
- 12. Healthgate Insurance System

Section 10 of this document 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 described 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 summarised and linked back to the contributing region or country.

In Italy, for example, we will have a clear handle on the Umbria datasets, but this may vary from other regions within the country, and 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.

10. Online Questionnaire

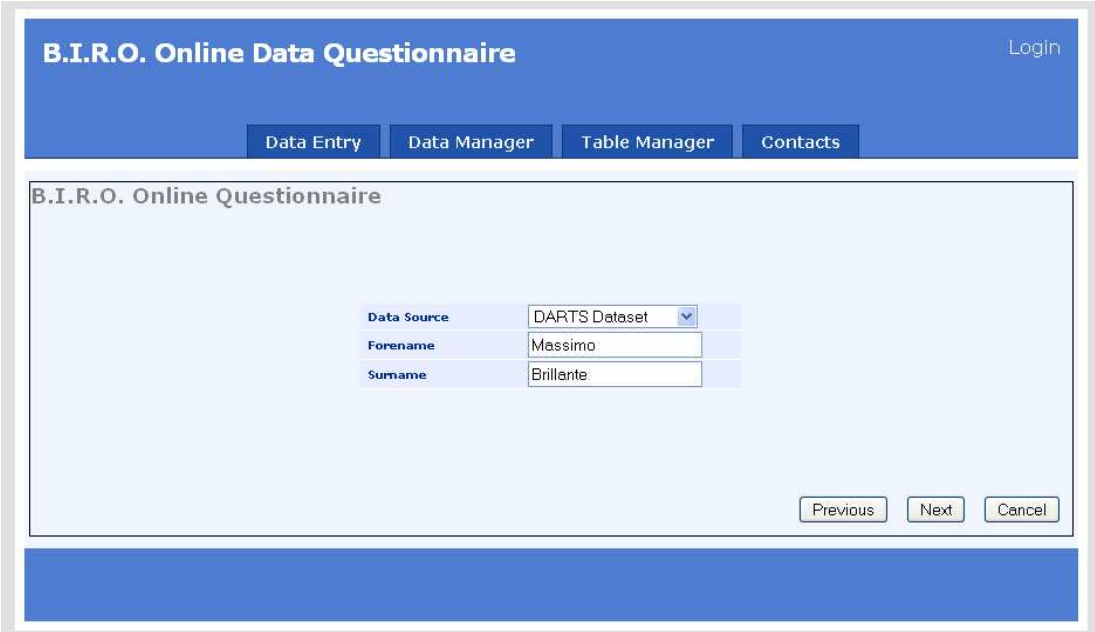
In order to provide a low-level description of confounding factors and to outline discrepancies in data collection, manipulation and data storage between partners, it is necessary to capture metadata. A series of metadata collection principles were outlined in Deliverables 4.1 and 4.2 and in order to 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.

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.



The screenshot displays the 'B.I.R.O. Online Data Questionnaire' interface. At the top, there is a blue header bar with the title 'B.I.R.O. Online Data Questionnaire' on the left and a 'Login' link on the right. Below the header, there are four navigation buttons: 'Data Entry', 'Data Manager', 'Table Manager', and 'Contacts'. The main content area is titled 'B.I.R.O. Online Questionnaire' and contains a form with the following fields:

Data Source	DARTS Dataset
Forename	Massimo
Surname	Brillante

At the bottom right of the form, there are three buttons: 'Previous', 'Next', and 'Cancel'.

Site Header

The second section 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.

The screenshot shows the 'B.I.R.O. Online Data Questionnaire' interface. At the top, there is a blue header with the title 'B.I.R.O. Online Data Questionnaire' and a 'Login' link. Below the header is a navigation bar with four buttons: 'Data Entry', 'Data Manager', 'Table Manager', and 'Contacts'. The main content area is titled 'B.I.R.O. Online Questionnaire' and contains a form with the following fields:

Address 1	<input type="text"/>	
Address 2	<input type="text"/>	
Address 3	<input type="text"/>	
Address 4	<input type="text"/>	
Post Code	<input type="text"/>	
Country	<input type="text"/>	
Clinical representative	<input type="text"/>	Clinical representative E-mail <input type="text"/>
Technical representative	<input type="text"/>	Technical representative E-mail <input type="text"/>
Comments	<input type="text" value="test comments"/>	

At the bottom right of the form, there are three buttons: 'Previous', 'Next', and 'Cancel'.

Site Profile

This section 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.

The screenshot shows the 'B.I.R.O. Online Data Questionnaire' interface, specifically the 'Site Profile' section. The header and navigation bar are identical to the previous screenshot. The main content area is titled 'B.I.R.O. Online Questionnaire' and contains a form with the following fields:

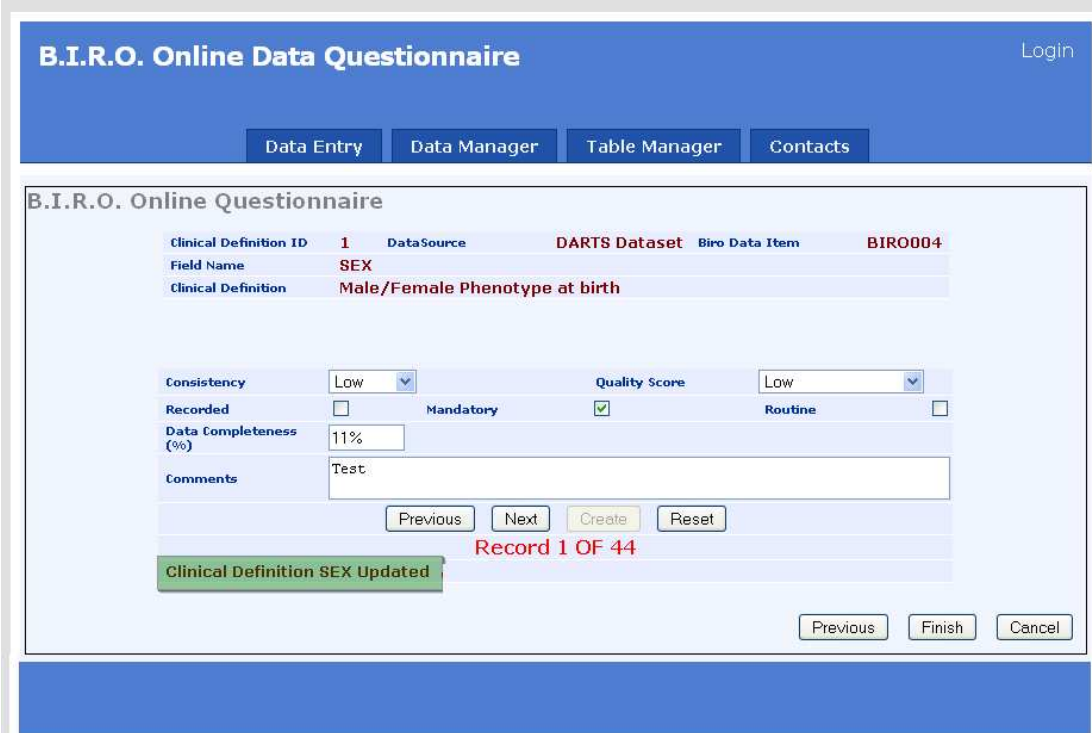
Population	<input type="text" value="10000000"/>	Area (Km2)	<input type="text" value="10000000"/>
Hospital Beds	<input type="text" value="10000000"/>	Physicians	<input type="text" value="100000"/>
Diabetologists	<input type="text" value="10"/>	Diabetes Nurses	<input type="text" value="1000"/>
Disease Management Programmes	<input type="text" value="1"/>	Doctors	<input type="text" value="100000"/>
Data Source Type	<input type="text" value="Disease Management Programme"/>		
Comments	<input type="text" value="Test"/>		

At the bottom right of the form, there are three buttons: 'Previous', 'Next', and 'Cancel'.

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 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.



The screenshot displays the 'B.I.R.O. Online Data Questionnaire' interface. At the top, there is a blue header with the title and a 'Login' link. Below the header is a navigation bar with buttons for 'Data Entry', 'Data Manager', 'Table Manager', and 'Contacts'. The main content area is titled 'B.I.R.O. Online Questionnaire' and contains the following information:

Clinical Definition ID	1	DataSource	DARTS Dataset	Biro Data Item	BIRO004
Field Name	SEX				
Clinical Definition	Male/Female Phenotype at birth				

Below this, there are several input fields and checkboxes:

- Consistency: Low (dropdown)
- Quality Score: Low (dropdown)
- Recorded: Mandatory: Routine:
- Data Completeness (%): 11%
- Comments: Test

At the bottom of the form, there are buttons for 'Previous', 'Next', 'Create', and 'Reset'. A status message 'Record 1 OF 44' is displayed in red. A green notification box at the bottom left says 'Clinical Definition SEX Updated'. At the very bottom, there are buttons for 'Previous', 'Finish', and 'Cancel'.

Summary of Data Entry

On completion of data entry, a summary screen 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 screenshot shows the B.I.R.O. Online Data Questionnaire interface. At the top, there is a blue header with the title "B.I.R.O. Online Data Questionnaire" and a "Login" link. Below the header are four navigation buttons: "Data Entry", "Data Manager", "Table Manager", and "Contacts".

The main content area contains a summary table with the following data:

Area	Beds	DS_DENOM	Diabetologists	Doctors	DS_DSN	Physicians	Programs	DS_TYPE	Date Modified	
Select	10000000	10000000	10000000	10	100000	1000	100000	1	1	08/04/2009

Below this is a table of data entries:

Name	Completeness	Quality Score	Consistency	Mandatory	Recorded	Routine	LastReviewed
BIRO004	11%	Low	Low	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	08/04/2009
BIRO005	11%	Low	Low	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	08/04/2009
BIRO006	11%	Low	Low	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	08/04/2009
BIRO007	11%	Low	Low	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	08/04/2009
BIRO008	11%	Low	Low	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	08/04/2009
BIRO009	11%	Low	Low	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	08/04/2009
BIRO047	11%	Low	Low	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	08/04/2009
BIRO010	11%	Low	Low	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	08/04/2009
BIRO011	11%	Low	Low	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	08/04/2009
BIRO012	11%	Low	Low	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	08/04/2009

At the bottom of the table, there is a pagination bar showing "1" and "2 3 4 5".

A yellow callout box on the right side of the interface contains the following text:

This design has to be refined.. just developed for testing purposes.
 a button to generate XML has to be added here.
 However the concept is this...
 On complete the user is presented with a summary of data just inserted/updated.
 there will be another interface from where the user can select each record and display full info. On these coming screens, the user will be able to update/delete records.

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.

11. Dynamic Documentation

As a result of the online data capture, it is now possible for all descriptive data source XML documentation to be generated dynamically. 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 below shows a sample extract generated by the software for the DARTS, Tayside data source area.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <CDDataSourceExport 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@tht.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_DSH>1</DS_DSH>
    <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>
```

The naming convention of these files is as follows:
Datasource_X_Extract.xml

'X' is the unique data source ID assigned to the clinical domain being described.

12. Deferred Updates

This section details points raised during review, evaluation and discussions within the BIRO project that have been deferred for review during the forthcoming EUBIROD project. These are updates deemed at the present time to be important, but not essential in order for the BIRO project to succeed.

Dataset Updates

The latest Clinical Review updates have amended the 'Core' BIRO Indicators and these continue to be supported as detailed in section 5 of this report. At the present time, the BIRO/EUBIROD dataset does not support data items referenced ONLY in non-Core data Indicators as agreed during project meetings. This approach will be reviewed as work progresses.

Data Schema Updates

As noted during the D4.2 evaluation, the current BIRO schema is overly flexible and requires a considerable amount of understanding by those involved in the data extraction process. A full schema review is planned in order to 'lock-down' and constrain output to those expected.

In addition, it has been suggested that data items should be separated more clearly based on logical groupings. For example:

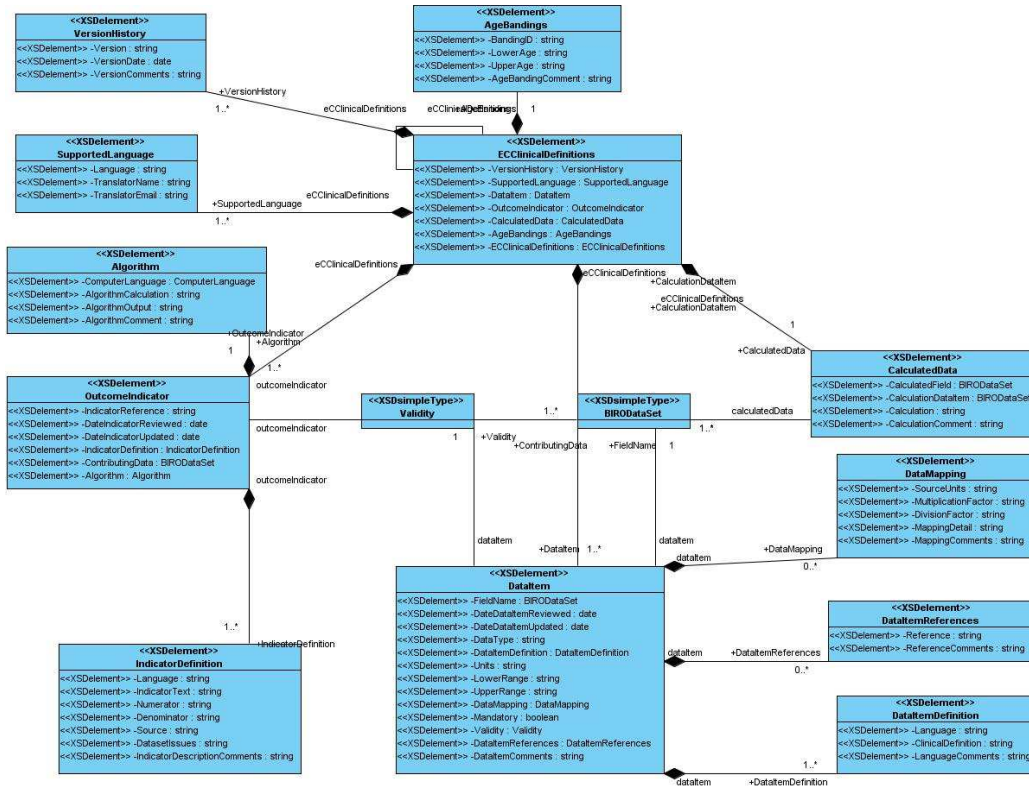
- Patient Demography
- Numeric Measurements
- Observations
- Treatments
- Interventions
- Clinical Events

These groupings would clearly make the logical data separation more obvious to the reader. In addition, although the schema was designed to enable consistent data capture, and not necessarily to influence BIRO Software development, a more structured approach to data design would have the benefits of covering both approaches.

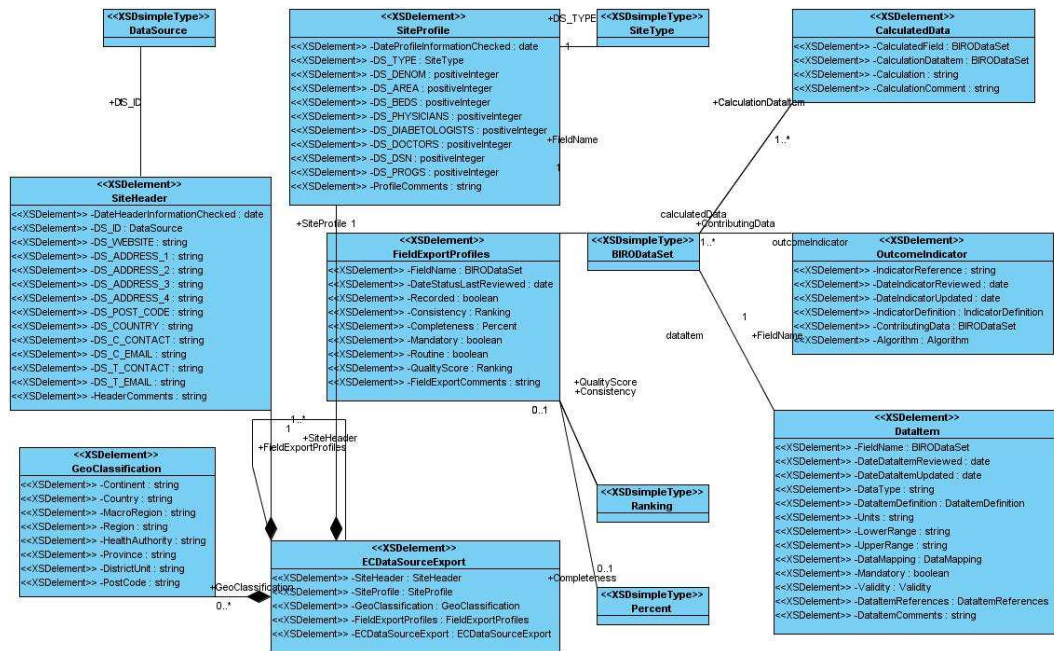
UML Class Diagrams

In support of the schema updates described above, a data model is also necessary to support partners as the XML Schema expands. New Modelling diagrams have been created to show the logical breakdown of the schema.

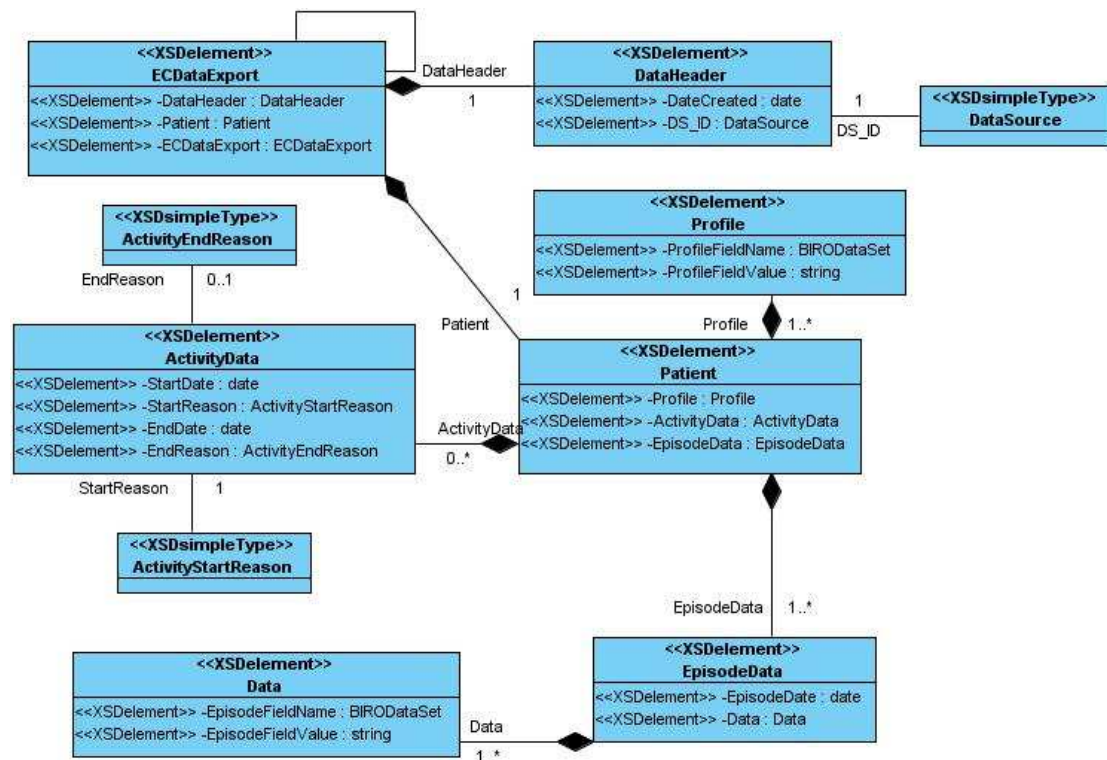
The diagram below refers to ECclinicalDefinitions.xsd:



The following diagram refers to ECDataSourceExport.xsd:



The final diagram details ECDataExport.xsd:



These diagrams will be review and described further after the EUBIROD schema review.

Appendix 1: Revised Dataset – Clinical Parameters

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 0 = Not Recorded
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	
BIRO047	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	Real	

BIRO020	HDL	HDL	Real	
BIRO046	LDL	LDL	Real	
BIRO021	TG	Triglycerides	Real	
BIRO022	RETINAL_EXAM	Retinal Examination	Enumerated	1 = Yes 0 = No
BIRO023	RETINA	Retinopathy Status	Enumerated	1 = No Retinopathy 2 = Background Retinopathy 3 = Referable Retinopathy
BIRO024	MACULA	Maculopathy Status	Enumerated	1 = No Maculopathy 2 = Referable Maculopathy
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 = Abnormal
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	Hypoglycaemic Drug Therapy	Enumerated	1 = Insulin Only 2 = Tablet Only 3 = Insulin and Tablets 4 = None (Diet Only)
BIRO040	ORAL_THERAPY	Oral Drug Therapy	Enumerated	1 = Sulphonylureas 2 = Biguanides 3 = Glucosidase Inhibitors 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	Real	
BIRO044	SELF_MON	Self Monitoring	Enumerated	1 = Urine 2 = Blood Glucose 3 = Both
BIRO045	EDUCATION	Diabetes Specific Education	Enumerated	1 = Yes 0 = No
BIRO048	DMP_ENROL	Patient Enrolment in DMP for Diabetes	Enumerated	1 = Yes 0 = No
BIRO049	AD_START_DATE	Data of commencement of period of patient activity	Date/Time	

BIRO050	AD_START_REASON	Reason for the commencement of activity period	Enumerated	1 = Birth 2 = Diabetes Diagnosis 3 = Transferred In
BIRO051	AD_END_DATE	Data of completion of period of activity	Date/Time	
BIRO052	AD_END_REASON	Reason for the completion of activity period	Enumerated	1 = Death 2 = Transferred Out 3 = Lost to Follow-up
BIRO053	LIPID_THERAPY	Lipid Lowering Therapy	Enumerated	1 = Yes 0 = No
BIRO054	ANTIPLATELET_THERAPY	Anti-platelet Therapy	Enumerated	1 = Yes 0 = No

Appendix 2: Revised Dataset – Clinical Site Parameters

Reference	Field Name	Parameter	Data Type	Enumerated Codes
BIRO002	DS_ID	Data Source ID	String(10)	
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 = National Data – Complete 12 = National Data – Sample 13 = Regional Data – Sample
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)	
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	
BIRO121	DS_PROGS	Disease Management Programmes	Integer	
BIRO122	DS_DMP_PHYSICIANS	Physicians Offering DMP's for Diabetes	Integer	

Appendix 3: Revised Dataset – Geographical Classifications

Reference	Field Name	Parameter	Data Type	Enumerated Codes
BIRO200	GC_CONTINENT	Level 0 Classification: Continent (BIRO Custom Level)	String(10)	
BIRO201	GC_COUNTRY	Level 1 Classification: Country (NUTS Level 0)	String(10)	
BIRO202	GC_MACRO_REGION	Level 2 Classification: Sub-National Area (NUTS Level 1)	String(10)	
BIRO203	GC_REGION	Level 3 Classification: Region (NUTS Level 2)	String(10)	
BIRO204	GC_HEALTH_AUTHORITY	Level 4 Classification: Local Health Authority (BIRO Custom Level)	String(10)	
BIRO205	GC_PROVINCE	Level 5 Classification: Province (NUTS-3)	String(10)	
BIRO206	GC_DISTRICT_UNIT	Level 6 Classification: District Health Unit (BIRO Custom Level)	String(10)	
BIRO207	GC_POST_CODE	Level 7 Classification: Post Code (BIRO Custom Level)	String(10)	