Facilitating Adoption of the NGOSS SID

Metabula Limited, June 2005

Intentionally Left Blank

Table of Contents

Table of Figures

Introduction

Metabula is a specialist provider of intelligent information management, integration and consolidation solutions. These solutions help to limit and reduce errors, promote consistency and increase efficiency thereby reducing project costs and times.

In particular, Metabula has specific competences in object-oriented representations, data modelling and mapping. This applies to standard and industry modelling techniques such as UML, ISO STEP representations and Application Protocol (AP)-models e.g. Oil / Gas, Telecoms, Shipping, Defence, Aerospace.

These capabilities have been applied to the telecoms industry where Metabula has focused upon the representation and management of the NGOSS Shared Information and Data Model (SID) model and comparison with client specific information models known as Shared Information Specifications (SIS).

There are however a number of challenges for organisations wishing to adopt the NGOSS standards and the SID in particular. This can vary from lack of a clear understanding of the organisation's current information models to absence of access to NGOSS knowledgeable resources.

Metabula believes that it is able to provide significant benefits in facilitating the adoption of the NGOSS (and the SID in particular) through its solutions and approaches. These include bringing together existing client systems and their respective data models into a single information model (if it is not already available), auditing the existing data usage, and even to present data directly from the source systems according to the SID model.

This provides an extremely powerful solution for those clients wanting to accelerate their adoption of the industry standard NGOSS SID model as well as understand, manage, and compare and map their business information model (an SIS) to the SID.

These capabilities are described further in the remainder of this paper and are supported by another paper from Metabula on 'Assessing SID Compatibility'.

Scenario

The particular scenario chosen utilises the SID Phase 5 model and a System Y in Company X where there are a variety of data tables constituting a data model.

The primary focus of the system is Subscribers and Equipment and how these asset groups are related. In particular there are three tables, DEM_SUBSCRIBER, DEM_SUBSEQUIP and DEM_EQUIPMENT. Although there are no joins between these tables in the database(s) it has been determined that there are links which can be held in the DataBridge system without affecting the source systems or their description.

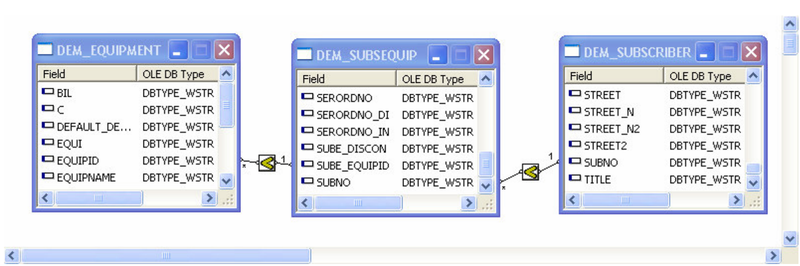

Figure 1 SystemY data

The Products

The solution described in this paper is implemented around a model repository that can hold both information and data models as shown in Figure 2. This repository will also contain the maps that link the SIS information model to the underlying physical data sources and any maps between the SIS and the SID.

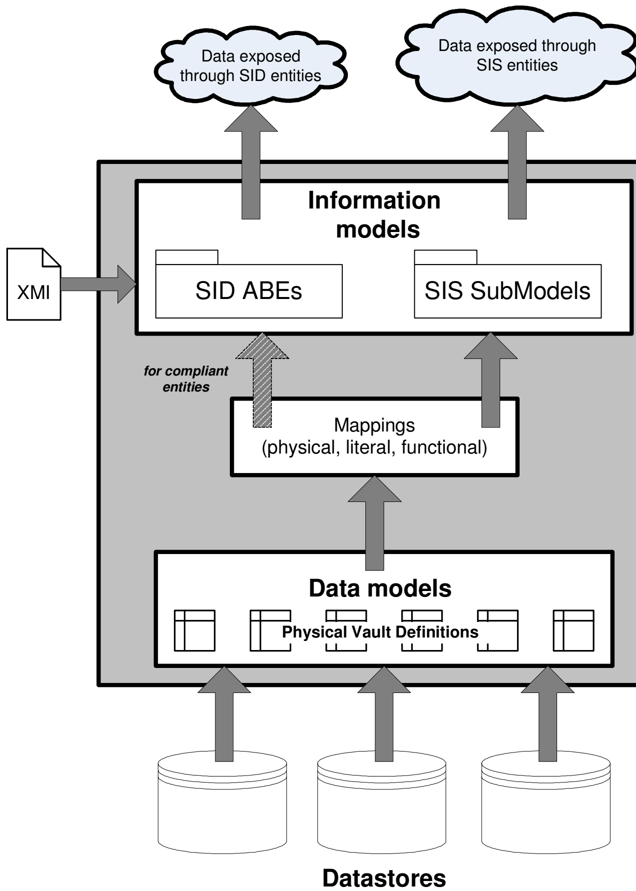

Figure 2 Information model repository

The Metabula solution utilises a variety of Metabula's products. The DataBridge product is used as the information model repository and is able to present the views on the SIS, SID, and SID extensions models. These views are then used by the Datanomic DataClean and DataAudit products to perform the overlap analysis and proposed matching with the results being loaded back into the DataBridge product.

However the DataBridge product can also act as an EII engine that is driven by three sets of model data –

- data models – defining the schema and location of external data sources
- information models – defining the entities to be exposed as "views"
- maps – linkages between data and information models.

This means that if the data models and maps have been linked to real external data sources then DataBridge will be able to present actual "instance" data "as" SID or SIS instances according to their particular entity type. This is represented in Figure 2 by the data "clouds".

The Process

The process of generating the solution requires the following steps –

1. Importing the SID information model
2. Loading the SIS information model
    a. Importing an existing SIS model
    b. Defining the SIS model from data models
    c. Representing the SIS model as subset of SID (based upon mapping data models to SID)
    d. Creating the SIS model from scratch
3. Mapping data models into SIS model
4. Testing the SIS information model against data source samples
    a. Validating the data model usage with real data
    b. Validating the information model
5. Mapping the SIS to SID
    a. Use of NGOSS Compliance levels
    b. Use of NGOSS Compatibility testing
6. Exposing data from source systems as SID data
7. Export definitions for comparison and use by other tools

The process is "mapped" against the key data structures in Figure 3 and described in more detail in the following sections.

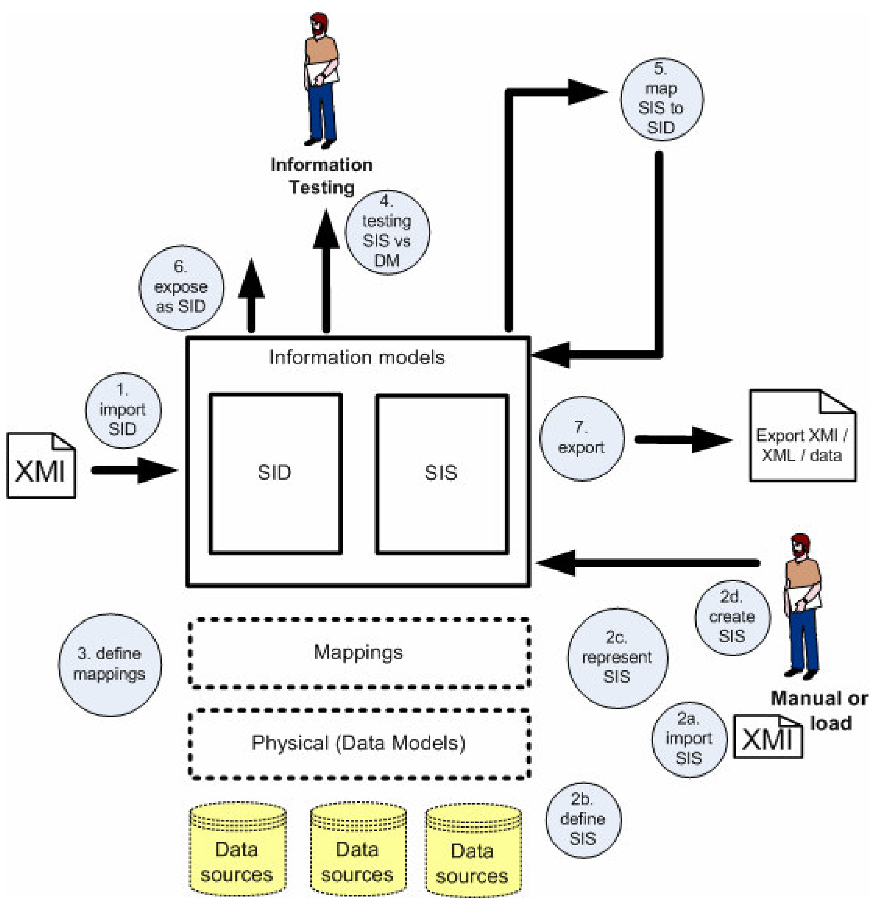

Figure 3 The mapping process

1. Import the SID information model

The first stage is to import the SID Aggregate Business Entities (ABEs) into the system using the Phase 5 XMI file.

The model is loaded into the system's Schema Management tool, Q-Schema as shown in Figure 4.

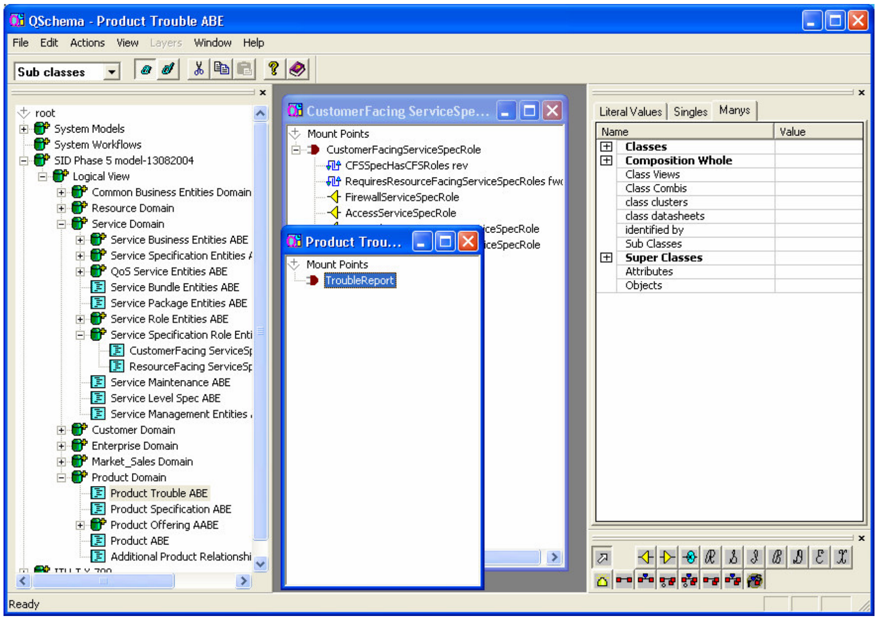

Figure 4 Q-Schema with SID Model

It should be noted however that the system will typically be pre-loaded with the SID information model before the start of any project.

2. Loading the SIS information model

This stage creates the client's information model i.e. the SIS model within the system. The client may already have an explicit information model or may need assistance in creating one from a set of data (or implementation) models.

<u>2a. Importing an existing SIS model</u>

It is possible that the organisation already has an information model representing the system to be modelled. In this case, the information model (i.e. SIS) can be loaded directly into the system. This is typically achieved through the use of XMI data files which are the representation of UML in XML according to a particular XML Schema.

The XMI files are loaded into the system using Q-Schema (see Figure 5). It supports import from a variety of XMI formats corresponding to different versions of UML.

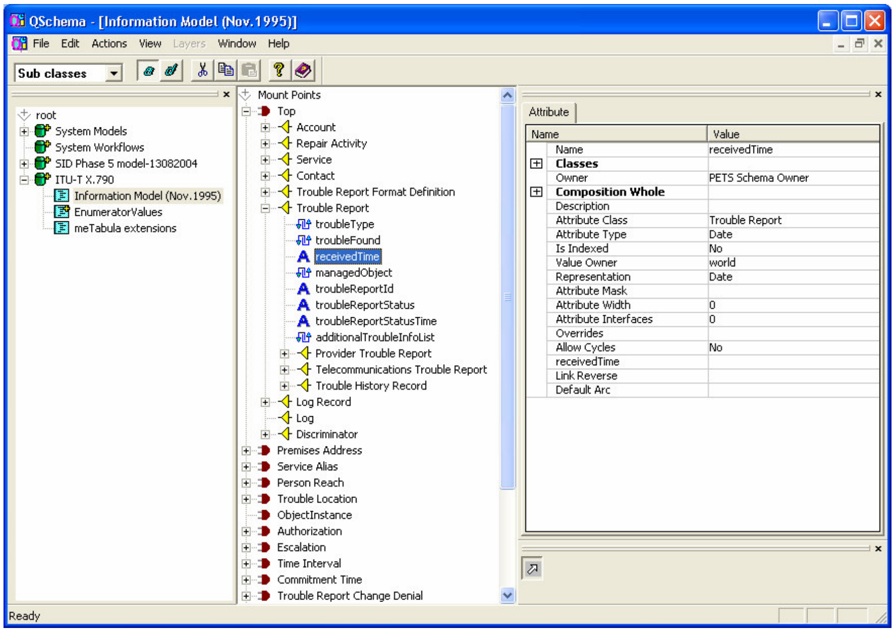

Figure 5 Q-Schema with SIS Model

Once the model is loaded then it can be viewed, managed, and exchanged as required.

2b. Defining the SIS model from data models

If the organisation does not have an information model then in this approach one needs to be generated. The information model is used as an interim staging post on the way to the SID. This means that the SIS can be fully defined or alternatively just treated as if it were only a virtual model.

This approach utilises the data models for the existing applications as the basis for generating the information model. The data models i.e. the schema for the applications are loaded into the physical tables section of the system.

This is shown in the Constructor view in Figure 6 where the three panes represent the following components

a. Top Pane – SIS model
b. Bottom Pane – data models
c. Middle Pane – mappings between the data models and information model

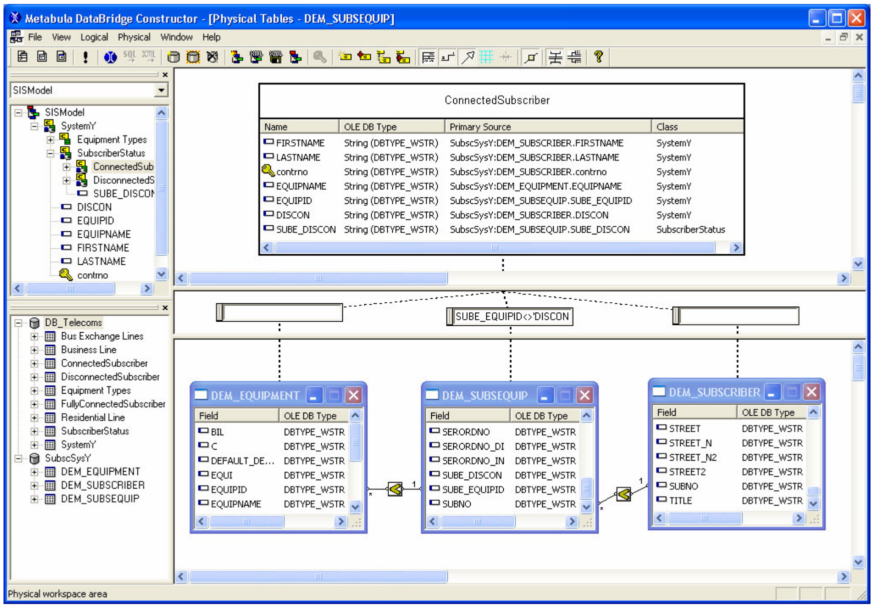

Figure 6 DataBridge Constructor UI

The description of how the mapping between data and information models is created is covered in stage 3 - Mapping data models into SIS model.

The data models can be loaded into the system in a variety of ways. These include connecting directly to the source itself if it is a suitable format such as relational database, Excel-file, or flat-file. In a number of cases, e.g. SAP or other applications with large numbers of tables this approach is not appropriate as the data available directly is meaningless in a business sense. The approach in this case is to connect through an application interface which presents a business level representation of the application's data (and model). The interface can be presented as XML, web services or even COM.

Once the data models have been loaded, then the process of generating the information model can be started. Each of the classes and attributes (as well as specialisations and generalisations) can be created manually, or by dragging an item from the physical (i.e. data models) section over to the logical (i.e. information model) section. This results in a direct representational copy of the item in the physical layer being created in the logical layer. Once this has been created then it can be modified, renamed etc.

As examples, Figure 7 and Figure 8 show the creation of a logical class as a direct representation of the DEM_EQUIPMENT table in the physical layer and the creation of a Quantity attribute directly from a DEM_SUBSEQUIP table. It should also be noted that a join has been created between the physical tables DEM_EQUIPMENT and DEM_SUBSEQUIP which does not actually exist in source systems.

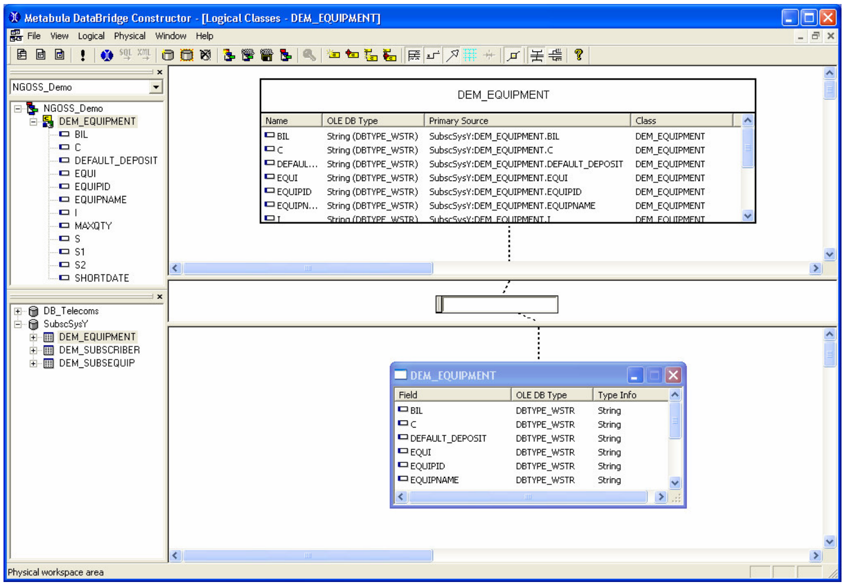

Figure 7 Create logical class

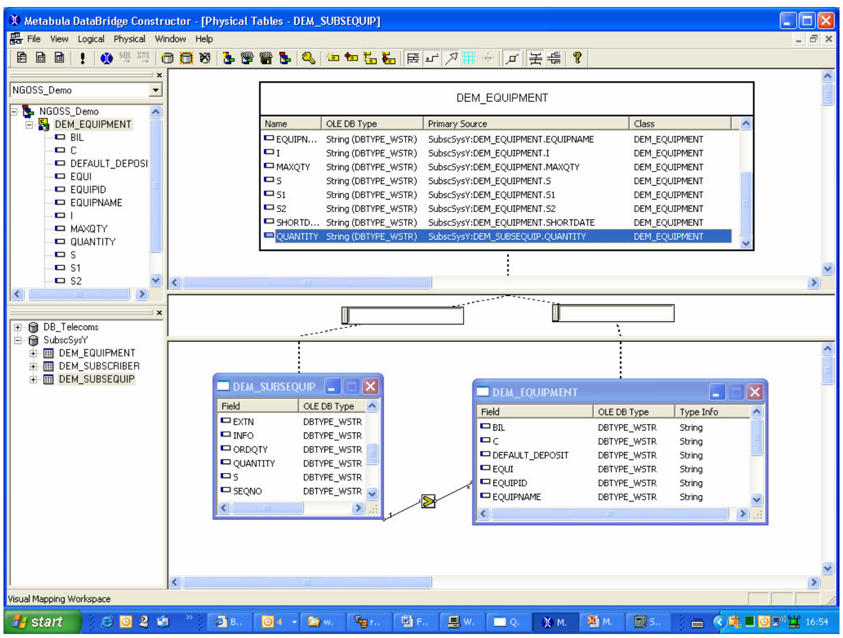

Figure 8 Create logical field

A fully populated class hierarchy can therefore be created as required to generate a representation of the important information entities and their properties in the information model.

The full details for this capability and its operation can be found in the appropriate product manual.

2c. Representing the SIS model as a subset of the SID

An alternative, where the organisation has a limited number of existing systems or where there are no distinguishable data models, is to utilise the SID directly.

This approach removes any requirement to map the SID to the SIS as this is already done. However it does mean that any existing systems will have to be directly mapped into the SID.

2d. Creating the SIS model from scratch

A final alternative is to manually create the information model in the system (using Q-Schema). This is achieved through the use of the toolbox which allows for creation of individual entities, attributes, models etc.

3. Mapping data models into SIS model

The task of mapping the data models into the SIS model can be complex. It is essential otherwise there is a complete disconnect between the data and information representations.

The Constructor application as shown in Figure 6 has a mapping layer. The mapping layer holds the maps between the physical space (i.e. data models) and the logical space (i.e. information model).

In the case where the information model was generated from the data models then the mappings, or a large majority of the mappings will have already been created.

In the other cases the mappings will need to be created manually. This can only be achieved by having a good understanding of how fields in the data models relate to attributes in the information model. In a simple case where there is a direct mapping, the relevant field in the physical table can be mapped to the correct attribute in the logical layer (as in Figure 8). The alternative, as shown in Figure 9, is to create more sophisticated mappings with functions, operations, or literals which can also combine a variety of fields.

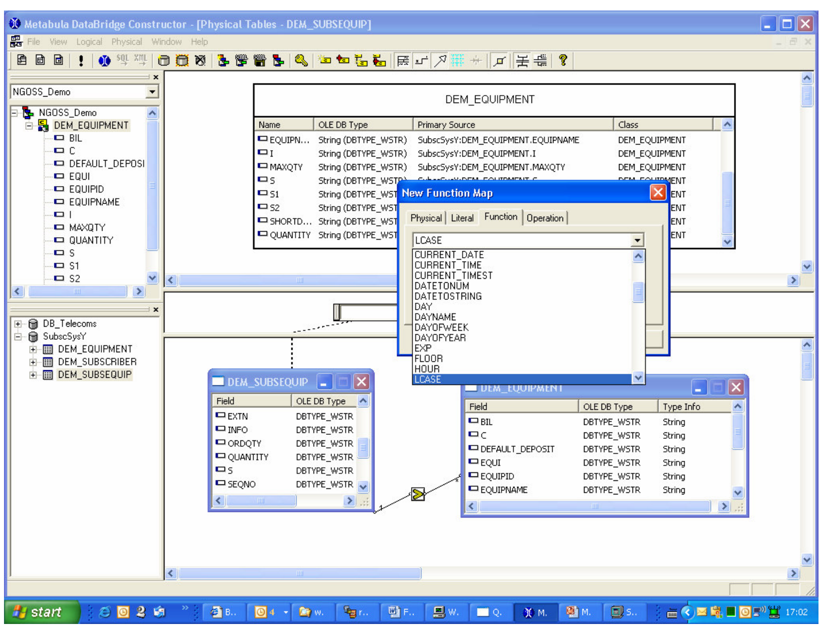

Figure 9 Create mappings

The resulting mappings are executed as SQL. Whenever a query is made for data according to the information model this query is deconstructed into queries into each of the source systems.

It is recognised that a more sophisticated mapping capability might be required in particular situations. A full mapping environment and language is available separately but it does require that all the source data is held locally in the system and the target data is then also generated locally. This is a different approach, but is necessary in particular cases.

4. Testing the SIS information model against data source samples

Once the SIS model has been created and a mapping to the data models generated it is appropriate to test the information model against the data residing in the sources. This is easy to achieve as the DataBridge system exposes two interfaces – OLEDB and XML and allows for the following tests

4a. Validating the data model usage with real data

Although the data should be validated and its quality certified, it is highly likely that the data quality of most of the source systems will be low. In addition, it is probable that there will be discrepancies between the data typing in the model (e.g. Integer, String) and the types of the actual data itself i.e. a phone number field which is expected to be numeric might contain text such as (UK) or a company name.

A tool in the Metabula suite (from Datanomic) allows for the auditing of data sources (as well as the auditing of the data generated from the information model). This is shown in Figure 10 and can perform standard and bespoke auditing tests on the sources to provide indications of typing usage, data patterns, likely key fields etc.

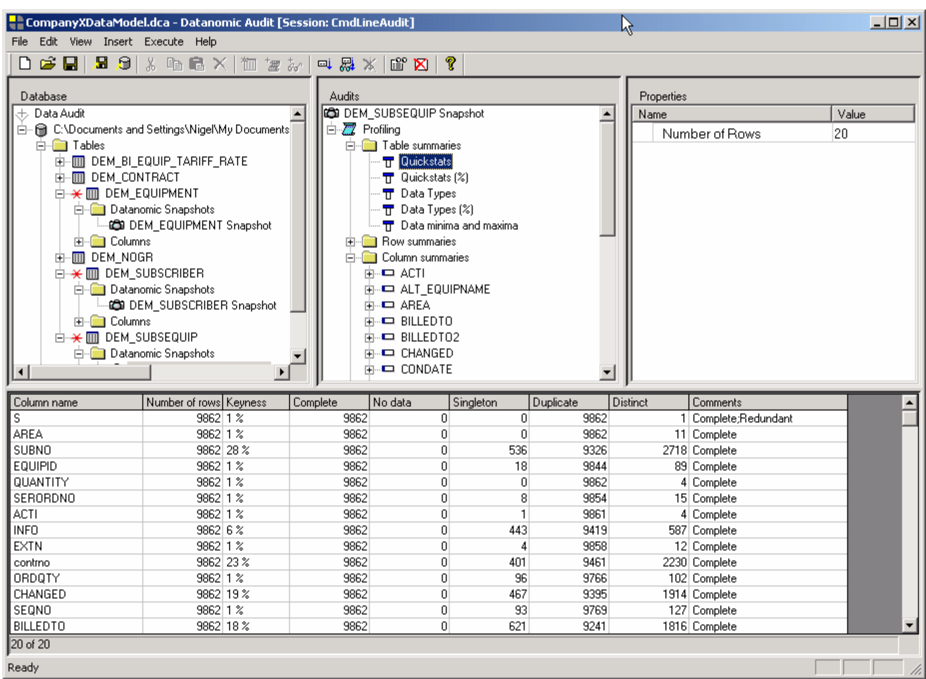

Figure 10 Audit data and information sources

This step could also be undertaken at the start of the whole process if the information model (i.e. SIS) is to be created from the data models. This provides an easy to understand analysis of the data, its actual use, and anomalies. These can then be used in constructing the identifiers, important fields, and classifications in the information model(s).

4b. Validating the information model

The importance of validating the information model representation should not be underestimated. The mappings from the data models into the information model allow data to be presented as if it were data native to the information model. Using this mechanism the types, values and validity of the mappings themselves can be tested on an ad-hoc or formal basis as required.

It should be noted that this capability allows for an evolutionary approach where class specialisations are created as data groupings are identified, and enhanced mappings created as increased understanding of the nature of the data is gained.

5. Mapping the SIS to SID

The step of mapping the SIS to the SID is covered in the related 'Assessing SID Compatibility' paper and is not repeated here. However two related areas are discussed briefly.

5a. Use of NGOSS Compliance levels

The NGOSS Compliance levels define a grading standard for the level of matching between an SIS and the SID. The levels are oriented around identifying the compatibility at class and attribute level.

5b. Use of NGOSS Compatibility testing

The result of performing an NGOSS compatibility testing task is to ascertain the NGOSS Compliance level. The task of mapping the SIS or aspects of the SIS, to the SID means that the effort required to perform the NGOSS compatibility testing is dramatically reduced. This is because a set of the entities which match across the systems is already available.

6. Exposing data from source systems as SID data

Once a mapping from the SIS to the SID has been created, which includes entities (i.e. classes) and attributes then this information can be made available through the DataBridge system.

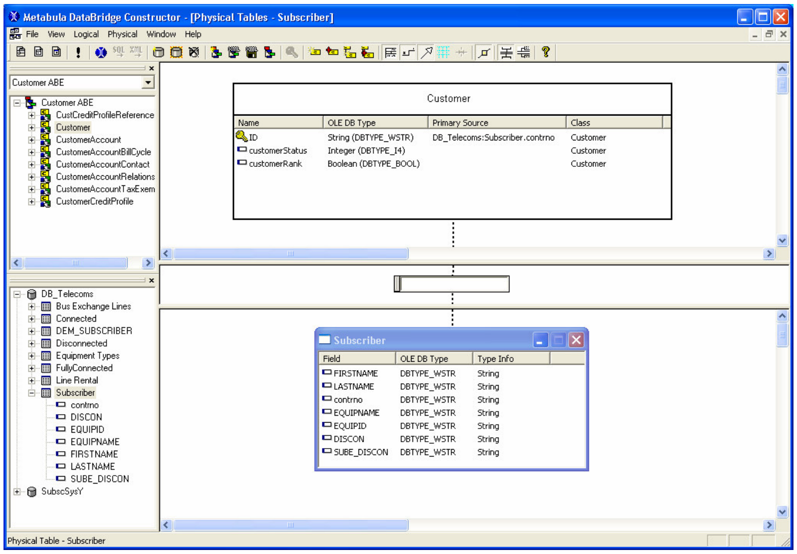

Figure 11 SIS to SID representation

In Figure 11, the Subscriber entity in the SIS has been loaded back in to the DataBridge system so that it appears as a physical system. It can then be mapped into the Customer entity in the SID. The class and attribute mapping is generated as a result of matching between the SIS and the SID.

The result of this mapping is that is the SID is mapped to the SIS and the SIS to the data models. Thus data from the source systems can be presented through the system as SID-compliant data.

7. Export definitions for comparison and use by other tools

All the information held in the system is held as data and meta-data and can therefore be exposed from the system as required for use in a variety of external systems. This might be for subsequent analysis or even for use in the definition of maps and structures to aid in the deployment of NGOSS compliant exchange environments.

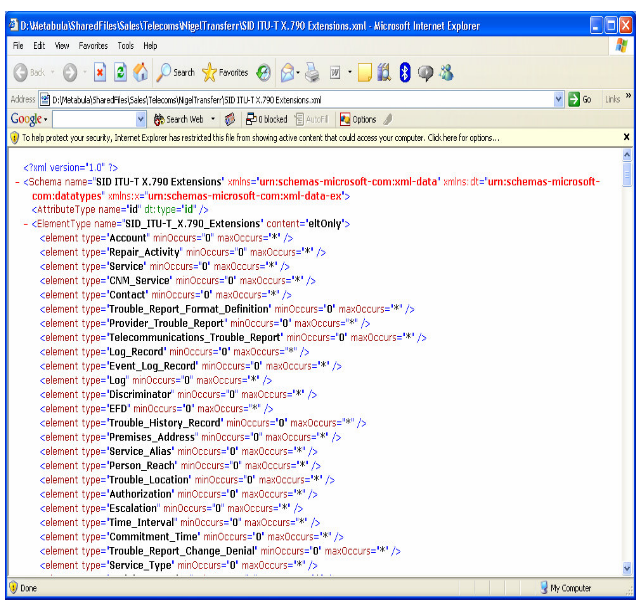

Figure 12 Export XML

The data exported depends upon the required usage, be it for comparison, be it for use in other applications etc. The format of the export can be defined also but is expected to be XML-based with the schema depending upon the requirements.

Summary

Metabula believes that the process described could provide significant benefit in the area of facilitating the adoption of the NGOSS SID. The ability to automate and electronically document the process is extremely powerful, but it also means that the resulting information models can be used to present SIS data and information according to the SID models as well as make the models available for use in other tools and subsequent phases.