STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE	COMMISSION OF THE EUROPEAN COMMUNITIES
CONFERENCE OF EUROPEAN STATISTICIANS	EUROSTAT
Joint UNECE/Eurostat Work Session on Statistical Metadata (6 - 8 March 2002, Luxembourg)	Working Paper No. 8 English only

Topic (i): Infrastructure issues for statistical metadata

USE OF XML AT STATISTICS SWEDEN: TODAY AND TOMORROW

Submitted by Statistics Sweden[1]

Contributed paper

SUMMARY

Since 1998 Statistics Sweden has had its own system for documentation of metadata, named Metadok. It is important that metadata storage be expanded and that it be accessible to as many users as possible. The intention is to facilitate the documentation work by giving the user the opportunity to import information from different sources.

XML makes it easier to export and import metadata from different software. This will result in an improved quality of metadata and will also make it a lot easier for the user.

This paper describes how Statistic Sweden will use XML to enable metadata to be imported from the modelling tool PowerDesigner to Metadok.

I.   INTRODUCTION

1.   At Statistics Sweden the tables, columns and values are documented for almost every register. The documentation registered with Metadok is stored in a Sybase database.

2.   Some of the metadata can be imported directly from other systems e.g. database schemas from Sybase, but the user still has to manually type in a lot of information.

3.   The metadata are important in the statistical process. It is used from the start when the database is created until the end when tables are created. Today, most of the metadata are registered at the end of the process. The metadata are often documented in different software and have to be typed in again. The user can easily make mistakes that can lead to lower quality of the metadata.

4.   The objective is for metadata to be registered more often during the statistical process. If registered early, metadata can be reused later in the process and the user does not have to register everything at the end of the process. It may also lead to an increase of metadata storage and that is more interesting to use.

5.   To attain this goal, it must be easy for the user to import metadata to Metadok. If it is easy to import data, the user makes fewer mistakes, and he/she can document their register earlier in the development process and update the metadata more often. At Statistics Sweden, XML will be used for exporting metadata from other softwares to Metadok. XML is structured information, containing content

---

[1]   Prepared by Catrin Saf and Eva Söderblom.

and some indication of what role that content plays. The XML specification defines a standard way to add markup to documents.

II.   METADOK

6.   Metadok is a tool that has been developed at Statistics Sweden. The first version was released in May 1998. Metadok is a tool for formalised metadata, machine-readable, of observation registers. Metadok is built in Visual Basic and the output is in Word. The information in Metadok is stored in several different local Sybase databases.

7.   Please see the graphical picture of Metadok as illustrated below.
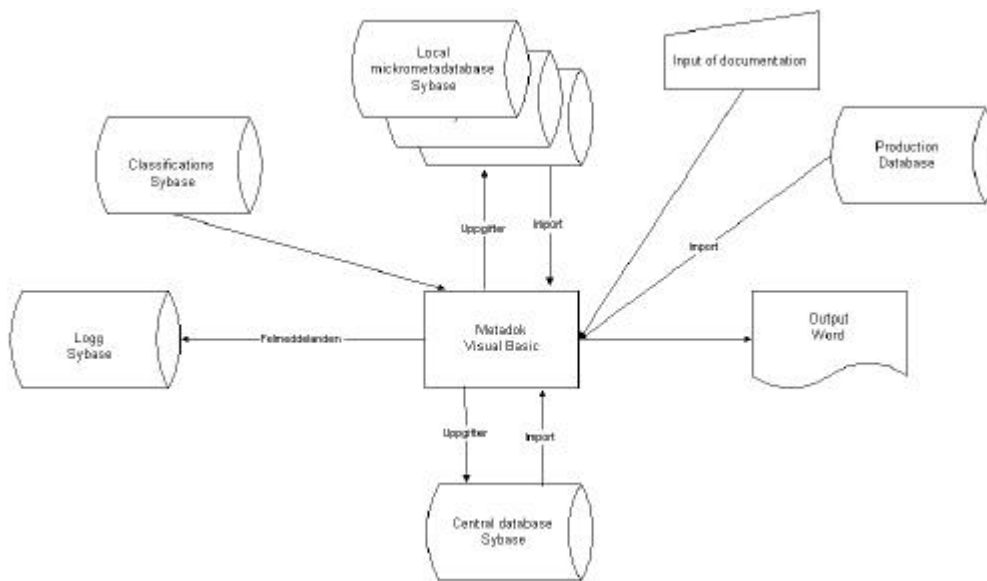
8.   There is a link between Metadok and the Classification database, enabling the importation of classifications into the documentation. Everything that happens in Metadok when users are documenting or errors occur is stored in the Logg database. The user can import information from local databases, the central database and production databases. The output from Metadok is in Word. When the documentation is finished, the data is exported to a central database. The central database is Statistics Sweden's central metadata storage. Different users and software use the metadata that is stored in the central database. The result, which is stored in the central metadatabase, is published on the Internet.

9.   Metadok describes the contents of a physical database for an observation register: versions, databases, tables, variables, value sets, values, indices, primary keys and foreign keys. There are also some other tables in the model containing codes. The database model is shown overleaf.

Mi_Register
Register
PresText
Description
Refperiod
Registertype
*SubjectCode
*ProductId
*StatResponseOrg
*Producer
Userid
Date

Mi_Version
*Register
Version
PresText
Description
*Contact1
*Contact2
Personal reg
FinalObservreg
LongTermStored
ArchivalRespite
*ContactArch
Userid
Date

Mi_Database
*Register
*Version
Database
FirstTime
LastTime
PresText
Description
Availability
Anonymous
Disclosercheck
LastUpdated
Published
Databasetype
Storagetype
Server
Path
DatabaseNamn
Userid
Date

Mi_Datamatrix
*Register
*Version
*Database
Datamatrix
PresText
Description
Objecttype
Population
Rows
FirstTime
LastTime
Userid
Date

Mi_IdKey
*Register
*Version
*Database
*Datamatrix
IdKey
Variable
SerialNumber
Userid
Date

Mi_RefKey
*Register
*Version
*Database
*Datamatrix
RefKey
Variable
SerialNumber
RefMatrix
Userid
Date

Subject
SubjectCode
SubjectArea
Fotnote
Userid
Date

Product
ProductId
ProductText
Database
Userid
Date

Mi_Valueset
*Register
Valueset
Description
ValueExists
KDBid
Userid
Date

Mi_Variable
*Register
*Version
*Database
*Datamatrix
Variable
PresText
Description
Definition
Source
Refperiod
Addable
Unit
Datatype
Startpos
Datalength
Decimals
StoreColumnno
*Valueset
Userid
Date
*x

Mi_Index
*Register
*Version
*Database
*Datamatrix
IndexNamn
Variable
SerialNumber
IndexType
Userid
Date

Organisation
OrganisationCode
OrganisationName
Department
Unit
Userid
Date

Person
PersonCode
PersonName
*OrganisationCode
Phoneno
Faxno
Email
X400
Userid
Date

Mi_Value
*Register
*Valueset
Valuecode
Sortcode
Valuetext
Userid
Date

Columncode
Metatable
Column
Code
PresText
Userid
Date

MetabaseInfo
Model
ModelVersion
DatabaseRole

10.   There are different functions to facilitate the progress of documentation work, such as importing metadata on several different levels. The user can import tables, value sets, values, indices and keys from Sybase databases. Today Metadok has a connection to the tool for designing databases, called Modellator. At Statistics Sweden, a modelling tool called PowerDesigner is used instead of Modellator and it is important that a connection between PowerDesigner and Metadok be established to smooth the documentation work.

11.   There is also a link between Metadok and Statistical Analysis System (SAS) software. This link makes it possible to import some metadata from Metadok to SAS. The aim is that it should be possible to import metadata from SAS to Metadok in the future.

III.   POWERDESIGNER

12.   PowerDesigner supports many different tasks performed by an IT-organization. PowerDesigner not only supports the design, maintenance, and generation of relational databases, but also the design of business logic used to access and manipulate database information.

13.   It provides an environment in which database designers and application developers can speak their own language and work with the model paradigm of their choice. PowerDesigner helps keep the models in synch with each other with its model generation and model merge capabilities.

14.   PowerDesigner is a modelling tool from Sybase that supports 30 different databases. PowerDesigner supports conceptual data modelling, physical data modelling, and object-oriented modelling with the Unified Modelling Language (UML). The UML support extends to use cases, and to sequence and class diagrams. PowerDesigner also supports forward and reverse engineering for SQL scripts, ERwin files, ODBC data sources, XML, and object language sources, including Java, C++, and Visual Basic.

15.   At Statistics Sweden, PowerDesigner is the recommended software. All databases at Statistics Sweden should be created and maintained by PowerDesigner. Statistics Sweden uses only two parts of PowerDesigner: the conceptual data modelling and physical data modelling. This package is called DataArchitect.

16.   DataArchitect has many additional capabilities: the ability to create conceptual as well as physical models, reverse engineering, merge models, compare models, modify database, make reports, generate test data based on the model objects, convert a physical model to a conceptual model and vice versa.

IV.   XML

17.   The Extensible Markup Language (XML) is the universal format for structured documents and data. XML is a set of rules for designing text formats that structure data. XML makes it easy for a computer to generate and read data, and to ensure that the data structure is unambiguous. XML avoids common pitfalls in language design: it is extensible, platform-independent, and it supports internationalization and localization.

18.   XML files can be created from different software, and can contain metadata. The XML file can be read by a computer or edited in any text editor software.

V.   ONGOING WORK

19.   Most registers at Statistics Sweden must be documented with Metadok. One problem with Metadok today is that the metadata is already documented somewhere, and the user has to document it once again. The metadata is important for Statistics Sweden and if it is documented in only one place, it is most likely the quality of the metadata will improve.

20.   When a system is developed, the metadata is first documented in PowerDesigner. Metadata about the physical database such as tables and columns are documented in PowerDesigner. When the database is completed in PowerDesigner it can be created in Sybase. XML enables metadata to be imported from PowerDesigner to Metadok.

21.   It is possible to obtain an XML file from PowerDesigner with metadata about the database. The model has to be saved in XML format. Metadok is able to show the information from the XML file. All or a part of the metadata can be imported to Metadok. This will also make it possible for the user to add a table in PowerDesigner and to update the metadata in Metadok with this information only.

22.   To enable metadata import from PowerDesigner, a new COM+ component will be built. The component will find the path to the XML file and will return metadata from the file. The user can then choose which tables, columns, etc. will be imported to Metadok.

VI.   Other projects

23.   Statistics Sweden is aware that XML will play a great role in the future. Besides the project described in this paper, there are two other interesting projects that involve XML as a part of the system solution. These two projects are not yet in production.

VI.1   Publishing on different mediums

24.   One project is to investigate how Statistics Sweden can use XML when publishing in different mediums. The idea is that a spreadsheet should decide how the output will look. The same information could be presented on paper, on the Internet or as an SMS-message.

VI.2   Electronic forms

25.   To gather information for a statistic examination, web forms are used. XML is used for sending data from the web form to the server.

VII.   USE OF XML TOMORROW

26.   The use of XML will probably increase. In the future, metadata can be used with other software if the import function of metadata from PowerDesigner to Metadok works well. The component that reads from the XML file could be used for other software. If metadata can be saved in XML format, it will be easy to import to Metadok.

27.   To build a central metadata storage, XML can ease the progress of gathering and storing information because XML is not dependent on any platform or software.

VIII.   CONCLUSIONS

28.   XML is making it easier to export and import metadata from different software. This will result in a higher quality of metadata and will also make it a lot easier for the user.

29.   The intention is to have a central database containing all metadata. A step in that direction is to develop the COM+ component. But even if the user can easily document the register, there is a lot more work to be done. It is a great challenge to assimilate metadata into daily work.

REFERENCES

www.w3.org
www.sybase.com