

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Luxembourg, 9-11 April 2008)

Topic 2 (iii) Metadata and the statistical cycle and Implementation

CASE STUDY: CENTRAL BUREAU OF STATISTICS OF THE REPUBLIC OF CROATIA

Submitted by Croatia¹

Revision History.....	2
Organization Details.....	2
1. INTRODUCTION.....	2
Metadata strategy.....	2
Current situation.....	2
2. STATISTICAL METADATA SYSTEMS AND THE STATISTICAL BUSINESS PROCESS.....	3
2.1 Statistical business process.....	3
2.2 Current system(s).....	4
2.3 Costs and Benefits.....	4
2.4 Implementa-tion strategy.....	5
3. STATISTICAL METADATA IN EACH PHASE OF THE STATISTICAL BUSINESS PROCESS.....	5
3.1 Metadata Classification.....	5
3.2 Metadata used/created at each phase.....	7
3.3 Metadata relevant to other business processes.....	7
4. SYSTEMS AND DESIGN ISSUES.....	7
4.1 IT Architecture.....	7
4.2 Metadata Management Tools.....	10
4.3 Standards and formats.....	13
4.4 Version control and revisions.....	13
4.5 Outsourcing versus in-house development.....	14
5. ORGANIZATIONAL AND WORKPLACE CULTURE ISSUES.....	15
5.1 Overview of roles and responsibilities.....	15
5.2 Metadata management team.....	15
5.3 Training and knowledge management.....	15
5.4 Partnerships and cooperation.....	15
6. LESSONS LEARNED.....	16

¹ Prepared by Maja Ledic Blazevic (majalb@dzs.hr) and Branka Cimermanovic (BrankaC@dzs.hr).

METIS COMMON METADATA FRAMEWORK (CMF) PART C CASE STUDY

CROATIA / CENTRAL BUREAU OF STATISTICS OF THE REPUBLIC OF CROATIA

Revision History		
Date	Section(s) updated	Comment
19/02/2008		Version 1

Organization Details

Organization Name Central Bureau of Statistics of the Republic of Croatia

Number of staff 422

**Contact person
(for Metadata)** Maja Ledic Blazevic
Senior IT Adviser
majalb@dzs.hr
+385 1 48 06 257

1. INTRODUCTION

Metadata strategy There were three principal reasons to implement metadata in the Croatian Central Bureau for Statistics (CBS):

- to standardize definitions across all statistical activities
- to move the production of statistics closer to the subject-matter experts in order to speed up the statistical survey life cycle
- to present statistics on internet along with its context in order to make statistics understandable and available to users of all types, i.e. to extend the use of statistics beyond the usual statistical publications

The strategy document was prepared in CBS already in 2001¹. In 2002, a framework agreement was signed between the Division for Western Balkans at the Swedish International Development Cooperation Agency (Sida) and Statistics Sweden's International Consulting Office (ICO). In this framework the Swedish Statistical Office (SCB) provided support for the creation of the public macro database and a central metadata repository in CBS. In its final phases (2006 - 2007) the project was extended to support the development of the Integrated Statistical Information System (ISIS).

1. Zdenko Milonja: Information System Development Strategy, CBS, July 2001.

Current situation The central metadata repository (CROMETA) is the essential part, the core of the Integrated Statistical Information system (ISIS) which is in the final stage of development. In other words, ISIS is developed upon CROMETA.

The original idea - to develop an automated statistical survey processing system on the client/server platform - resulted from the operational circumstances in CBS:

- the IT sector is strictly centralized, i.e. IT sector processes all statistical surveys upon the descriptions laid out by statisticians.

- the majority of statistical surveys are still processed on the mainframe
- the majority of surveys have similar processing stages (data entry, validation, correction, tabulation, dissemination); therefore the majority of corresponding data processing jobs have similar structure which could be incorporated in a generalized solution. Such a solution was developed in CBS for data processing on the mainframe in the 80s and is still in use.

The *metadata repository* must contain all the necessary information to be used as parameters for a general 'program' that produces specific *operating procedures* for particular surveys. Therefore it could be stated that centrally stored metadata could more or less automatically 'drive' the statistical production system. This is the basic purpose that initiated the metadata system development in the first place. Naturally, the idea was extended to cover all aspects of statistics, as laid out in the Reference ModelTM resulting from the MetaNet project within Eurostat (2000-2003).

The CROMETA model contains Reference ModelTM concepts extended and customized for CBS needs as well as specifics of a previous CBS metadata model and specifics needed to run PC-Axis as the main dissemination tool. Although the model is very complex and rather demanding to comprehend, it proved to be well conceived from the beginning, or rather from the moment we fixed the 'big picture'. Now the metadatabase is stable, with high tolerance for occasional changes that occur along with development of specific solutions for particular stages of the statistical life cycle.

The central metadata repository is presently rather empty since it is still in test phase, it contains data on just a few statistical surveys that were selected as pilots. We are well aware of the problems which may arise among statisticians with the obligations to enter or transfer all the 'knowledge' of all statistical activities.

2. STATISTICAL METADATA SYSTEMS AND THE STATISTICAL BUSINESS PROCESS

2.1 Statistical business process The majority of CBS's statistical surveys are processed in following steps:

1. Planning all activities
2. Survey design and description
3. Data capture and file transfer
4. Validity checking against preset rules and producing error-list to be presented to the statistician in charge
5. On-line correction
6. Tabulation, i.e. producing statistics for statisticians supervision
7. Publishing, i.e. producing first releases and other statistics
8. Archiving
9. Monitoring

It should be noted that mentioned steps refer to the *processing* stage of any particular statistical survey or, in other words, to the statistical survey as seen from the perspective of the IT sector. Mapped to the CMF model, the steps described above fit the model as follows:

Common Metadata Framework Generic Survey Life Cycle	CBS' Survey Processing Cycle	
Need		9. Monitoring
Develop & Design Build	1. Planning all activities 2. Survey design & description	
Collect	3. Data capture & file transfer	
Process	4. Validity checking 5. On-line correction 6. Tabulation	
Analyse		
Disseminate	7. Publishing	
Archive	8. Archive	
Evaluate		

2.2 Current system(s)

By dividing the main tasks as described in 2.1, CBS designed and developed a software that could produce (*generate*) programs needed, based on the on-line input of parameters. In line with that, CBS also designed a software solution called Planning that represented a rudimentary metadata system 20 years ago. The statistician would enter basic data on the survey (owner, periodicity, basic descriptions etc.) while programmers enter other data (record definitions, processing order of particular steps etc.). On the basis of all these data some programs are automatically produced and some programs still need to be done tailor-made. Once the necessary software had been produced, the programs are run either by the statistician(s) or by the IT expert in charge. For large surveys, steps 4 - 6 are carried out by a special data editing unit.

The major part of these processes are handled by the IBM mainframe, using a very old DBMS (CA Universe), PL/I programs, IBM Job Control and CICS transaction monitor. Since the client/server technology was introduced several surveys were shifted to the new technology using MS SQL Server and Visual Basic 6.0 development tools. However, all of this was handled in the tailor-made fashion.

The endeavour to develop and implement a similar solution to the client/server environment has been ongoing for several years already. The data entry module has been developed in MS Visual Basic 6.0 together with the data transfer module. Some of the tasks of the planning module have been covered by the comprehensive metadata project while other tasks have been handled through the project designated to establish a public macro database. In addition, a statistical business register has been developed and is in the final testing phase.

It is very relieving to realize that a comprehensive solution is almost ready now, that is, a metadata-driven, more or less automatic statistical information system, built upon a comprehensive metadata management system.

2.3 Costs and Benefits

Although the development of the new ISIS and the CROMETA in particular required a significant amount of resources, mainly from the IT sector, it is expected that positive effects on statistical production will by far outnumber the resource usage.

CROMETA will add functionality as well as quality to the new system through at least following aspects:

- well-described and uniformed metadata, i.e. all information on statistical surveys and statistics in one place
- using same classifications, registers and address lists (nowadays, they are different for each survey and not maintained centrally)
- speed up all the processes in the statistical survey life cycle and bring down the time needed to establish a new survey
- better control over statistical surveys in general and data processing in particular

Thereby CBS will provide a better image within the public sector in Croatia and international

statistical society as statistics will be more accessible.

It should be noted that a significant extra effort will be required from statisticians to get used to the new methodology of survey maintenance and especially to provide all metadata necessary. It is expected that some statisticians will regard the new system simply as extra work, while some will, hopefully, gladly meet the challenges and benefit from that.

The profitability of the system should be regarded through direct and indirect revenues against all costs (investments in HW and SW, staff, maintenance etc.). It is rather difficult to separate the costs/benefits of the CROMETA system from the whole ISIS; therefore the costs and revenues are calculated for the ISIS in whole.

Direct revenues will consist primarily of the savings from the eventual shutdown of the mainframe, i.e. nearly 300.000€ yearly in terms of software licenses and maintenance. If the solution proves to be successful, it might also provide some revenues from international statistical exchange, since it is a usual practice among NSO's to exchange solutions for a reasonable amount of money. Also, CBS's IT staff is already engaged in consultancy to other NSO's. This will be enhanced further by the experience gained through system development. Consultancy also brings direct revenues to CBS. Indirect revenues may be gained from the lower maintenance cost of the client/server equipment against the maintenance cost of the mainframe which will be shut down eventually.

A significant benefit will be as much as 20 IT experts made available for new tasks. This number of people is heavily engaged in regular statistical production now and will be freed for other tasks and further development.

The costs are spread from equipment to the in-house staff and technical co-operation. It is expected that about 250.000€ will be spent altogether on hardware and licenses, the majority of that is already installed. The cost of IT education in the regular budget should be increased by 50% since there are more young developers to replace those who left CBS. There are no extra staff cost for the in-house development since it is done as the regular duty, but more developers should be hired or educated for the same reason as above.

The majority of investments were applied through technical support by Sida and Statistics Sweden from year 2002 to 2007. Additional help in terms of hardware purchase will be received from programs PHARE 2005 and PHARE 2006.

2.4 Implementa- tion strategy

The new system will be introduced in a step-by-step manner. Five surveys were first selected as pilots. The statisticians in charge of these pilot surveys participated in the development teams and provided metadata to CROMETA. Upon presentations and introductions of the pilots new set of surveys will be discussed. Manuals, workshops and training will be provided. Eventually, a deadline will be set for all statisticians to enter metadata into CROMETA using the maintenance tool. The shift of surveys from 'old' to 'new' environment will depend on periodicity and urgency of particular surveys.

3. STATISTICAL METADATA IN EACH PHASE OF THE STATISTICAL BUSINESS PROCESS

3.1 Metadata Class- ification

Croatian Bureau of Statistics developed its own metadata model called CROMETA based on Reference Model™ from the MetaNet project of Eurostat. It includes also some specific metadata from CBS and has nine groups or sections of metadata. Metadata objects in each section are closely related, but there are relationships between metadata from different sections.

1. **Organizational structure** - all metadata as regards the statistical office, related organizations, persons working within the organizations and the responsibility of the latter, is kept within this metadata section.
Examples of metadata objects belonging to this section are *organization*, *person* and

contact person, etc.

2. **Variables and measurements** - this section contains metadata about the variables collected within the frame of the statistical activity, as well as the methods and ways of measuring them. In this section the variables are described from a more general point, regardless of use and implementation in different studies/surveys. Examples of metadata objects belonging to the section are *global variable, object variable, value domain, measure unit, etc.*
3. **Studies and questionnaires** - all metadata as regards studies, their versions, general methods for performing them and other, is kept within this metadata section. (*Study* is a term used for more broad definition of statistical activities, and includes surveys and other activities). Examples of metadata objects belonging to this section are *study, study version, questionnaire, question, interview method, population, coverage type, etc.*
4. **Classifications** - this section contains metadata as concerns classifications used in statistical office. Examples of metadata objects belonging to the section are *classification family, classification, classification version, classification item, correspondence table, etc.*
5. **Publications** - contains all metadata concerning publishing and dissemination of statistics. Examples of metadata objects belonging to this section are *publication, edition, publication series, etc.*
6. **Processing and validation rules** - the *Variables and measurements* section describes the variables and the way of measuring them independent on implementation in a study/survey, and the *Studies and questionnaires* section handles all the metadata about studies and their methods. This metadata section could in some sense be described as the instance of variables and measurements put in the context of a study (or study version). E.g. the section contains all metadata as regards the methods for processing the studies, including validation, production process, registers/cubes/tables created etc. Examples of metadata objects belonging to the section are *context variable, data collection, statistical process, rule, matrix, register, cube, table, etc.*
7. **General** - general characteristics of metadata objects (which are often metadata objects themselves) are kept within this metadata section. Furthermore there are some metadata objects that are used within all or at least several other metadata sections, and therefore cannot be placed directly in any of them. Also these metadata objects are kept within the *General* metadata section. Examples of metadata objects belonging to the section are *language, keyword, footnote, status, theme, statistical object type, etc.*
8. **Access and authorization rules** - this metadata section stores metadata about who can do what with metadata and the data owned by CBS. Examples of metadata objects belonging to the section are *user group, privilege, access condition, access form, etc.*
9. **History and version handling** - metadata has a certain life-cycle as determined by the methods defined by the statistical office. This means that all metadata objects may exist in an indefinite number of versions and that the history management is extremely important in order to keep consistency in metadata and data over time. The main part of the history and version handling is implemented through methods used by other metadata objects, but one example of a metadata object belonging to the section is *update information*, used for logging all changes made to a metadata object over time.

This classification of metadata could be mapped the other classifications: e.g. navigational metadata as keywords can be found in general section, or quality metadata is included in the section of studies and questionnaires.

3.2 Metadata used/created at each phase

Today metadata management is done at several places, but intention is to create one central metadata repository and use a tool for metadata browsing and maintenance that should cover all the aspects of management. Metadata was first collected to fill the repository in metadata sections of organizational structure, general metadata, studies and measurements. There were several sources of metadata for this initialization: Annual Implementation Plan, database about publications, so-called inventory of metadata that was done before going to developing metadata model, etc.

Then metadata is available for other other applications that are part of statistical data production. The most important consumer of metadata is so-called **Survey processor** application. This is a metadata-driven application for automatic survey processing. *Survey processor* is not only a consumer but also a producer of metadata. Metadata is captured during survey processing because it is the natural place where processing metadata is created. This metadata is immediately stored in the central repository. When a survey is processed, then following metadata is created: populations, questionnaires, questions, context variables, matrix variables, matrices, registers, cubes, tables, etc. This metadata can be easily used in metadata management tool to create new versions for next processing periods.

Data processing is the most important one from all phases of statistical business process where metadata is used and created. Initial metadata about surveys for processing phase are prepared in planning phase, but handled in Annual Implementation Plan that is independent from metadata management tools.

Additionally, some metadata are produced in dissemination phase (on publications). Hopefully more metadata will be collected in other phases of business process as the management tools develop in the future.

3.3 Metadata relevant to other business processes

Metadata sections *Organizational structure* contains metadata that could be used for management process. This means that persons who are responsible for specific processes are stored as contact persons for them (not only as contact persons regarding the statistical data and publications). Contact person has a role that can be for example *head of department*.

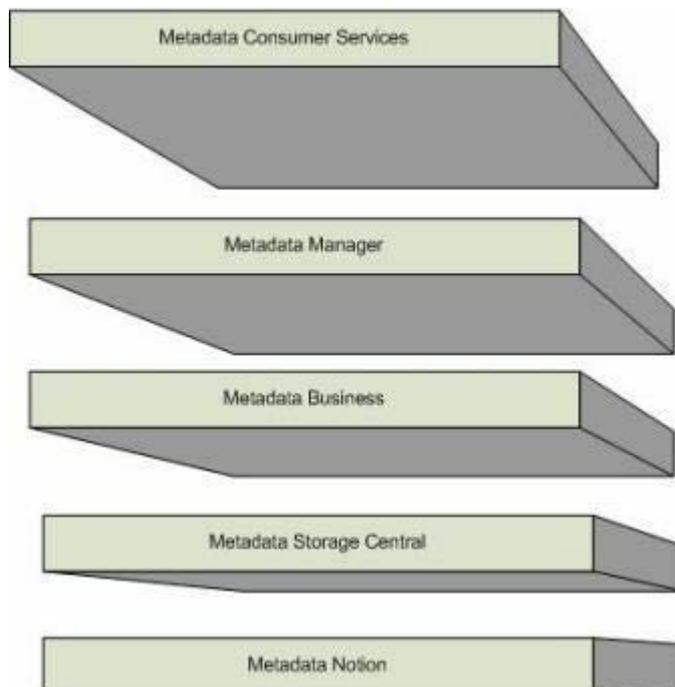
Section *Access and authorization rules* contains metadata that provide control over the access and management of statistical processes. User group has certain privileges that allow to group members to access and run specific process on statistical data.

4. SYSTEMS AND DESIGN ISSUES

4.1 IT Architecture

CROMETA METADATA SERVER

CROMETA Metadata Server is a comprehensive solution that covers all aspects of a modern metadata repository. The server is not designed for methodologists, statisticians nor systems developers solely, but provides each key unit of a statistical office with appropriate functionality through its various modules. Using the server, the methodologists can rest assured that a well-established and constantly developing theoretical model on statistical metadata has been brought into use. At the same time, the owners and producers of statistical surveys will enjoy a user-friendly Windows-like interface for maintaining and browsing their metadata. As far as concerns the system developers, an XML-based web service ensures a straightforward, practically code-free integration with existing as well as new metadata consuming applications using the centrally stored metadata. Even the system administrators are not needed to intervene during the implementation. The whole server uses the latest and most established tools and techniques around.



The comprehensive CROMETA Metadata Server

CROMETA Metadata Notion

Basically, CROMETA Metadata Notion is the *model* on which the complete metadata server resides. This generic model inherits its theoretical base from the MetaNet Reference ModelTM on Metadata, which was originally developed within EUROSTAT in order to reach a common model on metadata in general, and a widespread terminology in particular. All terms used in order to describe various metadata concepts within CROMETA Metadata Notion are directly derived from the EUROSTAT model. Given the fact that the terminology was elaborated by experienced statisticians representing major national offices, it is fairly easy for any statistician to rapidly grasp and understand the terms as well as their conceptual meaning and relations within the model.

CROMETA Metadata Notion is purely conceptual, hence could be implemented on any technical platform.

CROMETA Metadata Storage Central

The conceptual metadata model covered through CROMETA Metadata Notion is implemented in practice through the CROMETA Metadata Storage Central. This is the basic storage point of metadata for any implementation of the metadata server.

A huge advantage of the storage engine is that it is not limited to run on a specific platform, furthermore, it is completely open to use for any purpose. In practice this means that the system developers may use the underlying data source directly in order to connect and integrate it with other solutions. Avoiding encapsulation into some hidden storage format enables the office to use the DBMS already purchased or to choose one in line with the budget. Moreover, built-in functionality of the DBMS chosen could be used for replicating and synchronizing data all over the system. Obviously, this approach also provides the office with possibilities to further customize the metadata model, should there be a need for it.

As mentioned, the CROMETA Metadata Storage Central could be customized for any database platform; however, the current implementation is developed for the Microsoft SQL Server platform. At present, the system runs on the Microsoft SQL Server 2005 DBMS.

CROMETA Metadata Business

CROMETA Metadata Business is a business tier that implements the conceptual model from CROMETA Metadata Notion as well as storage-specifics from CROMETA Storage Central through an object-oriented approach. Included in the business model is a fully documented UML class diagram, clearly displaying the metadata classes, objects, methods etc. available through the metadata server.

Basically, CROMETA Metadata Business puts all features such as version handling, multi-lingual support, etc. into play by applying the appropriate business rules. For the ordinary user, the use and work of this module is obviously completely invisible. However, when using the maintenance tool - CROMETA Metadata Manager for browsing or editing metadata, or when using metadata for tabulation from the website, CROMETA Metadata Business takes care of all the background processing enabling the various features to be put into play.

For the developer engaged with integrating various metadata consumers with the central metadata server, the CROMETA Metadata Business is essential. By using the CROMETA Metadata Business together with the CROMETA Metadata Consumer Services, the task of creating a real Integrated Statistical Information System (ISIS) has been considerably simplified. Connecting surrounding applications with the metadata server is basically just one click away. For additional information regarding integration and how to put metadata consumers in connection with the central repository, see CROMETA Metadata Consumer Services.

CROMETA Metadata Consumer Services

The CROMETA Metadata Consumer Services offers an uncomplicated and swift way for integrating existing as well as future metadata consumers with the metadata server. By exposing all metadata objects with belonging properties through an XML-based web service, the whole metadata repository is made available. Virtually any application requiring metadata can reach it in a practically code-free manner. Using the CROMETA Metadata Consumer Services, the system developers can connect to the metadata server and use its metadata just by referencing the web service and use the predefined methods for retrieving and editing metadata.

Consumer services that are considered important to be implemented soon:

- Exposing classifications through the official web site
- Displaying the organization and areas of responsibility, contacts, etc
- Publishing statistics using metadata for browsing purposes
- Delivering publications including data and metadata

All of these are features that could be implemented in a straightforward manner using the CROMETA Metadata Consumer Services.

CROMETA Metadata Manager

Most people will access the CROMETA Metadata Server through the CROMETA Metadata Manager. This is the tool which provides a memorable, graphical user-friendly interface for adding, browsing, editing and generally maintaining metadata. Basically everything that could be done to metadata could be achieved here; given you have the appropriate privileges to the metadata you are aiming for.

The CROMETA Metadata Manager tool runs on the following platforms: Microsoft Windows 2000, Microsoft Windows XP, Microsoft Windows 2003 Server, and Microsoft Windows Vista.

Basically, CROMETA Metadata Manager features a Windows-like interface. Any user familiar with using Microsoft Windows should be able to acquaint with the tool in no time. For a more detailed description of this tool, see Metadata Management Tools (4.2)

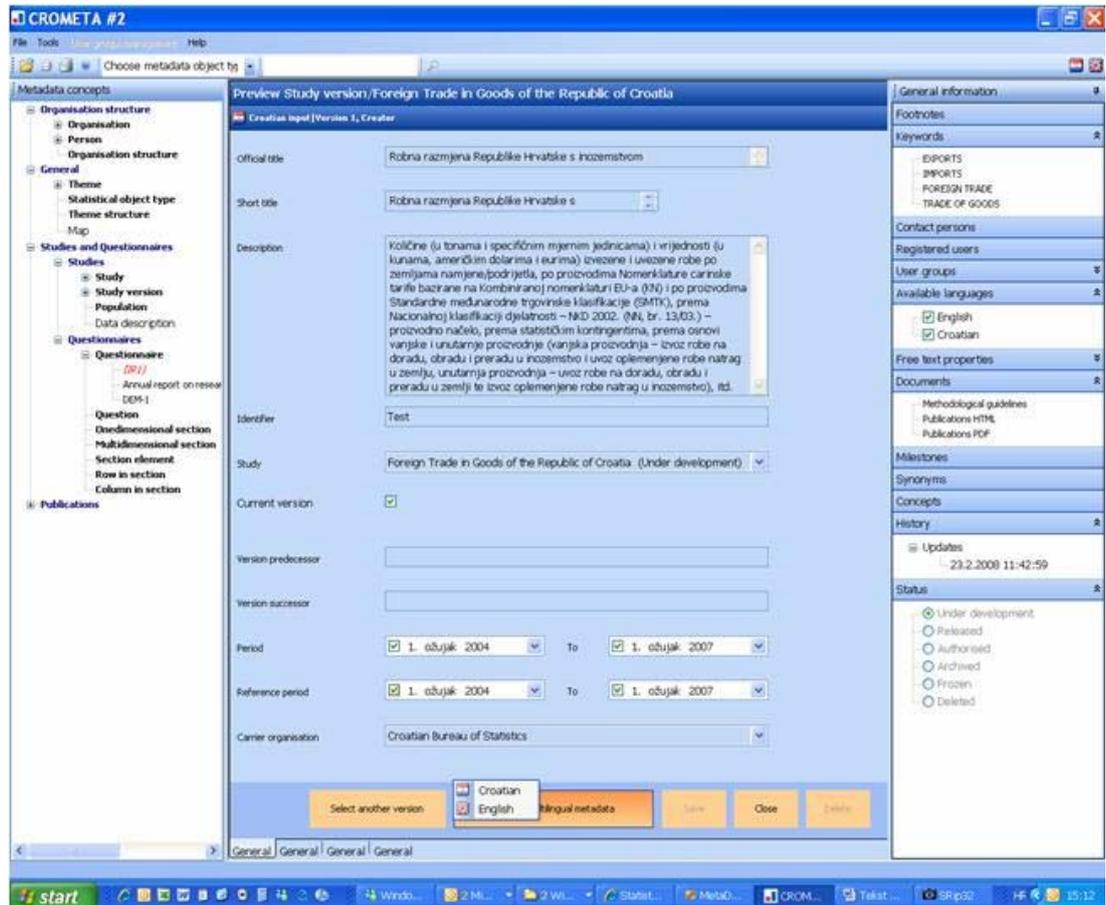
Technical Platform for CROMETA Metadata Server

As regards technical platform and development environment, the conceptual data model for CROMETA has been developed using Sybase Power Designer 9.5. The same software has been used for creating the physical data model that has been implemented on the Microsoft SQL Server 2000 RDBMS. Consequently, all database development has been carried out on the Microsoft SQL Server platform, while the CROMETA maintenance tool has been developed in Microsoft VB.NET. All object and use case modeling have been performed in Microsoft Visio. For managing the development in a shared environment, Microsoft Visual Source Safe has been used, while for project management purposes, Microsoft Project and SharePoint Portal Server have been applied. Since the development is still ongoing, the solution was upgraded and tested on Microsoft SQL Server 2005.

4.2 Metadata Management Tools

As part of the comprehensive CROMETA solutions, a maintenance tool has been developed, providing a user-friendly interface to the central metadata repository. It is expected for the tool to be developed further and extended with additional features that may be missing presently.

CROMETA Metadata Manager provides numerous functions, out of which the most obvious are to add new, edit and delete metadata from the central repository. It is difficult to describe the tool in words only; most probably the only way to get acquainted with it for real is to try it out in practice. However, some of the outstanding features of the tool are listed in this chapter.



Easy editing through graphical interface with multi-language support

Multi-language support

CROMETA Metadata Manager supports an indefinite number of languages, meaning that the languages desired for entering and maintaining metadata could be defined through the interface. All textual properties will then be open for browsing as well as editing in the languages defined. In practice this means that the office may decide to enter metadata in five languages, for example Croatian, English, Russian, French and Spanish. Obviously, all statistics may then be published dynamically in any of these languages. Furthermore, also the interface itself supports several languages; hence it could be customized for any language.

Versions management

All metadata objects may exist in numerous versions. Each version must be of one and only one state, however only one version of each object could be current/authorized. Versions will be described in detail in chapter 4.4.

General functions

All metadata, despite type or usage, could be added, edited or deleted using the exact same methods in the maintenance tool. Reusing methods between different types of metadata secures a tool that is recognizable and easy to use.

Add new version based on an existing version of a metadata object

Sometimes the differences between versions of a metadata object are minimal. For example, a questionnaire used within a survey one year, may only differ in terms of one or two new questions being added compared to the previous year.

To minimize the work within respect to this matter, CROMETA Metadata Manager supports "Add new based on", which means that all properties and connections of an existing version of an object is copied into a new object that could be further edited.

Taking the example with the questionnaire, it means that all questions, the questionnaire layout, validation rules, etc. are being copied from the previous questionnaire version. Obviously this saves a lot of time and effort. Basically the person responsible for the questionnaire only has to add the new questions in order to have it ready for processing.

History management

Besides the version management, the history of metadata could be displayed through the update log. Every time a change is made to metadata, this is logged with modifying user and date. The modifying user may also optionally supply a description of the change made. Through the history management, metadata may be studied and followed over time, including all changes, explanations, involved users and its lifetime cycle.

Authorization

In order to secure the metadata, a full-covering model for access and authorization is applied while using CROMETA Metadata Manager. First of all, the system supports a basic user group model where users can be divided into system administrators, object type experts, standard users etc. Secondly, all metadata created may be secured by the creator assigning privileges for how it may be used by other users. By default, all metadata kept within the system is read-only for all internal users.

In practice this means that when creating for example a new study, the creator may assign rights in any way he/she likes, using the tool, in order to ensure that only some persons are able to edit information, while others are only allowed to browse it, etc.

Search

CROMETA Metadata Manager allows searching and filtering objects in various ways. Quick search based on titles is available as well as a full-scale advanced search including descriptions, keyword-based search functionality, etc.

General properties managed in a general way

When working with any metadata object, the general properties will always be available in the same way. In practice this means that footnotes, keywords, documents, etc. could be connected to all metadata objects in the same flexible way.

Subscription to metadata

In some occasions users may find it useful to get notifications when objects like for example classifications are updated. This is supported by CROMETA Metadata Manager through subscription-functionality. A user is allowed to sign-up to any object, specifying with which frequency he or she would like to be notified on changes. As soon as the metadata object, for example the NACE-classification is modified; the registered users are notified according to their independent frequencies.

Locking and unlocking of metadata

In order to ensure that the same metadata are not edited simultaneously by two different users, CROMETA Metadata Manager has built in functionality for exclusively locking and unlocking of metadata.

4.3 Standards and formats

The CROMETA model contains Reference ModelTM concepts extended and customized for CBS needs. As CBS had already implemented several other software packages containing metadata, e.g. Crometa#1, PC-Axis, BRIDGE^{NA}, those other metadata had to be mapped to the Reference ModelTM. Thorough analysis provided specifications how to integrate specifics of Crometa#1 and PC-Axis SQL Macrometa in the new model. The result was CROMETA - a comprehensive CBS metadata solution consisting of metadata objects and their relations. Naturally, the original model has been extended and revised in last two years, in order to better support the needs of statistical processing. The model is constantly evolving as new needs are discovered among the users of the metadata server, and as areas of statistical processing are further elaborated. The CROMETA model is considered to be 'implementation independent', i.e. it is a conceptual description of metadata as well as its internal relations, and hence it could be implemented on any technical platform. The complete model is available as UML and other formats as part of the CROMETA Metadata Server;

The SDMX standards will be implemented as soon as possible; presently it is still under consideration.

4.4 Version control and revisions

Each metadata object can have several versions that are valid in exclusive periods of time. There is only one version valid in any moment. This is determined by the Status of the metadata object as follows:

1. Under development

When a metadata object is created it gets this status by default. (This could of course be changed by the person creating it). This is the only status that allows the user to completely delete the metadata object from the system. A metadata object having this status could be viewed, edited and deleted physically.

2. Released

When a metadata object has been completely created, it is normally being released. This means that it is finalized but not yet authorized, i.e. it has not been pointed out as the current version. This could also be described as ready for review. A metadata object having this status could be viewed, edited and deleted (not shown).

3. Authorized

After being released a metadata object is usually reviewed by a reference group, or the creator alone. When authorizing it, it is automatically displayed as the current version. The metadata object previously displayed as current version is at this point automatically moved into status 4, *Archived*. A metadata object having this status could be viewed, edited and deleted (not shown).

4. Archived

When a metadata object is being replaced as current version by a successor, the state of it is automatically changed to *Archived*. This shows that the metadata object has been used as current version, but has now been succeeded by another version. A metadata object having this status could be viewed, edited and deleted (not shown).

5. Frozen

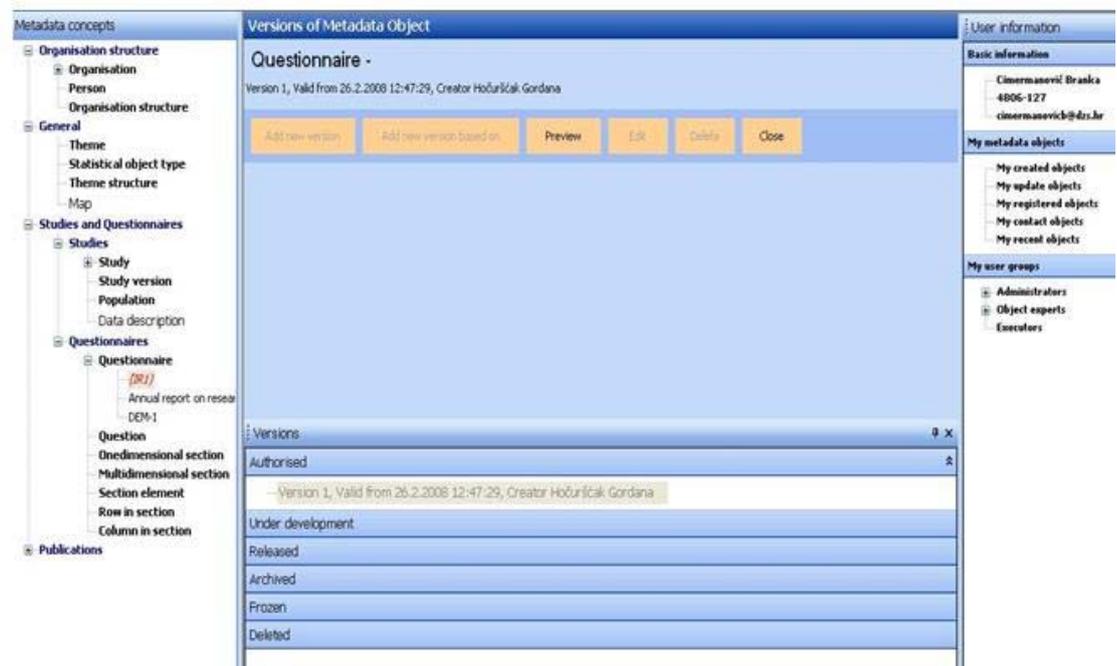
Metadata object that has been used as the current version is automatically given the status *Archived*. However, this object could still be edited. At some occasions you may not want to allow editing of old versions, and if so, the archived metadata object is given the status *Frozen*.

6. Deleted

Only metadata objects of the first status, *Under development*, could be physically deleted. However, metadata object of statuses two-five could be put into this status in order not to be shown in the metadata system. Metadata object of this status thereby appear to be deleted but

could in fact be restored by the administrator.

Whenever some significant change on metadata must be applied, it is recommended to create a new version. If changes are small, e.g. spelling corrections, they can be applied without defining a new version. Such changes are kept in Update information, a general property of all metadata concepts in CROMETA.



Full access to all versions of metadata data

4.5 Outsourcing versus in-house development

The detailed plan for CROMETA development was laid out in May 2004. Before that, in 2002 and 2003 plenty activities were performed such as on-site training courses, seminars, short-term visits, study visits, procurement, Crometa#1 development, etc. Participants in those activities were mainly IT people and some statisticians. It should be noted that CBS's IT people are more than well informed about statistical activities, methodologies and processes since they process or manage more than 100 statistical surveys on various platforms on a daily basis. The fact that the IT sector contained both knowledge about statistical methodologies and IT development made it natural that the central metadata repository should be developed in-house. That saved a lot of money since it was performed during regular working hours. On the other hand, the development time was significantly longer because there is no development team that could be appointed to this job exclusively.

There is also the crucial problem of 'brain drain' throughout CBS and particularly the IT sector. Some twenty people left IT sector in last three years, most of them were involved in CROMETA and ISIS development. This influenced a great deal the activities laid out in the plan since there was a constant shift of duties and responsibilities along with the constantly growing amount of regular production work.

Outsourcing would take a year or so for the software development itself, but it would take another year of work with statisticians to sort out, explain and document all the requirements. Given the complexity of CROMETA concepts it seems unlikely if not impossible to outsource the development and receive a comprehensive solution to be used and maintained by statisticians.

5. ORGANIZATIONAL AND WORKPLACE CULTURE ISSUES

5.1 Overview of roles and responsibilities

In CBS there are 25 subject-matter departments with *statisticians* responsible for respective statistical surveys. They manage statistical surveys according to the periodicity and they are responsible for the timelines and quality of statistics. Data of almost all statistical surveys are handled and processed in the IT sector where appropriate software is applied by *application developers / information system developers / programmers*. The SW is developed upon the written requirement issued by the statistician in charge. The requirement document is a template based on the needs of IT experts to process any survey.

By implementing CROMETA we expect some changes of roles, namely more work should be completed on locations where and when the need originates, instead of following the much more time-consuming procedure of central data processing.

5.2 Metadata management team

The CROMETA development project has been divided in two parts, as defined in July 2004:

- the *Metadata Methodology* sub-project with the task to form and define the conceptual metadata methodology for CBS, based on mapping of all existing metadata models and definition of processes needed to reach a comprehensive metadata solution. The result of this project should be the basis for technical implementation.
- the *Technical Implementation* sub-project with the goal to develop the solution to physical implementation of the conceptual model of CBS' metadata on MS SQL Server 2000 (later upgraded to the MS SQL Server 2005) and metadata management tool. The result of this project should be the implementation of the 'empty' metadata container and the user interface.

Each sub-project had its own development team consisting of 7 - 8 people; between them one project leader, 1 - 2 full time developers, others as supporting members with 30-75% engagements in the respective project. Unfortunately, the project suffered a significant loss of resources, nevertheless the interesting and challenging work. By the year 2007 the metadata repository was practically completed by three developers.

5.3 Training and knowledge management

There are several lines of education necessary for CROMETA implementation. First, the IT developers should be trained and experienced in database design, Microsoft SQL Server DBMS, Microsoft Visual Basic.NET development tools and they should also have a rather thorough knowledge of statistical activities. Newly employed developers usually attend Microsoft MOC seminars.

Second, the users within CBS should be trained in using the metadata maintenance tool to enter and maintain their metadata, to create a new survey upon another one, to handle versions and history etc. Apart of detailed user manuals, workshops and hands-on trainings are planned for statisticians. The training will be prepared and performed by CROMETA developers.

5.4 Partnerships and cooperation

The cooperation with Swedish International Development Cooperation Agency (Sida) and Statistics Sweden (SCB) was established in CBS in 2002 upon an agreement on technical help signed by Croatia and Sweden. The first project was named *Support for the creation of public macro databases and a central metadata repository at the Central Bureau of Statistics of Croatia*. The project included consultations carried on through numerous missions, study visits, IT education, procurement of hardware, software and literature etc. The progress was monitored by Sida assessment missions twice a year and an independent expert mission once a year. All assessments proclaimed the project as very successful and the work done was considered as a model for other countries in the region.

In 2006, the final stage of the project was named *Development of an Integrated*

Statistical Information System (ISIS) at the Central Bureau of Statistics of Croatia. It was divided in two parts, one performed by Sida and SCB and the other by a private company Conscious™ from Stockholm, Sweden - procured directly by Sida. The SCB part ended in April 2007 and the Conscious part in September 2007.

6. LESSONS LEARNED

The most important questions are still unanswered since the central metadata management system is not deployed yet. The complete ISIS information system in general and CROMETA system in particular will force big changes upon the overall culture of CBS. The degree of content or discontent by the majority of statisticians when they start using the metadata maintenance tool is yet to be learned. In any case, we expect resistance from a number of subject-matter experts, especially those who cherish very much the legacy from past times when it was usual practice to order a tailor-made data processing system from the IT department.

Therefore the 'Lessons learned' here apply mainly to software development activities and this is by far less important than the overall cultural changes that will be met by the deployment of the new ISIS.

- 6.1** The most important lesson learned is that there is no serious development when there is no development team appointed to this and only this project. This applies to IT developers as well as statisticians. Of course we knew that even before we started the project but we could not afford to have experts unavailable to regular production for a longer period of time. So we entered a vicious circle: we wanted to develop software to make production easier but we could not develop because we had to handle the production. For this reason the development lasted much longer than planned.
- 6.2** The support from top management is crucial.
- 6.3** Teamwork is very important and now it is enhanced with practical tools such as SharePoint etc.
- 6.4** It is obvious from experiences of other NSOs that the involvement of statisticians is crucial for the project. Therefore we tried from the beginning to include selected statisticians in the development through various forms of cooperation but somehow it always ended after two or three meetings (see item1.).
- 6.5** Strict project management. It is the responsibility of the project management that statisticians fell out of development activities sooner or later.
- 6.6** The most painful lesson learned was that there is no project interesting and challenging enough to keep young and well educated IT experts from going to better paid jobs. IT experts in government bodies are paid two or three times less than in private sector and this needs no further comment. The CROMETA project started with 16 people (more or less involved) and ended with 3.

*** END ***