WP.5
ENGLISH ONLY

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

UNECE Workshop on the Common Metadata Framework
(Vienna, Austria, 4-6 July 2007)

Topic 1: Models of the Statistical Cycle

IMS (INTEGRATED METADATA SYSTEM) – AN ARCHITECTURE FOR AN EXPANDABLE METADATA REPOSITORY TO SUPPORT THE STATISTICAL LIFE CYCLE

Statistics Austria[1]

I. INTRODUCTION

1. In international discussions it is generally acknowledged that metadata play a decisive role both in satisfying statistics users' growing quality requirements and in increasing the efficiency of the internal production processes within an NSI (national statistical institute).

2. In recent years relevant technical publications have repeatedly stressed that the implementation of metadata systems must be founded on a comprehensive and general model of statistics production and on a long-term master plan (the term „metadata strategy" is often used in this context). Paying too little attention to these preconditions leads to metadata systems which are neither linked with each other nor with the data they document and which lack the ability to cooperate with each other. Often, the same information is stored repeatedly, rendering it difficult to keep the metadata consistent and causing unnecessary effort and costs. In the worst case, the resulting applications rely on mutually incompatible concepts and models, making integrating them *ex post* an extremely demanding if not impossible task.

3. The problems mentioned above are not unknown in Statistics Austria either. As presumably in every other NSI, numerous metadata-related cross-domain projects have been implemented over time. These projects were often carried out without relation to each other, thus creating more or less isolated solutions.

4. Even when they are isolated solutions, however, explicit metadata systems do have the advantage that the information they contain is administrated centrally. This distinguishes them from metadata which are generated during the planning, implementation or execution of a statistical project[2] within the separate

[1] Prepared by Guenther Zettl, guenther.zettl@statistik.gv.at
[2] The term "statistical project" is here taken to denote all types of statistical production systems – primary and secondary surveys, registers and analytical projects.

(stovepipe) production systems and are stored only within the latter. Such metadata often occur redundantly and in non-structured form in project documents or are distributed over various print publications. It is therefore difficult for internal and external users to locate them; in the worst case, they cannot be accessed at all.

5. Concurrently with the separation of the Austria Central Statistical Office from the federal civil service and the creation of Statistics Austria as an independent, nonprofit federal institution under public law, the Federal Statistics Act 2000 installed the so-called "Statistikrat" ("Statistical Council"). This functions as the highest-level body for quality assurance of federal statistics. One of the tasks of the Council's 15 members is to elaborate comments and recommendations pertaining to the work programmes which the general directorate presents.

6. In several of its comments the Statistical Council has explicitly drawn attention to the importance of delivering comprehensive metadata and of increasing the statistical system's coherence, and has demanded the development of a metadata repository. In this context it has also underlined the central role which the IT department should fulfil in "implementing the requirements repeatedly voiced by the Statistical Council for uniform information delivery, increased quality, enhanced timeliness, easier data access and provision of more comprehensive metadata"[3].

7. In no small part due to the Statistical Council's emphasis on these matters, an IT project was commenced in 2006, the goal of which was to conceive an **integrated** metadata repository and to prepare an overall plan for implementing such an information system in the form of sub-projects. The system architecture elaborated in this project goes under the working title of IMS („Integrated Metadata System") and will be presented in abbreviated form in this paper.

II. METADATA DIMENSIONS

8. It is not a new discovery that the subject of statistical metadata is an extremely complex one. Even now, almost three and a half decades after Bo Sundgren first used the term, different individuals may still mean quite different things or place emphasis on different aspects when speaking of metadata. This phenomenon is even more pronounced when these persons stem from different areas of expertise: subject matter statisticians, methods specialists, IT experts etc.

9. Statistical metadata can thus be investigated from varying points of view. The IMS project team differentiated between the following dimensions:

[3] Quote from the position paper of the Statistical Council pertaining to the work programme 2007 (p. 12)

A. Dimension "Function":

10. This dimension describes metadata's purpose. Basically, metadata are required for the following reasons:

- for searching and finding statistical information
- in order to interpret statistical data
- in order to access data
- for processing data and producing statistics
- for managing statistical projects

B. Dimension "Statistical Life cycle":

11. Statistics production can be described as a process which transforms input data into output data via several steps and using statistical methods. In Statistics Austria, this statistical life cycle is structured in the form of "statistical projects" of different types (surveys, registers, analytical projects or systems). This is described in more detail in Chapter III.

C. Dimension "User":

12. Statistical metadata are no end in themselves, but are required by different groups of users for varying purposes. "User" must here be understood in a broad sense, comprising not only persons but also IT systems.

13. We can distinguish roughly between two main groups of users: external ones (which do not belong to the NSI) and internal users. External users are mostly "consumers" of statistics, they may however also be providers of raw data (respondents). Internal users are often both "producers" as well as consumers. Among others, external users may be politicians, scientists, economic enterprises, journalists, private persons or international organisations. In general, the users in these groups differ in the amount of their previous knowledge, the level of detail they wish for in the statistical information they are seeking, and the resources at their command. From the point of view of the amount of metadata they require, one must keep in mind that this may also vary within the relatively heterogeneous groups. Furthermore, the requirements may evolve with time.

D. Dimension "active / passive":

14. This dimension treats the degree to which metadata play an active role in statistics production, i.e. controlling the process or automatising processing steps (e.g. when an electronic questionnaire is generated automatically based on the specification of a survey's questions). With regard to efficient production of

statistics one should aim at letting as many active metadata elements as possible be defined directly by the statistical subject matter experts.

E. Dimension "formatted / unformatted":

15. A distinction can be drawn between formatted and unformatted data. The structure of the former is agreed beforehand (e.g., every record in a file consists of the same sequence of data fields, which in their turn exhibit prearranged characteristics such as data type, length, etc.; or a data file conforms to a predefined XML schema) and thus easily lends itself to automated processing with computers. Unformatted data on the other hand – texts, graphics, voice etc. – are much more difficult and cost more effort to process, especially with regard to IT programs "understanding" their contents. Statistical metadata often occur in unformatted form, e.g. as text in documents.

F. Dimension "manual / automatic":

16. The criterion by which this dimension classifies metadata is whether they are recorded manually by the persons entrusted with planning and implementing statistical projects, or whether they are created automatically by tools.

17. The development of an integrated metainformation system which incorporates all these aspects of metadata is certainly a demanding task for any NSI. With regard to the limited budgetary and personnel resources and in order to make quick progress in the project, Statistics Austria had to reduce the scope of the conceptual tasks in the IMS project. The goals and consequently the basic focus of the project were thus specified as follows:

- The goal of the system to be developed is to deliver to Statistics Austria's customers (various external users, national and international organisations) that functionality which they require in order to satisfy their needs with regard to statistical information (e.g., to understand statistical results and to have the means to judge their quality). One can start from the assumption that the functional range implemented internationally as "best practise" in various statistical offices will cover the customers' requirements.

- Not only external users should profit from the metainformation system. Internal users of statistics also require the metainformation relevant to statistical products and processes (e.g., in order to be able to efficiently reuse statistics produced by the Office or to process them further for specific projects). It can be assumed that a metadata repository will also generate internal benefits with respect to efficiency and quality of the statistical production process.

18. With the above as the fundamental goals, the main focus of the IMS project was placed on passive metadata, which are required both by external and internal statistics users, in particular for the functions "finding" and "interpreting" statistical data. With an eye to this, the metadata repository was conceived as a comprehensive **documentation system** for statistical data and production processes. The concept provides for collecting the metadata which are hitherto scattered over various production systems and (working) documents, storing them in structured form according to a general model of statistics production, and integrating them by allowing links to be created between the individual elements of documentation and the data they describe.

19. The IMS team was aware of the fact that restricting the focus to the area of statistics documentation would in the short or long term mandate further stages of expansion in order to remain in step with international developments (keywords being horizontal and vertical integration, standardisation, metadata-driven processes). Special attention was therefore paid to laying the foundations for a platform which would remain usable in the long term and be capable of further extension. The demands as to quality and efficiency of modern statistics production, which have risen sharply and can be expected to rise even further, should thus be satisfiable through successive functional expansions – then also including active metadata.

III. STATISTICS AUSTRIA'S MODEL OF THE STATISTICAL LIFE CYCLE

20. Within the framework of a strategy project, a model of the statistical life cycle was defined in Statistics Austria in 2002 which we will describe shortly. The model distinguishes the following types of statistical projects: surveys, registers and analytical systems.

21. **Surveys** are the most "typical" and most commonly occurring form of a statistical project in Statistics Austria. One can differentiate between primary surveys (in which the collection of raw data is one of the steps of the process) and secondary surveys (which process data which already exist and were often collected for non-statistical purposes). There are also mixed types, e.g. surveys in which data from secondary sources are used to augment the data collected by questionnaire. Some surveys are only undertaken once, others are repeated at regular or irregular intervals – although the sets of variables collected in each repetition do not have to be identical. It is therefore useful to further subdivide the survey structure: each survey consists of one or more survey versions, and each survey version consists of one ore more survey instances, i.e. concrete executions. E.g., in the case of a survey with monthly periodicity each year might be seen as a new survey version with twelve survey instances.

22. In contrast to the data of a survey instance, which pertain to a certain reference date or a certain period, **registers** are usually updated continuously. Maintaining a register is thus a core process distinguishing a

statistical project of type „register" from projects of type „survey". Another fundamental difference is that register data are used as resources for workflows in other statistical projects, e.g. when drawing samples, for addressing, registration of incoming questionnaires and administration of reminders. Commonly, specific (database) applications are developed to carry out these functions.

23. **Analytical projects and systems** (as, for instance, national accounts) characteristically do not collect raw data on specific observation units, but use data from other statistical projects and evaluate them or combine them into a coherent, integrated model.

24. Data which form the input to a statistical project (or which are collected in an early phase of processing, in the case of primary surveys) often pertain to individual observation units, e.g. individual persons, households, enterprises, events, etc., and are termed "microdata". However, the input may also consist of macrodata, i.e. data pertaining to collectives. In addition to these, metadata also enter into a statistical project and form an important resource for the steps carried out in its processing.

25. The output of a statistical project consists predominantly of macrodata – cross-classified tables, multidimensional data cubes and time series being the most important categories – and metadata. More rarely, microdata may also be produced, e.g. anonymised sets of microdata. Macrodata and certain related metadata are often combined into an „information object" (e.g., a press release consisting of a table and descriptive text). Such information objects may also be composites of smaller information objects (as with a printed publication containing several tables and descriptive metadata – e.g., analytical texts).

26. The Statistics Austria model of the statistical life cycle distinguishes the following phases in the production of statistical information:[4]

Phase 1: Planning, design and system development

27. The survey is set up in this phase. Given specific requirements (e.g. EU regulations) and the information needs of internal and external parties (e.g., statistics users in government and the economy), the survey must be prepared to satisfy these as best possible while simultaneously minimizing the effort of the statistics producers and the burden on the data providers.

28. Output of this phase are metadata of various kinds – e.g., description of the survey's goals; description of the characteristics to be surveyed; definitions; value domains; classifications; questionnaires and comments

[4] This description applies to statistical projects of the type "survey". Registers exhibit different core processes (creation, maintenance, and use of the register), although the contents of a register may also form the basis for production of statistics and information (which can be identified with the relevant phases of a survey). The line between surveys and analytical projects is not always clearly defined – in fact one could certainly argue that analytical projects are a special type of statistical survey creating statistical information from input data, the difference being mostly that methods are applied which differ from those used in "typical" surveys.

explaining them; list of validation rules, etc. The metadata created in a survey may form the input to other phases or other surveys and be reused there.

29. The development of tools for actually conducting of the survey (e.g., electronic questionnaires, editing software, programs for checking consistency and plausibility) is also a component of the first phase.

Phase 2: Data production

30. Whereas the decisions taken in the design phase and the metadata and tools which are created there apply to the whole survey or at least a survey version, the focus of the activities undertaken in the following phases lies mostly on the current survey instance (excepting activities in which data from more than one survey instance are processed, as in the creation of time series).

31. Data production can be subdivided into three subphases:

- In **pre-production** activities such as drawing the sample, printing and posting paper questionnaires, loading Web questionnaires with respondent-specific initial data, etc. are undertaken.

- The actual survey/measurement/observation of the statistical raw data is termed **core production**. This includes conducting interviews, filling in paper or electronic questionnaires, registering and roughly checking questionnaires which arrive, mailing reminders, data entry from paper questionnaires, etc. In secondary surveys, this subphase includes acquisition of the secondary data and, if necessary, reformatting or recoding it. The collected data are stored in the so-called Raw Data System (RDS).

- Post-production includes all activities necessary to correct the raw data which has been collected. Among these are validation and consistency checks, examination and correction of dubious information, and imputation of missing values. The results of this phase are the “authentic” data (ADS: Authentic Data System); of these several versions may exist, especially in complex and voluminous surveys (e.g., preliminary version at a certain cut-off date and final version at a later date).

32. A large part of the metadata created during the design phase enters the second phase as input. As also in later phases, however, new metadata also are produced (e.g., the answer rate, which is an attribute of the survey instance).

Phase 3: Statistics production

33. In this phase, the contents of the Authentic Data System (consisting mainly of microdata) are processed further. To do this, data from other surveys may occasionally be accessed. Some of the processing steps

undertaken are aggregation into macrodata, calculation of statistical measures and indices, diverse methods for increasing quality and comparability of statistical information (e.g., seasonal adjustments), and creation of time series. The results are data sets which are stored in the Statistical Data System (SDS) and are at the disposal of internal, often also external users. The SDS primarily contains multi-dimensional data cubes, although anonymised sets of microdata may occasionally also be created in this phase.

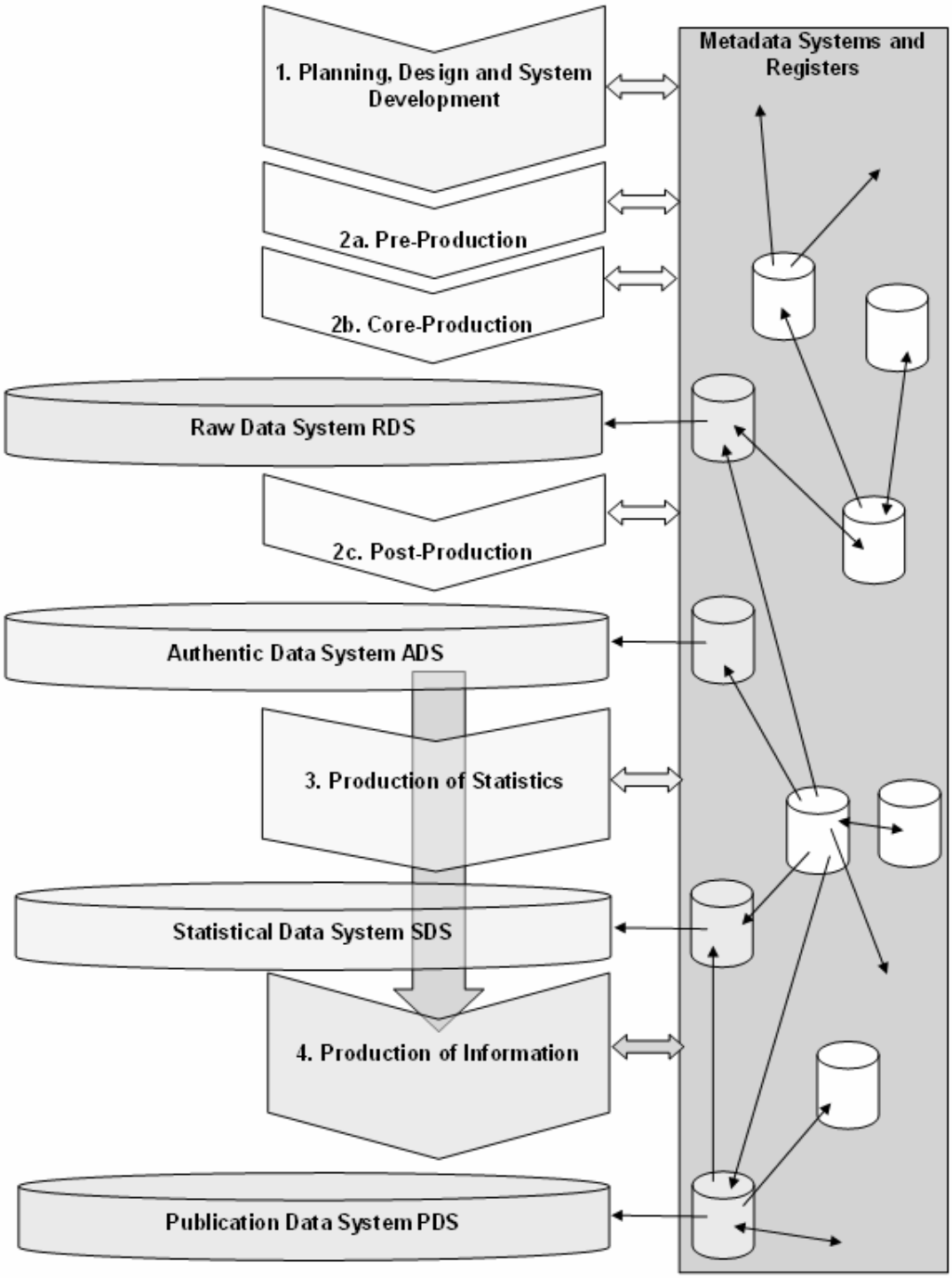
34. In part, the transformations which are carried out here have already been planned in the design phase and are applied to the data of each survey instance. Partly, however, ad hoc analyses may also take place, which use the existing data material in ways not foreseen when the survey was planned. This underlines the importance of comprehensive and easily accessible documentation of all a survey's design decisions, of the data sets and of the transformation processes (in whatever phase of the statistical life cycle they may be created or executed).

Phase 4: Information production

35. In this last phase, „information objects" such as tables, charts, articles, press releases etc. are created from the data stored in the ADS and the SDS and their metadata and disseminated via various media (internet, print publications, etc.).

36. The following figure presents the phases described above, the data systems, the registers and the meta-information systems. On the one hand, the latter provide input and various services for the activities carried out during the statistical life cycle, on the other they also accept the metadata created as output from each phase. The data systems are drawn as broader than the "process arrows" in order to point out that they contain data from various surveys and that an individual phase of a survey may accept input data from more than one statistical project.

37. In actual fact the workflows are of course not quite as linear as the figure might suggest; on the contrary, complex control flows (branches, loops) often occur. Moreover, events in later phases may have retroactive effects on the survey's design and lead to adjustment of the current or future survey instances (e.g., changes to the validation rules).



38. Individual sectors of this life cycle are already covered by cross-domain applications or metadata systems in Statistics Austria. Some examples are:

- **ISIS:**

39. ISIS (short for Integrated Statistical Information System) is a statistical database which was already developed in the early 1970's and has been consistently maintained and developed further since then. It

contains thousands of multi-dimensional data cubes as well as metadata of various kinds (e.g., short descriptions of the data cubes and the underlying surveys; keywords and a hierarchically structured topic tree are furnished for data searching) and implements a large part of the Statistical Data System SDS in the life cycle model. Although ISIS is still very modern from the point of view of the conceptual design of its contents, the software itself has reached the end of its life span, as only one programmer now still possesses sufficient technical know-how to maintain the mainframe Assembler and PL/I programs. Because of this, a successor system is currently being developed on the basis of the Australian company Space-Time Research's SuperSTAR product range.

- **e-Quest:**

40. e-Quest is a system consisting of several tools for metadata-driven generation of electronic questionnaires, administering them and preliminary processing of the incoming questionnaires. The subject matter experts can design the questionnaires with a user-friendly graphical editor. The active metadata thus specified are stored in XML format and then used to represent the questionnaires dynamically in a Visual Basic rich client application (which must be installed by the respondents) on the one hand; on the other hand, however, they are utilized for generating Java and Javascript source code, JSP pages and SQL table definitions for electronic questionnaires accessible via a uniform Web questionnaire portal. e-Quest thus covers important areas of phase "data production".

- **Publication object database:**

41. Using the document management software of the company Stellent (which has recently been acquired by Oracle) the publication data system PDS was created in the last few years. This stores all information objects (i.e., documents of various types, from tables over print publications and press releases to the so-called standard documentations) together with metadata relating to the documents[5]. Since the Web relaunch on June 1st 2007, Stellent is also being utilized as a Web content management system. The subject matter experts now create Web pages in the form of standardized Word documents which are automatically converted to HTML and copied to the correct position in Statistics Austria's website on the basis of associated metadata (in particular a hierarchical topic and navigation structure). The navigation structure is also used for generating links to related documents with data and metadata. The new catalogue of print publications (many of which can be downloaded free of charge as PDF files) was also implemented in the Stellent system.

[5] Compared to the publication object database the concept of „information objects" was broadened in the IMS. In the latter the term does not only denote documents of various types, but quite generally objects which carry certain types of statistical data and/or metadata and which may be related to each other along predefined paths. The terms "survey", "survey version", "survey instance", "type of statistical unit", "characteristic", "ISIS data cube", "code list", "table", "print publication", "document", "press release", "definition of concept", "authentic data set", "validation error" are just a few examples of object types (classes) whose concrete instances are termed information objects.

- **Classification database:**

42. A few months ago the classification database KDB was released. This allows Web access to (hitherto) 16 voluminous classifications such as PRODCOM, NACE, COICOP, SITC and CPA, including comments and correspondences. More than one version is available for several classifications.

- **Statistical Table Format STF:**

43. STF is an XML specification which permits cross-classified tables to be stored together with extensive metadata in a hardware- and application-independent format – for long-term storage, among other uses. Converters from STF to Excel and HTML and from Excel tables to STF are supplied. When Excel tables are checked into the Stellent publication object database, they are automatically converted to STF format. ISIS query results can also be stored in STF format.

- **Standard documentations:**

46. The standard documentations – which can be downloaded as PDF documents over the Web – serve as the most important source of metadata about statistical projects and the quality of the statistical results they produce. The documents exhibit a standardised chapter structure and hitherto describe almost 100 statistical projects or survey versions, in part in great detail (they number between 9 and 68 pages; in many cases further documents are provided as attachments which can be accessed via hyperlinks in the text). The standard documentations prove that extensive IT applications are not always required to support metadata systems which may be very rich in informational content. However, among other things they do carry the disadvantage of usually being written and made available to the statistics' users in a separate and additional work step after the fact, although they contain many documentation elements which come into existence in the early phases of planning and preparing the statistical project.

IV. IMS ARCHITECTURE

47. Although the (meta)data systems listed in the previous chapter already cover several essential sectors of the statistical life cycle, other sectors still show large gaps. In comparison to international best practice, a series of weak points were identified in the IMS project:

- The pronounced stovepipe architecture of statistics production in Statistics Austria leads to major differences between the workflows and tools utilized in different statistical projects. Thus, for example, RDS and ADS have no standardised data storage: the formats range from sequential files with varying record structures on the mainframe over relational databases to data sets in varying PC formats. Apart

from the difficulty that information about the various raw and authentic data sets and their structure can often be discovered only by asking the responsible subject matter experts and/or programmers, the lack of standardisation in this area also requires important processing tools such as editors and validation programs to be tailored to each stovepipe separately, making development and maintenance inefficient.

- The principle often underlined in relevant technical literature, that metadata should be captured in a standardised form at the moment they come into existence, and thereafter should be reused, is only inadequately put into practice at present. On the contrary, metadata are often stored redundantly in work documents or appear in the continuous text of bulky documents. In this form, it is difficult or impossible for internal and external users to access them, and IT access to individual documentation elements or attributes of metadata objects is normally also impossible. This not only causes opportunity costs and additional effort, but also damages the coherence of the statistical system.

- Most of the existing metadata systems are isolated applications. Additionally, essential metadata systems such as databases for definitions, datasets, variables, or value domains are missing, although many other NSIs have had these for quite some time.

- A further weak point identified by the IMS project team concerns searching for data and metadata. Since the recent Web relaunch we now have a full text search facility, but this can only locate documents in Stellent (including Web pages), but not, for example, data in ISIS or classifications in the classification database. Searching may therefore be cumbersome and time-consuming for the user in some cases, as it may entail multiple search requests to multiple systems (which may not even be known to the layperson) with differing search mechanisms and user interfaces.

- As regards other types of searching, a list of keywords (index) and a thematic search based on a hierarchical topic tree now exist since the Web relaunch, but again these can only locate information objects in Stellent. Other information systems such as ISIS possess a differing thematic structure. Searching for data based on a list of statistical projects or on types of statistical units and their characteristics is not supported.

- With respect to linkages between data and metadata (which would allow users of statistics to navigate quickly and easily within a „semantic net" between objects of various types) the Web relaunch and the use of Stellent as content management system have brought considerable progress – for example, starting from a Web page, links to print publications, standard documentations or press releases are displayed. However, these links are created indirectly by associating topics from the topic/navigation tree defined in Stellent to the individual documents. Specific information objects cannot be connected explicitly via various types of relationships (e.g., the relationship „is published in" between a table in

Excel format and a print publication). Links to information objects which are not checked into Stellent as documents are also impossible.

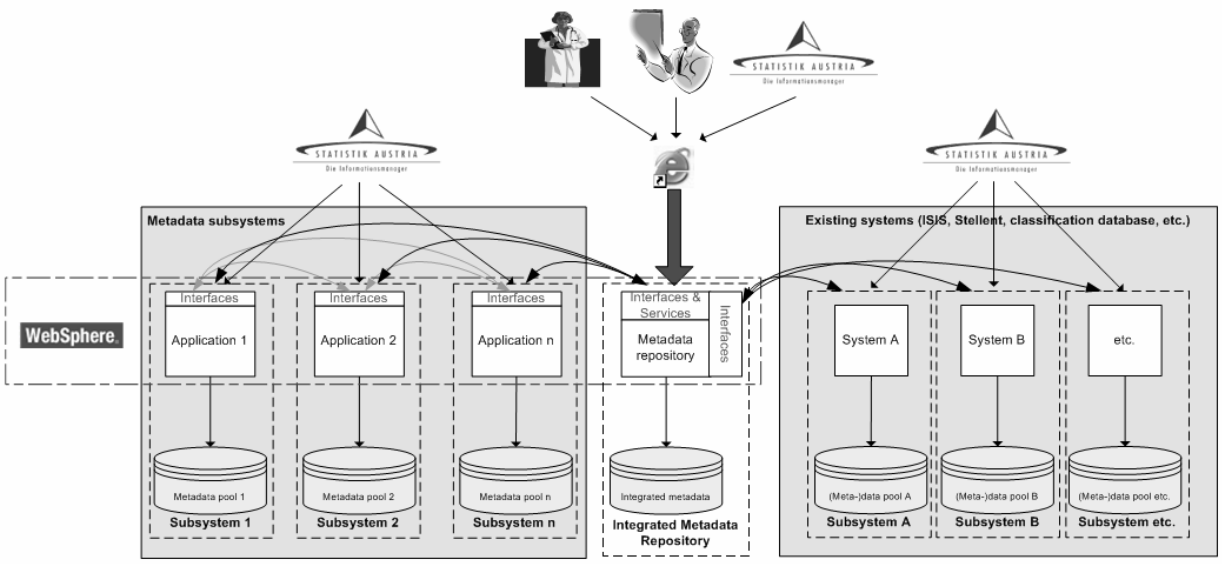
46. In the conceptual design of the IMS the attempt was made to construct an architecture for an integrated metadata system by which the weak points mentioned could be improved or eliminated over time, in the sense of evolutionary development of the existing applications and information systems.

47. Especial emphasis was laid on facilitating gradual commissioning and expansion of the IMS. To this end, the overall system was partitioned into subsystems which can to a large extent be implemented and used independently of each other, but which work together via precisely defined interfaces in the form of web services and via a central component, the so-called registry. It must also be possible to integrate the previously existing (legacy) systems into the overall system, with the linkage more or less strong (loose or tight coupling).

48. Work on the IMS proceeded in two phases and in cooperation with external partners, among them – to ensure scientific monitoring and supervision – Professor Karl Froeschl (who is known for his longtime engagement in the field of metadata and for collaboration with numerous pertinent projects such as METANET, IDARESA etc.). In phase 0 (analysis/design) a rough architecture was defined. In order to gain an overview over the various tasks performed by the specialist departments, the availability of metadata, IT systems used etc., workshops were carried out with the persons responsible for four widely varying statistical projects. In these workshops the current statistical production workflows were described, the availability of metadata was analysed, these metadata were categorised and the potential for improvements which might lead to better satisfaction of internal and external users' requirements was discussed. A further focus of the project efforts was laid on the analysis of existing products and standards (in particular ISO 11179 and DDI) and on research as to current developments in various best-practice NSIs (especially Statistics Canada).

49. In phase 1 (detailed analysis) the concepts were subjected to extensive further detailed analysis. Amongst other things, UML class models were constructed for the individual subsystems, solutions were drawn up for overall aspects such as access authorization and quality control (status administration), the requirements were specified for Web and rich client applications to be developed in the future, and important functionality was illustrated in the form of use cases. In addition, an analysis was carried out as to in what degree metadata from ISIS and Stellent could be transferred automatically as the initial filling of future IMS subsystems, and a prototype for a rich client application for the input and search for definitions of concepts was developed.

50. The following figure presents a schematic overview over the architecture of the integrated metadata system IMS.



51. The right side of the picture symbolises the cross-domain applications which already exist or are currently being developed (ISIS, classification database, publication object database, etc.). The left part shows new subsystems which must be developed. These do not replace any existing productive systems, but are responsible for central and structured entry and consolidation of metadata which at present are scattered over various sources.

52. In detail, these subsystems are as follows:

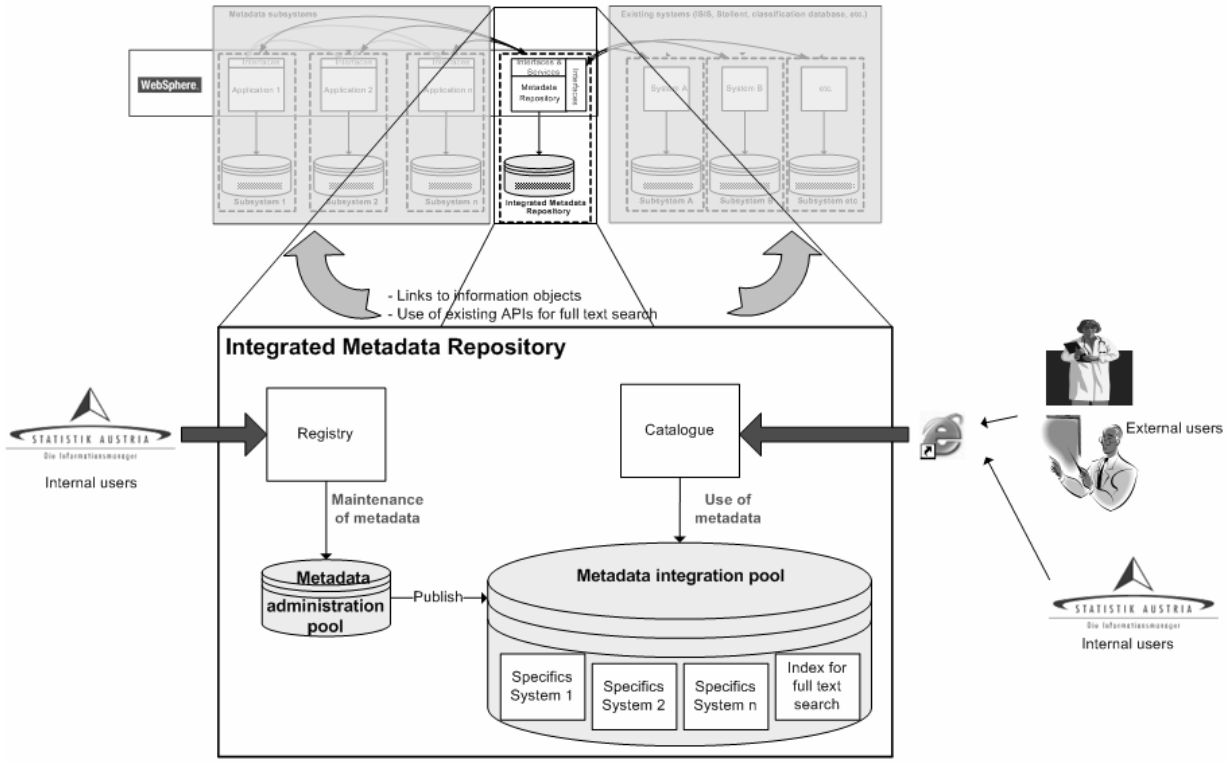
    1. Definitions and concepts
    2. Statistical projects
    3. Types of statistical units and their characteristics (variables)
    4. Value domains

53. The latter two subsystems are based on ISO 11179, although during the modelling process some areas were simplified and others enhanced as compared to this standard.

54. The subsystems which are to be developed will communicate with each other (and also with the Integrated Metadata Repository IMR, particularly with its component “Registry” – see below) via web services. In this way the mutual interdependencies between the individual subsystems can be minimized, with an eye to the concept of encapsulation.

55. The Integrated Metadata Repository (IMR) is positioned centrally between all these systems. This consists of two parts, the “Registry” and the “Catalogue”. The latter replicates all those metadata which should be accessible for external users over the Web. These will normally consist of a subset of the metadata administered in the IMS subsystems, as the latter will also contain information which is only of interest to

the subject matter persons responsible for statistics production. The catalogue will also allow a comprehensive full text search over all subsystems, i.e. with a single search request the user should be able to locate not only the documents and Web pages stored in Stellent, but also data and metadata in ISIS, the classification database and the future IMS subsystems.



56. The second component of the IMR, the “Registry”, is responsible for a Statistics Austria-wide unique registration of all the information objects contained in the individual subsystems and in the connected legacy applications. In addition, it administers links between information objects, which will be of various types (e.g., “contains data from” between a document or an ISIS data cube and one or more surveys). These two core functions are prerequisites for later allowing the users to navigate from one data or metadata object to another along predefined paths. E.g., starting from a list of types of statistical units such as enterprise, household, birth, etc., one might quickly locate the characteristics which were collected or created during statistics production, “surf” from there to the corresponding value domains or to the surveys / survey versions, to definitions etc.

57. Further tasks of the Registry are to provide central services required for more than one subsystem (such as administration of users and access rights, status of the registered information objects) and the publication of metadata to the Catalogue.

V. NEXT STEPS

58. To elaborate and implement a strategy for stepwise development of an *integrated* metadata system founded on a stable and extendable overall architecture is a major challenge for statistical institutes. With the conception of the IMS, Statistics Austria has undertaken an essential and promising step.

59. Statistical metadata are however an interdisciplinary field which affect not only IT, but also and in even greater measure the specialist statisticians. International experiences have repeatedly shown how important it is for metadata-related projects that the specialist departments are involved at an early phase of development and that the commitment of top management is ensured.

60. Hitherto, the IMS project has been very IT-centred. The next steps must now be to subject the concepts which have been worked out to a broad discussion within the whole of Statistics Austria and to conduct a cost-benefit analysis of the proposed subsystems.