

Distr.
GENERAL

CES/AC.71/2003/6
20 December 2002

ENGLISH ONLY

**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 ECE/Eurostat/OECD meeting on the management of statistical information systems
(Geneva, 17-19 February 2003)

Topic II: Impact of technical measures and standards on data quality

PUBLICATION CONCEPT (PRINTING AND WEB)

Invited paper

Submitted by Statistics Austria¹

Summary

I. STARTING POINT

1. The PC and Internet generation has a completely different access to the search, retrieval and processing of information. In the foreground is not the procurement by others (information service, written order etc.) but it has become natural to recall information (inter)actively, online and on demand. This group of clients – which is continually increasing – expects an offer from the providers of information which corresponds to their way of dealing with information.

2. To Statistics Austria, which is a modern service company and provider of information, the necessity arises to search continually for means and ways to present statistical results to a group of customers as large as possible and make them accessible in a competitive and efficient manner.

II. VISIONS AND TARGETS

3. All information objects (text documents, tables, charts etc.), which are produced in-house, are stored within a data bank. This enables internal as well as external recycling.
- All publications (tables, documents, charts, press releases, etc.) are, first of all, stored within this data bank and may be retrieved via Internet. Publications on other media follow and are exclusively based on the information available from the information object data bank;

¹ Prepared by Dieter Burget (dieter.burget@statistik.gv.at).

- For this reason, Statistics Austria's publication strategy focusses on this data bank (or publication object data bank). This way, publications (in the broader sense of the term) may be generated automatically.
4. As a big part of these objects is also presented on the homepage of the Institution, the possibility to recover them must be regarded as very important. This is feasible only by means of a sophisticated content management. A single modular solution is planned for the two problem fields of:
 - publication object data bank, and
 - content management.
 5. Basically, it must be possible to store an information object in different formats within the publication object data bank. In the medium and long run, the use of the XML format will be obligatory and the XML variant will be the first version to be stored.
 6. On every information object, sufficient meta information must be available, with the majority of metadata not being stored redundantly but directly accessible from a central metadata repository.
 7. As, presumably, free access is not intended for all information objects user administration, access control and electronic payment facilities are desirable.
 8. The publication object data bank must be implemented in a way that application will be possible as well in case of in-house documents to be accessible to Statistics Austria or its individual organizational units, only.
 9. The users must have the opportunity to register their e-mail addresses so that they are automatically notified whenever an information interesting to them is available (personalization of the offer).
 10. Other examples of new forms of publication are e-mail subscriptions automatically distributing information objects via e-mail.

A. General targets

11. We intend to apply standard software, which will be adapted to the existing functions with possible extensions, as well as to use and apply already existing functions. We do not envisage changes of the standard package.
12. Modules in addition to the standard package must be analysed and drafted and will be possible only after the agreement of the project leaders and of the project steering committee. If the application of standard software is not possible the corresponding specifications must be worked out.

B. Starting point

13. To Statistics Austria, a modern service enterprise and provider of information, it becomes necessary to inquire continuously into the possibilities of making statistical results accessible to a group of customers as large as possible in a competitive and efficient way. In accordance with the structure and the information requirements of our customers – including in-house officials – provision of information may be structured into three levels:
 - (i) The highest level is formed by publications including frequently requested data in the form of prefabricated tables, expert articles and charts;
 - (ii) The level below comprises detailed information supplied by retrievals from statistical databases;
 - (iii) On the lowest level we find specific evaluations often requiring programming activities and sometimes going into micro-data level.

14. Higher levels include a selection and aggregation of data, which contain more details in the lower levels – but not exclusively, as level 1 may also comprise information from other sources (e.g., the National Bank, WIFO, international organizations).
15. Data quantities increase with the lower levels. At the same time the costs and complexity of access rise.
16. The frequency of inquiries for information, however, declines with the lower levels. While the highest level covers the information requirements of the majority of interested persons (and, thus, according to the concept of the European Statistical Systems has millions of potential customers) the levels below are of greater interest to experts, political and economic institutions as well as to big enterprises of relevance.

C. Points of contact to other *ST.AT+* sub-systems

17. PubObjDB is in close contact to the central metadata repository, as to every information object there should be links to the corresponding metadata (e.g., keywords are attributed to a table in PubObjDB, in-depth information on every keyword may be retrieved from the metadata repository).

D. Catalogue of requirements

F1. Functional requirements

- F1.1. Facility to store data of diverse formats (at least MS Office formats, PDF, graphic, XML, ASCII delimited).
- F1.2. Define and use for retrieval user-defined metadata of these files.
- F1.3. The storage of a document in different file formats should be supported organizationally.
- F1.4. The filling of the DB by means of automated batch processing must be possible.
- F1.5. Different search methods should be supported:
 - F1.5.1. Hierarchical search (similar to ISIS)
 - F1.5.2. Search by keywords (from the metadata)
 - F1.5.3. Metadata-based search (e.g., all *.xls of a certain survey)
 - F1.5.4. Full text search
 - F1.5.5. Search by wildcard (*, ?)
 - F1.5.6. Unclear search (Maier, Mayer, Meier etc.)
 - F1.5.7. Search with Thesaurus
- F1.6. Administration of database objects (check in – check out)
- F1.7. Facility to define (simple) work flows.
- F1.8. User administration with hierarchical access rights.
- F1.9. Facility to create “personal“ databases.
- F1.10. Facility of automated information systems depending on defined events.
- F1.11. Facility of e-mail subscriptions.
- F1.12. Electronic payment of external users, if possible.
- F1.13. Justifiable reply times of search processes even with large data quantities.
- F1.14. Web surface (HTML), as external use is intended apart from internal use.
- F1.15. Support of the in-house website production, therefore at least rudiments of a CMS.
- F1.16. Basic technology for individual developments (no monolithic software programme) enabling Rapid Application Development.
- F1.17. Cost allocation/accounting facilities to (internal as well as external) users.
- F1.18. Multilingualism
- F1.19. System must be open to combinations with other project (parts) of *ST.AT+*, for example:
 - F1.19.1. Project transgressing search system
 - F1.19.2. Metadata repository

F2. Technical requirements

- F2.1. Best suited storage of text-oriented large DB objects.
- F2.2. Database component shall be up on the mainframe and not on individual servers.
- F2.3. DB should either be the strategically applied DB2 or require no peculiar DB administration (apart from F1.6 or F1.8), that is, be well encapsulated.

F3. Commercial requirements

- F3.1. Solution with methodologically sound software products, which have been frequently applied, already.
- F3.2. Supplier should have years of experience in the area of text databases.
- F3.3. Relevant reference installations (in the German speaking area).
- F3.4. Competent support as close as possible.
