STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 15
English only

CONFERENCE OF EUROPEAN STATISTICIANS

Joint UNECE/EUROSTAT Work Session on Electronic Data Reporting
(Geneva, Switzerland, 13-15 February 2002)

Topic (iii): Metadata, conceptual models and standards

e-QUEST: A METADATA-BASED SYSTEM FOR ELECTRONIC RAW DATA COLLECTION

Submitted by Statistics Austria [1]

Invited paper

ABSTRACT

It is very likely that electronic means of data reporting will become a standard in the near future. In the case of organizations with a wide range of surveys to manage, however, the effort involved in developing, disseminating and maintaining electronic questionnaires tailored to specific surveys is prohibitive. What is needed is a complete infrastructure covering all phases of the data collection process from the preparation of electronic questionnaires by subject matter specialists up to the management and initial processing of the transmitted raw data.

In 1998 Statistics Austria initiated a research project with the mandate to develop a generic solution of this kind for electronic raw data collection that can be used for any survey by specifying all survey-related meta-information – including questionnaire forms and validation checks – in extended markup language (XML) format. The core component of this system, the "electronic questionnaire management system" e-Quest, is a product that especially suits the requirements of complex statistical surveys.

The paper presents the architecture of the system (including system components that are used internally at Statistics Austria for the preparation of new electronic questionnaires and for the processing of incoming data) and describes basic design principles and how they were implemented in e-Quest. Also discussed are different types and layers of metadata in the context of electronic data reporting.

I. INTRODUCTION

1. Like many statistical institutes, Statistics Austria aims to reduce the burden placed upon respondents, in particular those of the business sector. In order to attain this goal a package of correlated instruments, principles and measures is needed. These measures form one of the basic dimensions within the Total Quality Management (TQM) concept which has been developed as a central part of the management policy of Statistics Austria. [1]

2. The new Federal Statistics Act, which came into force on 1 January 2000, has set a modern framework in which official statistics is to be elaborated and disseminated in Austria. This law gives high priority to the alleviation of the respondents' burden. The most important rule in this context is stated in article 6 which lays down that before any surveys are conducted, one must examine whether existing administrative data can be used instead.

---

[1] Prepared by Wolfgang Koller, Günther Zettl (Statistics Austria) and Frederick Rennert (CSC Austria).

3. Not all required information is available in registers or other sources, however, so primary data collection – often carried out in the form of self interviews – will still be necessary in the future. Apart from principles such as preferring sample surveys to censuses (on the basis of voluntary co-operation whenever possible) and undertaking rotation of the respondents of periodical surveys, the provision of electronic questionnaires is often considered a major step toward minimizing the burden.

4. The extent of the benefits offered by electronic questionnaires largely depends on the features that are implemented in the questionnaire software. As a rule of thumb it can be said that the respondents will profit most when the reporting task is automated as much as possible. On the other hand, if the electronic questionnaire offers almost no advantages over paper forms (or even has massive disadvantages, for example shortcomings regarding security, or very strict validation rules which make it difficult to complete the questionnaire), there will be little incentive to use it.

5. Electronic data reporting can be of benefit not only to the respondents, but to the statistical institute (or in general to any "collector of statistical information" or CSI) as well. Most likely the quality of the collected raw data will increase because the software can immediately apply certain validation checks and because a potential error source – namely typing in the answers at the CSI – is avoided. Moreover, the procedures related to data reporting and to processing of incoming data can be speeded up, leading to time-savings in the production of statistical results.

6. But there are no pros without cons. For the time being, electronic questionnaires are usually just another reporting instrument in addition to the conventional paper forms. The respondents are free to choose whichever technique is most convenient for them – and as experience shows, it often takes years before a significant percentage of them takes the plunge. For the CSI this means that two different modes of reporting have to be set up and maintained in parallel.

7. Another drawback is the costs involved in the preparation of an electronic questionnaire. Development and maintenance of questionnaire software (including tasks like eventually conducting a call for tenders, analysis, design, implementation, tests etc.) are time-consuming and expensive. Beyond that there is always the risk that the developers underestimate the requirements and complexity of statistical surveys and are not able to finish the product in time and within the given budgetary constraints.

8. Statistical surveys are not static. In many cases the information requested from the respondents changes every year, often due to influences beyond the control of the CSI. If an electronic questionnaire is tailored to a specific survey, every modification of the statistical inquiry calls for an adaptation of the program's source code by an IT specialist.

9. Another important issue concerns the efforts required for the integration of electronic responses into existing processing systems. The statistical production of many CSIs is organized in the form of "stovepipes" [2] where individual surveys are to a large extent implemented by different organizational units and are supported by tailor-made information systems. Getting rid of these historically grown stovepipes by combining similar processes within different surveys is no simple, short-termed matter. Thus, even if an integrated, survey-independent system for the reception of electronic raw data is built, it still has to come to terms with the existing stovepipe applications.

10. All these problems notwithstanding, it is very likely that electronic means of data reporting will soon become a standard. In the case of organizations with a wide range of surveys to manage, however, the effort involved in developing, disseminating and maintaining electronic questionnaires tailored to specific surveys is prohibitive – regardless, whether they are implemented as stand-alone CSAQ (computerized self administered questionnaire) software or in the form of web browser applications.

11. In order to cope with these future requirements, it is not enough just to develop questionnaire software. What is needed is a complete infrastructure covering all phases of the data collection process

from the preparation of electronic questionnaires (by subject matter specialists with no or only limited help from IT experts!) up to the management and initial processing of the transmitted raw data. This system should be suited even for complex surveys, and it should be so flexible that it can be adapted to the particular needs of any specific survey. The core element should be a generic "electronic questionnaire management system" (or multi-questionnaire CSAQ system, as it has been called [3]) that would avoid the above-mentioned trap of individually developed single-survey applications.

12. In 1998 Statistics Austria initiated a research project with the mandate to investigate the feasibility of developing a generic, metadata-based system as outlined above. After specifying the requirements, investigating several design alternatives, acquiring know-how in the fields of object-oriented software development and Microsoft's Component Object Model, analysing the system in the form of UML diagrams, writing prototype programs as proof of concept and conducting a call for tenders, in November 1999 CSC Austria was retained to develop the SDSE ("System zur Durchführung Statistischer Erhebungen"3), which is now usually just called e-Quest after its main component.

13. Working closely together, the two project teams succeeded not only in implementing the metadata- and XML-based software system in less than two years, but also in preparing and deploying function-rich electronic questionnaires for four different surveys in the economic field (among them the monthly short-term surveys with about 18.000 respondents and seven different questionnaire types in April 2001, and the annual structural business survey with almost 45.000 respondents and five questionnaire types in August 2001). They also accomplished the integration of the electronic responses into the existing stovepipe systems that carry out the actual statistical evaluation.

II. BASIC PRINCIPLES

14. At the beginning of its work, the Statistics Austria project team elaborated a set of basic principles which functioned as guidelines when the e-Quest system was designed. These principles are:

a) First and foremost, electronic questionnaires must benefit the respondents, or else the respondents will have no incentive to use them;
b) e-Quest must be usable for different surveys, including the most complex economic surveys carried out by Statistics Austria;
c) It must be possible to change an electronic questionnaire without modification of the software's source code;
d) e-Quest must be capable of extension;
e) Data security must be given very high priority;
f) For surveys in the economic field the software must be practicable for enterprises of all sizes as well as for third party declarants (for example accountant firms) who are empowered to fill in the questionnaires on behalf of their clients;
g) Regarding the design of the system, present and foreseeable future IT trends should be taken into consideration;
h) Last but not least, e-Quest must be useful not only to the respondents, but also to Statistics Austria.

How these principles have been realized, will be described in section III.

15. Why is e-Quest a CSAQ program which has to be installed on the respondents' computers, and not a browser-based online application deployed over the web? This decision had to be made quite early in the course of the project because several other design issues would depend on it. After evaluating the pros and cons, the first alternative was chosen because of several reasons.

16. In the first place, the complexity of some economic surveys conducted by Statistics Austria spoke against a browser solution. Depending on the size of the enterprise, a respondent may be obliged to complete several questionnaire forms, each of them containing at least a hundred questions (but possibly a lot more, if a large number of different goods are manufactured). Moreover, which fields have to be

filled in for every product, is determined by the respective PRODCOM classification code. All in all, collecting the data and entering them into the forms may take quite a long time. It would task the patience of the respondents if they were required to do the work online – especially if they would have to pay the telephone fees in case of dial-up connections to the internet.

17. Another reason was our conviction that the much-hyped browser applications are a nuisance if more than a few fields have to be typed in. "Let's face it, in terms of developing friendly, responsive user interfaces, browser-based programming is a huge step backwards. HTML and the browser were not designed to be a platform for application user interfaces. They were designed for hyperlinked documents. Interfaces developed in a browser pale next to what we can do with tools like Visual Basic." [4] Admittedly, apart from pure HTML other technologies like Javascript, Java and Dynamic HTML could be used for more convenient user interfaces, but they would either ruin the advantage of being platform- and software-independent or they may bear the risk of security leaks (as the Nimda virus has shown just a few weeks ago, even the automatic execution of Javascript statements embedded in web pages or e-mails can lead to infections).

18. Questionnaires which comprise a large number of fields are often answered in more than one session, in some cases even by different people. Therefore the software must offer the possibility to store and load unfinished forms. If a browser-based system uses a central database accessible over the internet for this function, again potential safety problems (like, for example, hacker attacks) have to be taken into consideration. Certainly it is easier to reach a high level of security if data are stored locally.

19. Another matter that has to be taken care of concerns the performance of the electronic questionnaire solution. In case of a browser-based online application, the capacities of web and database servers must be sufficient for high peak loads near the reporting dates of surveys with many respondents. While the issue of server capacities can be solved – it's just a question of money – other aspects (e.g. the bandwidth of the respondent's and his/her service provider's connections to the internet, technical problems of the internet service provider, etc.) are beyond the influence of the CSI. If every page of a multi-form questionnaire suddenly takes minutes to load, the users will quickly return to paper questionnaires.

20. These and some other reasons led to the conception of the e-Quest system as described in the next section.

III. THE e-QUEST SYSTEM

21. Fig. 1 on the next page shows the components of the e-Quest system and how they are connected by data flows.

22. The preparation of a survey begins with the definition of metadata which describe – among other objects – the collector of statistical information, the survey and its version, the types of observation units (for example "enterprise", "establishment" and "local unit of work", or "household" and "family member"), the possible hierarchical relationships between them and the questionnaire types associated with them. Furthermore, the questionnaire forms – in detail, the logical structure and the properties of the questions, their representation by GUI elements, layout, event-driven flow of control, intra- and inter-questionnaire validation – are designed. All this is done by statisticians with the e-Quest Metadata Manager software. Finally, these structural metadata, as they are called, are exported as an XML file.

23. If the respondents and the observation units for which they are responsible are pre-defined by the CSI, these respondent-specific metadata as well as initialisation data (for example, name and address of the observation unit, NACE codes etc., which will automatically be imported into actual questionnaire forms) must be prepared in XML syntax too. This happens outside of the e-Quest system (e.g. at Statistics Austria the respondents of business surveys and their observation units are managed in the business register database on the mainframe computer).

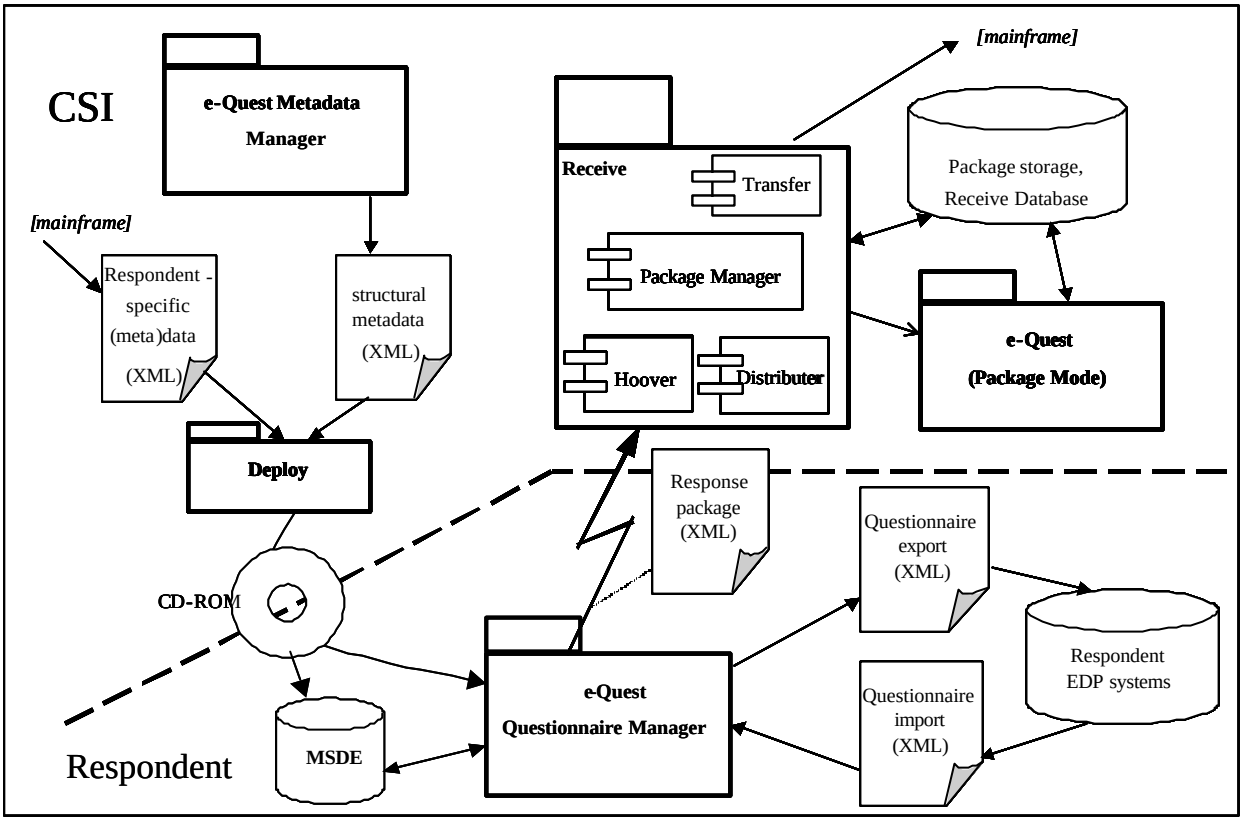

Fig. 1

24. The metadata files are compressed and encrypted by means of the Deploy software. The resulting file is distributed to the respondents, for the time being usually on CD-ROM, but other methods like e-mail or downloads from the web are possible too.

25. On the CD-ROM the respondents also receive the core component of the system, the e-Quest Questionnaire Manager software and the relational database system MSDE (a version of Microsoft's SQL Server which can be deployed free of charge). This software must be installed. As long as there are no software updates, a one-time installation is sufficient, even if the respondent is obliged to report for several surveys (which, by the way, could even be carried out by different CSIs). Next the structural and respondent-specific metadata are loaded into the application which then manages all aspects of questionnaire initialisation and presentation, data management and validation, response data encryption and transmission to the CSI.

26. The e-Quest software offers an export/import interface in a simply-structured XML format. This feature also allows the export of metadata describing certain properties of the questions, for example data type, minimum and maximum values, a short description etc. Such an export file can be used by a programmer as a template, when he/she writes a program to create import files containing data from existing EDP systems of the respondent in order to automate the data reporting task.

27. Completed questionnaires are compressed, encrypted and transmitted by e-mail or FTP to specific servers and addresses defined in the structural metadata. If no internet connection is available, the response package can be exported as a file and sent on diskette by "snail mail".

27. The Receive subsystem on the CSI side consists of programs to listen for incoming response packages containing the data from one or more questionnaires in a standardized XML format (Hoover), to backup, decrypt, register, check and forward them to the responsible organizational unit (Distributor) and for their administration (Package Manager, also called Pot Application). For viewing and eventually correcting the data the e-Quest Questionnaire Manager is used in so-called "package mode", offering additional functions like logging the change history of every input field (so that the original respondent data can be reconstructed if necessary). The questionnaire forms may differ from those presented to the respondents. Finally the Transfer program, supplemented by components which must be written specifically for each survey, is responsible for the conversion and transmission of the data to the stovepipe systems where further processing and evaluation are performed.

28. This much as a short overview of the e-Quest system. Let us now look back to the basic principles listed in section II. How have they been realized in e-Quest?

a) Electronic questionnaires must benefit the respondents:
i)Many features have been implemented to increase the practical value of the software. Although the initial task of installing the application takes quite a while (apart from the installation of e-Quest and the MSDE – which is a fully-featured relational database server, after all – the database tables have to be created and loaded), the respondent gets an integrated system for the presentation and management of electronic questionnaires in return for his/her efforts. If he/she is obliged to report for another survey, only the metadata describing it have to be imported into e-Quest. In such a case the user is not required to adapt to a new user interface or to learn to handle new software.

ii) The standardized export/import interface implemented in e-Quest makes it possible to automate the completion of questionnaires by importing data from the respondent's EDP systems. The "auto-import" feature allows a simple configuration so that answers which stay unchanged in different instances of periodical surveys are filled in automatically when a new questionnaire form is created. Questionnaires can also be copied, so that only changes from the previous survey period have to be filled in.

iii) Regarding the dynamical creation of questionnaire forms according to the structural metadata prepared by the CSI's staff, e-Quest is able to present plenty of graphical user interface elements including but not limited to text fields, buttons, multi-choice questions and tables. These items can be arranged on one or more pages (tabs), grouped hierarchically to any depth with single or repeating groups (arrays). A wide range of functions are provided, such as immediate or delayed input checks, visibility and activation management (the user only sees those questions that must be filled in) and an event-driven flow of control depending on user interactions with the questionnaires (implementing, for example, the automatic calculation of row and column sums in a table).

iv) It is not necessary to complete questionnaires in one session. The answers can be stored in the database at any time, even if they are causing validation errors (but as long as these errors have not been corrected, it is not possible to transmit the response data to the CSI). Speaking of validation, questionnaire data can be checked for plausibility against standard or freely defined conditions. The former include data type, length and value range checks of single field values, the latter can involve any number of values, utilizing subroutines and functions written by the questionnaire form designer in Visual Basic Script language and embedded within the structural metadata file. Three levels of severity (error, overridable error and warning) can be associated to validation rules.

v) In addition to intra-questionnaire validation, all the respondent's questionnaires relating to a certain survey instance may be checked together by inter-questionnaire checks (for example, the

number of employees in the enterprise questionnaire must be equal to or greater than the sum of the employee numbers entered into the business unit questionnaires).

vi) Other important characteristics of e-Quest are an extensive help system comprising several levels of information, and plug-in components alleviating the task of searching for an appropriate classification code (more on these two features in section IV). The user can print selected questionnaires at any time but these printouts are not accepted by Statistics Austria as responses. Some additional highlights of the e-Quest application designed to increase the benefits for the users will be mentioned in the next paragraphs.

b) e-Quest must be usable for different surveys, including complex economic ones:
This principle has been realized by designing e-Quest not as yet another tailor-made electronic questionnaire software, but as a generic questionnaire management system. Most of the analysis work was based on the structural business survey and the short-term survey. The idea behind this procedure was that – if e-Quest is able to satisfy the requirements of these two surveys – the implementation of less complex ones would constitute no problem. After finishing the development of the software (at least of version 1) and deploying it for four different surveys (others are in preparation), it can be stated that this premise has been proved correct.

c) Changes of a questionnaire must not require modifications of the program's source code:
In any complex application, metadata describing the structure and contents of other data are of course omnipresent. All too often, however, much of the metadata are implicit in the source code and distributed throughout the system. In contrast, a generic system like e-Quest aims to manage explicitly as much of the metadata as possible, creating program and data flows dynamically, based on the actual metadata active at any given time. This conception makes it possible to modify existing questionnaire forms and to deploy new ones just by loading XML files into the application, without any need to change the source code.

d) e-Quest must be capable of extension:
Apart from being applicable to statistical surveys of (almost) any degree of complexity, new classifications like PRODCOM, NACE or the Combined Nomenclature can be installed as plug-in components conforming to a general COM interface (see section IV).

e) Data security must be given very high priority:
i) By keeping all data locally on the respondent's computer, an open internet connection is only required during the short period of time when the responses (which are encrypted by RSA, an asymmetric public/private key algorithm) are transmitted by e-mail or FTP – or not at all, if the response packages are exported as files instead. Data and metadata are stored in encrypted form in a relational database management system installed either on the same PC as e-Quest or on a dedicated database server. All security measures offered by this RDBMS contribute to guarantee the safety of the data.

ii) In order to maintain confidentiality of respondent-specific metadata and initialisation data, they are encrypted using a symmetric algorithm. The key necessary to decrypt and load them into e-Quest is sent to each respondent in a separate envelope.

iii) Further features enhancing security include the optional definition of user-ids and passwords, the optional management of access rights (so that a user can only work on those surveys, respondents, observation units and questionnaire types where he/she has been authorized by an administrator) and an auxiliary program capable of creating and reloading backup files containing all contents of the e-Quest database tables.

f) e-Quest must be practicable for enterprises of all sizes as well as for third party declarants:
i) Most users install e-Quest in stand-alone mode (with MSDE and client software on the same PC), but a network client/server installation for multi-user operation is supported as well. Of course it is guaranteed that a questionnaire which is being edited by one user is protected against simultaneous write access by others. An existing relational database server (Microsoft SQL Server, IBM DB2 UDB, IBM DB2/390) can be used instead of the MSDE.

ii) e-Quest is not only able to manage different surveys but also a large number of respondents, so that third party declarants such as accountant firms can administer the questionnaires of all clients within the same installation and database. In order to enhance usability, special functions have been implemented, e.g., the definition of an optional client level within the "navigator" window where all objects like respondents, surveys, observation units and questionnaires are displayed in a tree view, and the option to transmit the questionnaires of all respondents with a single command.

g) Present and foreseeable future IT trends should be taken into consideration:
i) e-Quest utilizes XML for the transmission of all data and metadata as well as for the export/import interface. As there were no existing XML standards in the field of electronic data reporting, all XML schemas have been defined within the project work, forming the non-official e-Quest standard.

ii) Other IT trends that have been integrated in the systems design are a component-based software architecture (Microsoft's Component Object Model COM) and the usage of the internet for data transmission (namely SMTP and FTP).

h) e-Quest must be useful to Statistics Austria as well:
i) As already described above, the e-Quest system provides a standardized, integrated and survey-independent infrastructure for the development and distribution of function-rich electronic questionnaires and for the initial processing of the received response data packages. Adaptations to existing stovepipe processing systems are supported by the definition of several COM interfaces, where survey-specific modules can be plugged in (e.g., the Distributor program can be configured to call software libraries performing additional validation checks on the incoming raw data).

ii) Another benefit for Statistics Austria is the reuse of software developed within the e-Quest project. Two plug-in components for displaying the NACE and PRODCOM classifications have been deployed to the respondents together with the e-Quest Questionnaire Manager application. These modules have also been integrated in a data editing program written and used at Statistics Austria – at practically no additional cost, as only a few source code statements were needed.

IV. METADATA IN THE CONTEXT OF EDR AND e-QUEST

29. During the last decade, in particular, discussions on metadata and metainformation systems have become very important in the world of statistics. Papers prepared for conferences like the UN/ECE METIS work sessions often focus on the dissemination phase of the statistical life cycle. In this context metadata are primarily seen as data needed by end users for the search, access and understanding of statistical information. A second main discussion topic is the creation and usage of metadata within the organizations responsible for the production of statistics. But the important role that metadata play in the phase of raw data collection should also not be overlooked – especially when a tool like e-Quest (which can be described as "metadata-based") is used for this task.

30. In the context of electronic data reporting by means of CSAQ software, several types of metadata can be identified. In the following we will distinguish three main categories:

a) descriptive metadata designed to enhance the respondent's understanding of the survey and questionnaire contents,
b) "collected" metadata originating from the respondent (in addition to the raw data entered into the questionnaire fields) or from a statistician,
c) internally used metadata, which the user is only aware of in an implicit fashion, as he/she is presented with the end product of its interpretation.

31. Types a) and c) tend to overlap, however. For example, in e-Quest survey- and survey-version-specific information like periodicity and validity dates are necessary to generate survey instances and to calculate the deadline for the responses to be sent to the CSI, but they can also be presented to the users.

32. On considering how to format and transmit data and metadata between the various e-Quest components and subsystems, the XML format quickly sprang to mind as a self-describing character-based data format currently becoming a de facto standard and easily adaptable to varying structures and contents. Thus XML became the format of choice for all data transfers and external interfaces, except where technical considerations led to the use of other alternatives. XML schemas describing the XML formats were defined and of course form a further level of metadata.

IV.1 Descriptive metadata

33. The first layer of descriptive metadata consists of information that is presented within the questionnaire form: the captions of the questions, answering hints, footnotes and so on. This layer exists in paper questionnaires as well as in electronic ones. In e-Quest this information is defined within the XML definition of the questionnaire.

34. The second metadata layer is specific to e-Quest. Whenever an edit field is focused, a short help text is displayed in the info window at the bottom of the screen. These texts are stored in a file which is installed together with the survey-specific help pages.

35. The next layer consists of comprehensive explanations with regard to answering the questions, including definitions of terms and, sometimes, examples. For surveys carried out on paper, these descriptions are usually printed on separate pages. In e-Quest they are distributed as an extensive help system in HTML format. When the user presses the F1 button, the browser is launched and a help page relating to the active question is presented. In comparison with printed explanations, this help system offers all the advantages of a hypertext system. If Java is enabled in the browser, a hierarchically structured content frame is displayed in addition to the help page, and the respondent can find a topic with the help of an index or by using full-text search. At Statistics Austria a commercial software called Robohelp is used for the preparation of the help pages.

36. Metadata layer 4 comprises general information about the survey, e.g. its objectives, its periodicity, its sampling methods, and so on. For paper questionnaires this information is provided in printed form. In e-Quest it is part of the HTML help system, but it is considerably more detailed. The complete text of the European Union's regulations on a survey as well as those of the responsible Austrian ministry and the Federal Statistics Act 2000 are included for those respondents who are interested in the legal basis of the survey.

37. Classifications are a special type of descriptive metadata. While some classifications are just short code lists (like "Gender" with the items "male" and "female"), others are quite complex hierarchically structured "information objects" where items may be linked to items of other versions or other classifications or to additional metadata (for example a list of synonyms allocated to classification codes). With paper questionnaires where specific codes have to be entered (such as PRODCOM codes for the classification of produced goods), sometimes dozens of pages describing the classification items have to be provided to the respondents. Obviously, such voluminous classifications do not facilitate the job of completing questionnaires.

38. In e-Quest short code lists can be defined within the description of the questionnaire in XML format. For extensive classifications, however, the format is not predefined. Specific classification modules, which must conform to a certain COM interface and which can be deployed separately and installed as plug-ins, are responsible for the management of their own data. Currently two of these components (for NACE and PRODCOM) are in use. They store their data in relational databases, but this is no mandatory design. At some time in the future, a component will perhaps be capable of querying a (in this form not yet existent) classification server of Statistics Austria or even of leaving the choice to the users whether the classification data should be accessed online or downloaded for local storage.

39. A last type of e-Quest metadata designed to enhance the respondent's understanding of the questionnaire contents is not aimed at users filling in the statistical declarations, but at IT experts. It is possible to export questionnaires so that programmers can apply these files as a model when they develop extensions to the respondent's EDP systems which will automatically generate import files for e-Quest. Options are provided to include short explanations of the XML tags and further information such as data type and value ranges into these template files.

IV.2 "Collected" metadata

40. In order to avoid being called back by a staff member of the CSI, respondents who are completing paper questionnaires sometimes write additional comments on the form – what of course can be seen as another type of metadata in the context of EDR. When the e-Quest Questionnaire Manager application was designed, subject matter statisticians expressed an important requirement that such notes must also be possible in electronic questionnaires. For this reason, e-Quest offers several features intended for the collection of (so to speak) "raw metadata" – apart from the obvious way of including a "comment question" into the questionnaire form, of course.

41. The info window at the bottom of the screen can be switched to a "comment view". Here the user is able to attach a text note to any question which will be transmitted together with the field value to the CSI.

42. Validation rules, as has already been mentioned, can be of type "overridable error". If such an error occurs, the respondent is allowed to insist on his/her input, but is required to add a comment why he/she thinks that the answer is correct. Without this feature there would be the risk that the user just modifies the answer in such a way that the error disappears, lowering the quality of the collected raw data by doing so.

43. When the command to transmit a response package to the CSI is executed, a dialog window is displayed where a further comment can be entered. This comment can be seen by statisticians in the Package Manager client/server application without opening a questionnaire.

44. Finally, statisticians can open and edit incoming questionnaires in e-Quest. When the software is used in this mode, the complete history of changes to every field is kept, together with metadata showing date and time of the modification as well as the user-id of the statistician. It is also possible to attach a note to a modification.

IV.3 Internally used metadata

45. With regard to metadata used internally in e-Quest, two types can be distinguished. The first type are called structural metadata because they describe the survey structure: the collector of statistical information, the survey, survey version, observation unit and questionnaire types, permitted hierarchical relations between observation units, etc. These data are loaded into e-Quest and stored in database tables. The respondent is aware of them only in the sense that the survey structure is presented in the e-Quest "navigator" window and governs the actions permitted.

46. Also belonging to structural metadata are the questionnaire definitions which govern the appearance and the behaviour of questionnaire forms and are thus implicitly apparent to the respondent. Questionnaire definitions are stored as encrypted XML documents in database tables to prevent that respondents with “hacker mentality” can modify them (eliminating, for example, certain validation rules).

47. A questionnaire definition starts with the <QuestXML> tag and consists of five mandatory parts:

?? the logical structure (describing first and foremost the questions);
?? the GUI elements assigned to the logical items;
?? the layout;
?? the event handling;
?? and the validation rules.

48. These parts are identified by the tags <QUDEF>, <GUIDEF>, <LAYOUTDEF>, <CTRLDEF> and <PLAUSDEF>. Each of these tags comprises exactly one child element, which is the top level node for all subsequent definitions.

49. The logical structure of an e-Quest questionnaire is a hierarchy consisting of items which can either be “elementary items” (i.e. primarily questions; however, sometimes elementary items are used as buttons or labels as well), “groups” or “arrays”. The latter two can contain further items which again are of one of the aforementioned types. A codelist can be attached to an elementary item. The following UML diagram (fig. 2) shows the relations between these (meta)classes.

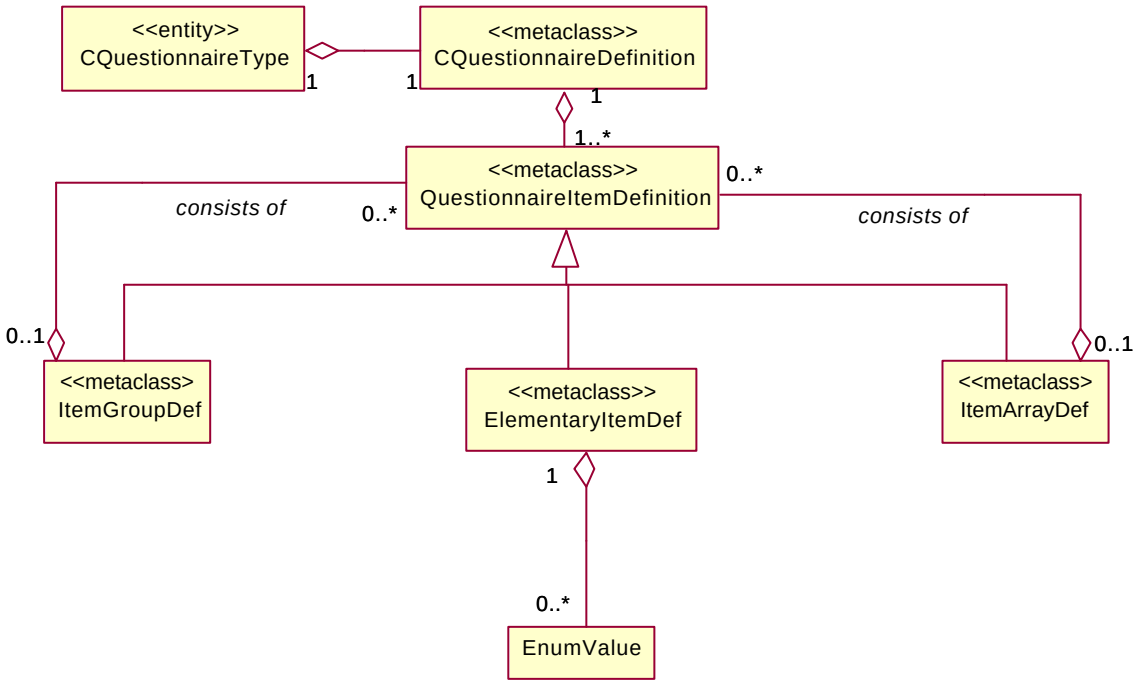

Fig. 2

50. Structural metadata are defined by statisticians with the e-Quest Metadata Manager, an interactive tool which provides – among others – the following features:

?? a tree view in which the logical structure of the questionnaire can be created;
?? a toolbox for the selection of graphical user interface element types;
?? a preview window on which the GUI elements can be placed;
?? a properties window which allows for the input of all property values pertaining to the currently selected element;
?? a font management window;
?? a script editor which allows the definition of variables, event triggers, subroutines, functions and validation rules;
?? a utility for generating the structural metadata file (which is then encrypted by means of the Deploy application).

51. Questionnaires can be imported directly into the e-Quest database where they can be tested immediately.

52. The second type of internally used metadata are respondent-specific metadata describing the respondent, the actual observation units for which the respondent must fill in questionnaires and the relationships between the units. These metadata are closely connected to so-called initialisation data which provide initial values for certain questions in specific questionnaires. These data (which are encrypted by a symmetric algorithm) are also loaded into e-Quest, often along with the structural metadata, and also influence the structures visible in the "navigator" window.

REFERENCES

[1] Ewald Kutzenberger, Werner Holzer and Norbert Rainer: Introducing Total Quality Management (TQM) in Statistics Austria. International Conference on Quality in Official Statistics, Stockholm, 14-15 May 2001

[2] G. Priest: Issues of Meta Information and Integration. UN/ECE Work Session on Registers and Administrative Records in Social and Demographic Statistics, Geneva, 11-13 November 1996

[3] Uwe Kunzler, Electronic Data Reporting (EDR) Strategy for the European Statistical System (ESS). ETK-NTTS Conference 2001, Hersonissos (Crete), 18-22 June 2001

[4] Billy Hollis, Death of the Browser? Column published on 14 October 2001 in the MSDN library (http://msdn.microsoft.com/library).