Distr.
GENERAL

CES/AC.71/2004/5
5 May 2004

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (ECE) 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 (MSIS)
(Geneva, 17-19 May 2004)

Topic (i): Web technology in statistical information systems

SAVING TIME AND MONEY: WEB QUESTIONNAIRES FROM THE "ASSEMBLY LINE"

Invited Paper

Submitted by Statistics Austria[1]

I. INTRODUCTION

1. In October 2002, the results of the *survey on the availability of electronic public services on the Internet*, carried out for the third time by Cap Gemini Ernst & Young, were published. This study – commissioned by the European Commission and the General Directorate "Information Society" within the framework of the eEurope programme – measured the availability of public services on the Internet and the level of online sophistication of the delivery process. The benchmark's objective was to enable participating countries to analyse progress in the field of eGovernment and to compare performance. As one of 20 basic public services, the study defined "Submission of Data to the Statistical Office". Austria and eight other EU states were rated as reaching online availability of 100% for this service. At first glance, this result may appear extremely satisfactory. We must, however, not overlook the fact that the definitions underlying the study will result in a rating of 100% as soon as one publicly accessible website allows the submission of at least one statistical questionnaire.

A. The Problem: the Multitude of Statistical Surveys

2. National Statistical Institutes (NSIs) carry out a multitude of surveys for which the Internet would afford a convenient reporting medium suitable to lessen the burden on respondents. From the citizens' and the statistical institutes' point of view, availability can thus only be graded 100% when every respondent affected by a survey with paper questionnaires also has the option of transmitting his or her data electronically.

3. Such a goal, however, cannot be attained in the traditional way – by "hand-crafting" Internet applications – when we bear in mind that statistical surveys are just as affected by ever shorter production cycles and shrinking budgets as most other projects are. Apart from the high cost of software development which accrue to each survey in addition to the paper questionnaire production costs, the provision of a timely

[1] Prepared by Wolfgang Koller, wolfgang.koller@statistik.gv.at, Günther Zettl, guenther.zettl@statistik.gv.at, Frederick Rennert, frederick.rennert@statistik.gv.at

electronic alternative to each new or modified survey presents an almost insurmountable barrier. The NSI may carry out software development in-house, thus eliminating the time periods inherent in a tender procedure and shortening the whole process. Even so, developing a user-friendly, safe and adequately tested Internet questionnaire application for a survey takes much longer than the traditional production process for paper questionnaires.

B. The ST.AT Solution: "Automatising Software Production"

4. In mid-2002, Statistics Austria (ST.AT) carried out a feasibility study on this subject. Just as for the PC solution "e-Quest"[2], which has been used successfully since 2001, the main focus for the development of a Web solution was laid on the motto "Expensive software experts should only be used where they are really necessary". We developed an approach which will allow us to manufacture Web questionnaires quickly, flexibly and – last but not least – extremely cost-effectively for the future. The solution proposed in the study starts with the subject matter statistician, who uses a comfortable PC program, the so-called "e-Quest Metadata Manager", to specify the metainformation relevant to a survey. With this tool, the required questionnaires are specified and designed with regard to contents, layout and logic. The specification is exported as metadata in XML format and is input into software production tools or generators which can produce – "at the push of a button" – a complete web-service-based Internet questionnaire application.

5. In this paper we will delineate this approach by outlining the scenario we envisage for its use. A separate section (III) is reserved for questions of application and security architecture. Section IV gives a brief description of the project – currently still in progress – in which the system (called e-Quest/Web) is being implemented. The Appendix provides further technical information including a short outline of the generator tool „UML2Web".

II. THE GOAL: FUTURE CREATION AND DEPLOYMENT OF WEB QUESTIONNAIRES IN ST.AT

6. The preparation of a survey begins with the definition of metadata which – among other objects – describe the collector of statistical information (CSI), the survey and its version, the types of observation units and the questionnaire types associated with them[3]. The contents of the questionnaires themselves must of course be specified and designed in a manner which will lend itself to an easily understandable representation for the respondents.

A. Create the form

7. A major goal of the aforementioned e-Quest system was to create tools which would allow these tasks of specification and design to be carried out by subject matter experts with only minimal external assistance. To this end, the efficient and user-friendly graphical editor „e-Quest Metadata Manager" was developed. This allows all aspects of an electronic questionnaire to be defined and exported in XML form without the user having to intervene manually in the creation of the XML representation. Figure 1 is a screenshot of this „Metadata Manager" demonstrating specification of an – in this case extremely simple – questionnaire.

---

[2] e-Quest is a generic, metadata-driven system for statistical raw data collection by means of self interviews with electronic questionnaires. It provides a standardized, integrated and survey-independent infrastructure for the development and distribution of function-rich electronic questionnaires and for the management and initial processing of the received response data. The core component of this system, the multi-questionnaire, multi-respondent, multi-survey and multi-user software "e-Quest Questionnaire Manager", is a product that especially suits the requirements of complex statistical surveys. The e-Quest system avoids the problems of individually developed single-survey applications. It is completely metadata-driven, based on a common object model describing the structure of a survey as well as the questionnaire forms, thus allowing the creation and maintenance of electronic questionnaires by statisticians with no or only limited help from IT experts .

[3] The complete survey structure object model as adapted to use in e-Quest/Web can be seen in Appendix B.

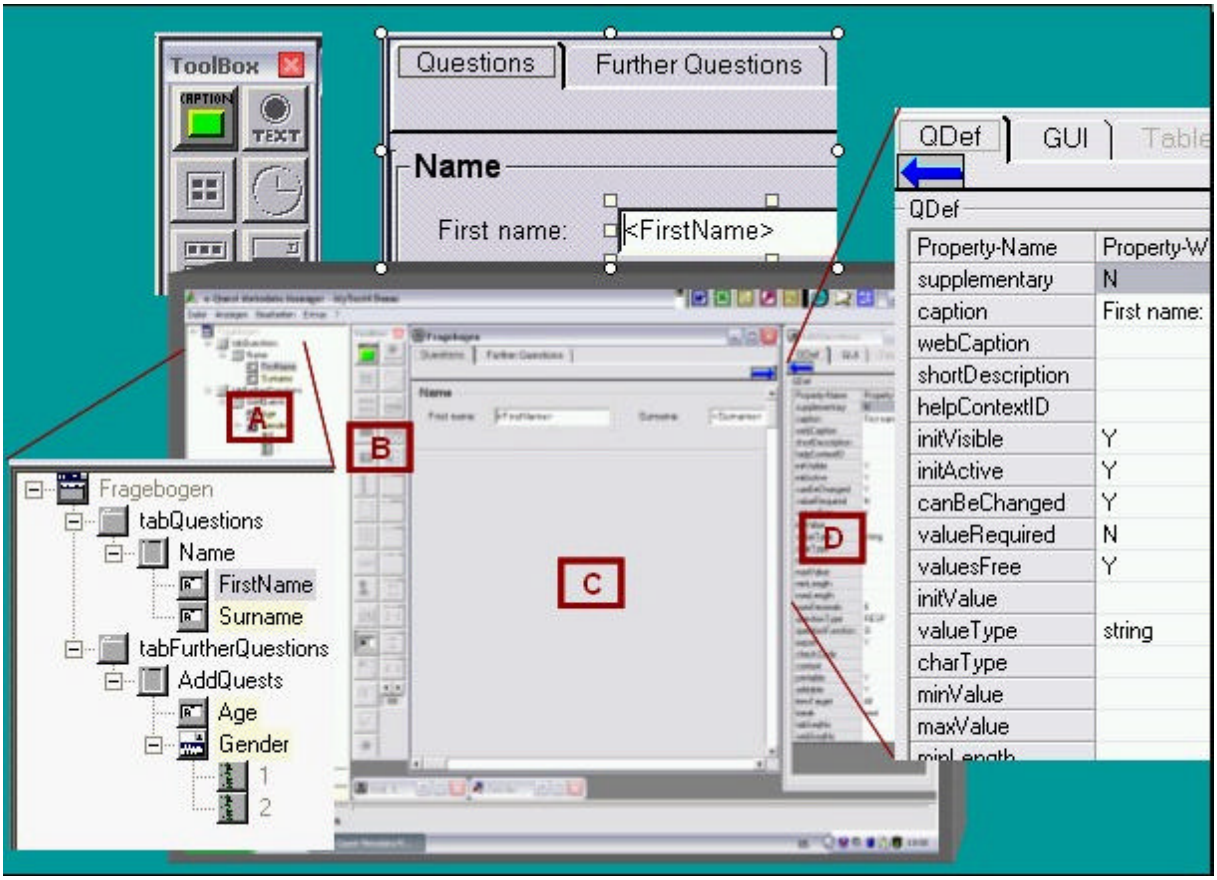


Figure 1: Creating a questionnaire with the Metadata Manager

8. Area **A** specifies the questionnaire's logical structure. Here you can specify the basic questions such as "Firstname" and "Surname", which can be gathered into groups ("Name"). The elements at the highest level (here "Questions", "Further Questions") represent separate screen pages which are visualized as filing cards and are used by ST.AT to structure larger questionnaires into thematic chapters and afford a better overview of the contents. Area **B** provides a tool box containing representation and formatting elements used to place the elements defined in area **A** into the window **C** by drag and drop, thereby creating the questionnaires' layout. Thus, questions may be represented on-screen as edit boxes, radio button groups, combo boxes etc. In the Property window **D** we can specify numerous attributes for each element (such as data types, minimum and maximum values, etc.). Over and above these windows the "e-Quest Metadata Manager" contains several other ones which are not visible in the figure. One of these is a script window which allows the designer to specify short scripts for complex consistency checks, non-standard error messages, navigation and event handling routines.

9. After specifying the questionnaire in this manner in the „e-Quest Metadata Manager", we store it in XML format – the XML resulting from the above example is shown in Appendix C – which forms the input for the generator in the next step.

B. Generate the software

10. The XML file resulting from the Metadata Manager not only contains the layout information but also all the information on contents (fields and structure), field attributes, check logic etc. A Web form containing both aspects – contents and layout – can be generated from this representation with little effort. In the course of the project e-Quest/Web a type of Internet application "Web questionnaires for Statistics Austria" was defined and standardised as a general schema (including application structure, work flow, construction rules, behaviour, security mechanisms etc.). Thus all required information needed to create the

desired software for a survey inside the ST.AT framework in the form of Web services[4] "at the push of a button" is at hand. Not only the web application software, HTML and JSP pages are generated, but also all necessary database definitions, an import application for initial data and a comparable Intranet application for Statistics Austria. The latter provides auxiliary functions which allow our statisticians to check and process the reported data.

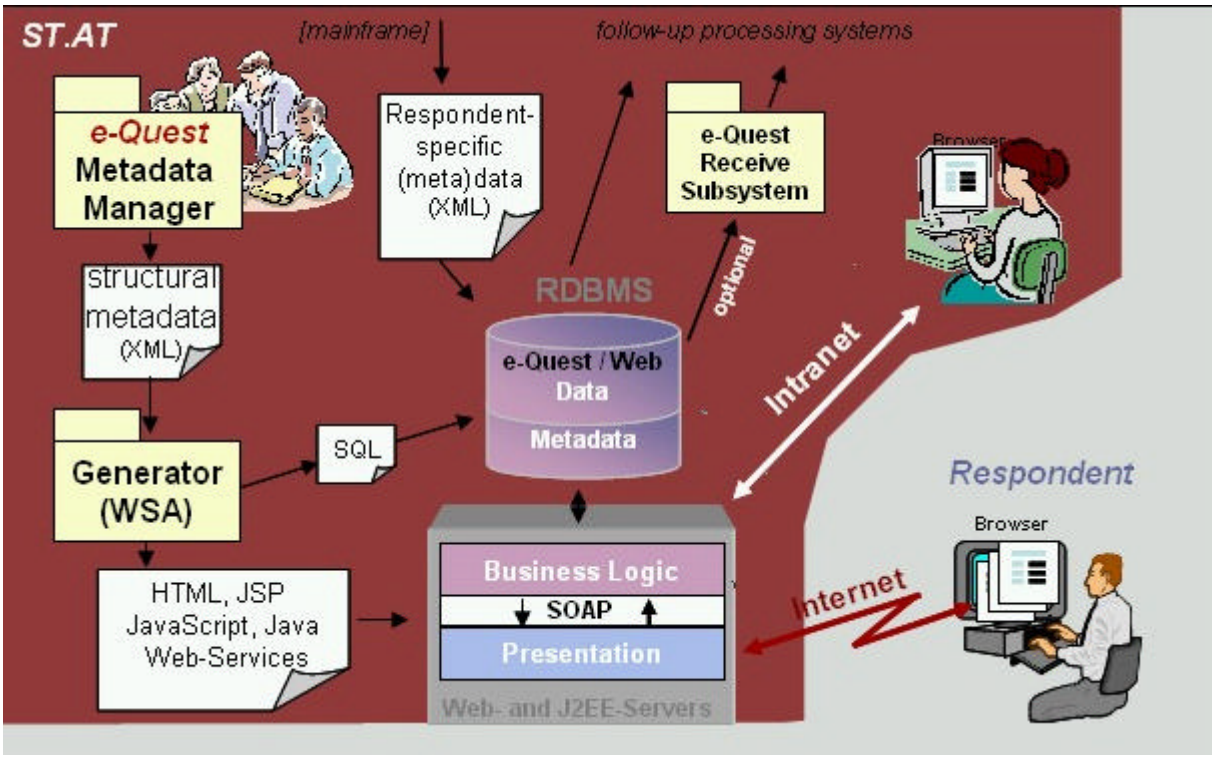


Figure 2: Survey using e-Quest/Web

C. Deployment

11. Before the Web application can be used by respondents, its software must be deployed to the target Web and application servers. For surveys aimed at a predefined list of respondents, respondent-specific data are loaded to the e-Quest/Web database. In most of the economic surveys carried out by Statistics Austria, the questionnaires are pre-filled in with the name and address of the enterprise, the code of its economic sector, its size code etc. These initial data originate from the enterprise register on our mainframe and are transformed into an XML interface format. In order for the respondents to access the application, user IDs and passwords are also generated and loaded to an LDAP server. These tasks done, the web questionnaire application can be released.

D. How can the user report over the internet?

12. The user accesses the web questionnaires via a link in Statistics Austria's homepage. A secure SSL connection is created and the user is asked to provide his or her user ID and password. These have been transmitted by post together with the paper questionnaires when the survey was first sent out.[5] After successful authorisation, the user is presented with a list of the questionnaires due to be filled in. These may stem from more than one survey, if the respondent has to report for them (this is often the case in the

[4] Web services are a promising technology. Able to mediate business logic via internet techniques independently of the user's system environment, they form an ideal foundation for service-oriented application architectures.

[5] It is a matter of policy with Statistics Austria to provide more than one reporting channel for each survey. Thus the respondent who lacks computer hardware or has qualms about the safety of his or her data on the internet always has the option of filling in the forms by hand and sending them by post.

economic sector). So-called “third party declarants”, such as tax accountants who may be responsible for several respondents’ questionnaires, are first presented with a list of all the respondents for which they are registered. This of course presupposes a prior agreement with Statistics Austria.

E. What functions are available to the user?

13. The selected questionnaire is opened for processing. The initial data are already filled in.

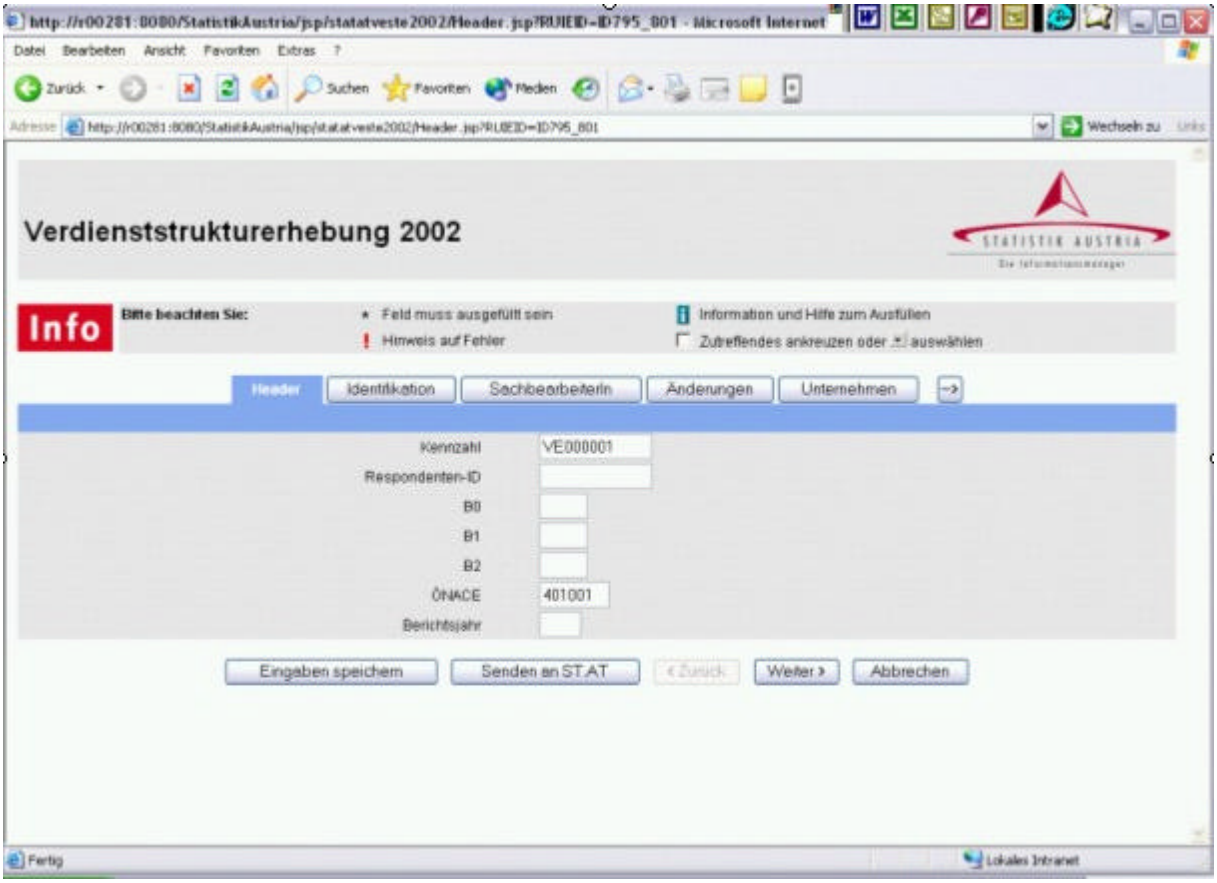


Figure 3: Screenshot of a generated questionnaire

14. The user can input the required data in any sequence. Functions such as sums, validity checks etc. are supported and are first carried out in the Internet browser via Javascript, if the user has enabled the latter. In any case, the calculations and checks are done on the server, so that error messages become visible after the next interaction. In accordance with the users’ greater security concerns when transmitting enterprise or personal data, no cookies are stored on the user’s computer. At any time, the user can store his or her data temporarily on the server and continue with the filling in at a later time. After completion, the button “Send to ST.AT” transmits the data to Statistics Austria where they are subjected to a final server-side validation. If no major errors are found, the questionnaire is removed from the list of open questionnaires.

F. How does the ST.AT subject matter statistician use the system?

15. Reported questionnaires are transferred to the Statistics Austria-internal system and can be checked and processed there in the Intranet. Excepting some auxiliary functions, this side uses the identical Web services from the Internet application. Any changes made to the data by the statisticians are logged. In certain cases, a reported questionnaire can be reassigned to the respondent for corrections. When any necessary processing has been finalised, the questionnaire is marked “finished” and passed to the respective successor system. There is also the option of transforming the data into the e-Quest XML format and forwarding it to the existing “Receive” subsystem.

III. ARCHITECTURE AND SECURITY

16. The system's architecture results from a study of the high requirements of official statistics with regard to security and data protection, within the constraints of ST.AT's strategic infrastructure. After completion, this study was subjected to an external security audit.

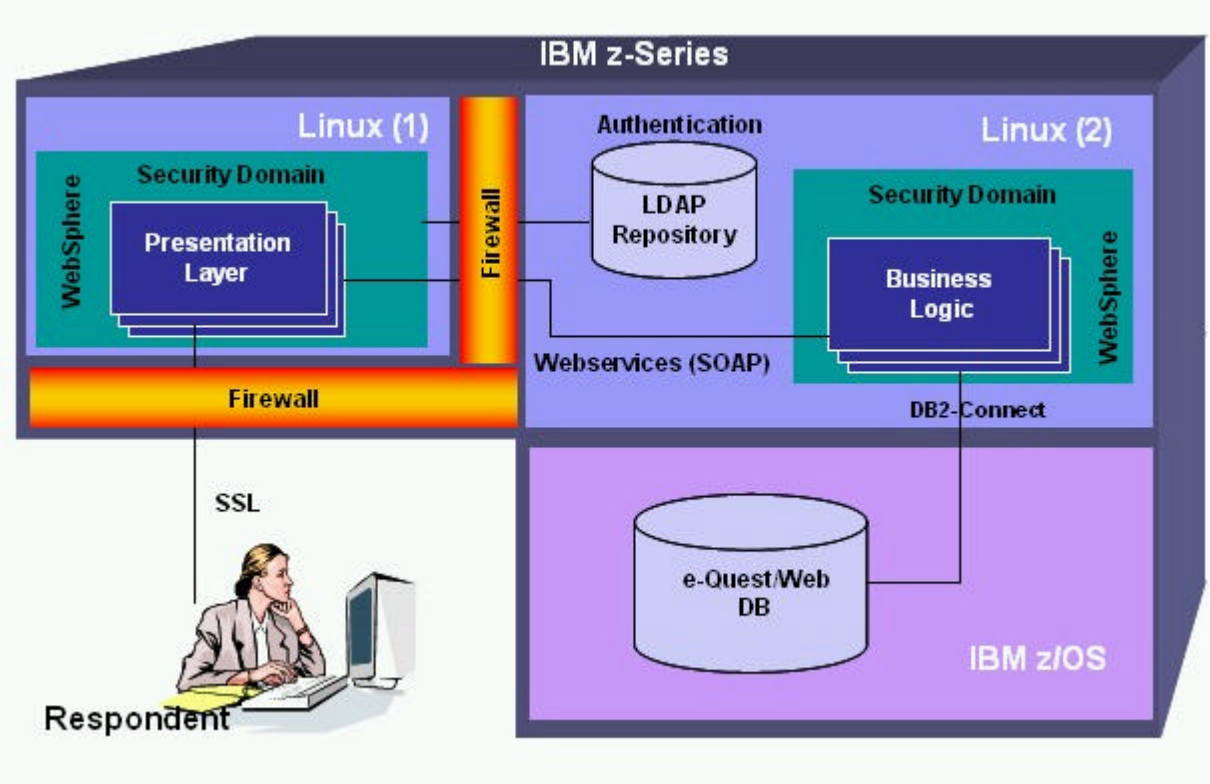


Figure 4: System architecture

17. As can be seen from the figure above, e-Quest/Web's architecture utilizes two virtual Linux environments and the relational database management system DB/2 under IBM z/OS. The user interface consists of a standard Web browser and HTML forms. The frontend is a server-side application on IBM WebSphere containing the presentation layer. Web browser and frontend communicate via HTTPS (http over SSL). The service layer with the business logic is separated from the presentation layer by operating system, detached from the Internet, and runs in its own Linux environment, as does the authentication mechanism LDAP. This second environment only accepts communications from the first one using the SOAP protocol. The second Linux environment's Web server is secured by a firewall which prohibits all requests via other than a specific port.

18. e-Quest/Web is designed with a service-oriented architecture in order to guarantee today's application requirements with regard to re-usability, openness and platform independence. Services allow for a flexible application and permit many separate applications to access the core system. The services form the interface between the application layers and communicate via the SOAP protocol.

19. An important factor in the construction of a secure system is the utilisation of generators and the use of a state machine for the request sequence. Generating the code – especially the code for database access – avoids security leaks due to manual programming errors. The resulting code is of uniform quality and can be regarded as very secure due to the tests to which the generator itself has been subjected. If the system design is error-free, we can therefore trust that the application itself will not be easily attacked. The state machine ensures that a possible assailant can only make requests which are permissible in the current application context. This makes it very difficult if not impossible for an attacker to succeed with a fatal assault at a chosen time.

IV. PROJECT IMPLEMENTATION

20. On the basis of the aforementioned study, the project was initialised in February 2003, together with an external software partner (Software AG in cooperation with CFC), after we had become acquainted with their product „Web Services Accelerator“ (WSA). This system generates Web services based on UML models which must be specified according to certain conventions and are converted to XML (see Appendix A). In the course of the project, the generators were adapted to accept the XML input produced by the e-Quest Metadata Manager. Their functionality was augmented to not only include the questionnaire contents but also the layout information and many of e-Quest's functions. The use of concepts and components which had already proven themselves in practical use resulted in project time and cost savings.

21. **Project Phase 1** encompassed analysis and design and lasted until the end of May 2003. This phase resulted in a detailed technical concept including all important aspects of the targeted solution: from the logical, technical and security architecture of the whole system to the metadata schemata to the specification of all components of a standard Internet application "Web questionnaires for Statistics Austria".

22. To define and standardise this Internet application type was of course the first step on which the targeted solution was founded. Its structure, construction rules, behaviour, security mechanisms, etc. had to be formulated clearly. We reduced the logical flow of a Web application for raw data collection to a general schema, implying that general components and interfaces could be extracted. Rules and regulations were formulated that determine and when necessary restrict function and appearance of the common tasks and elements of Web forms.

23. For the questionnaires themselves, we defined standard patterns, not only for their appearance but also for their contents and behaviour, up to and including the technical objects required for implementation. Whenever possible, we referred to existing standards or adapted them.

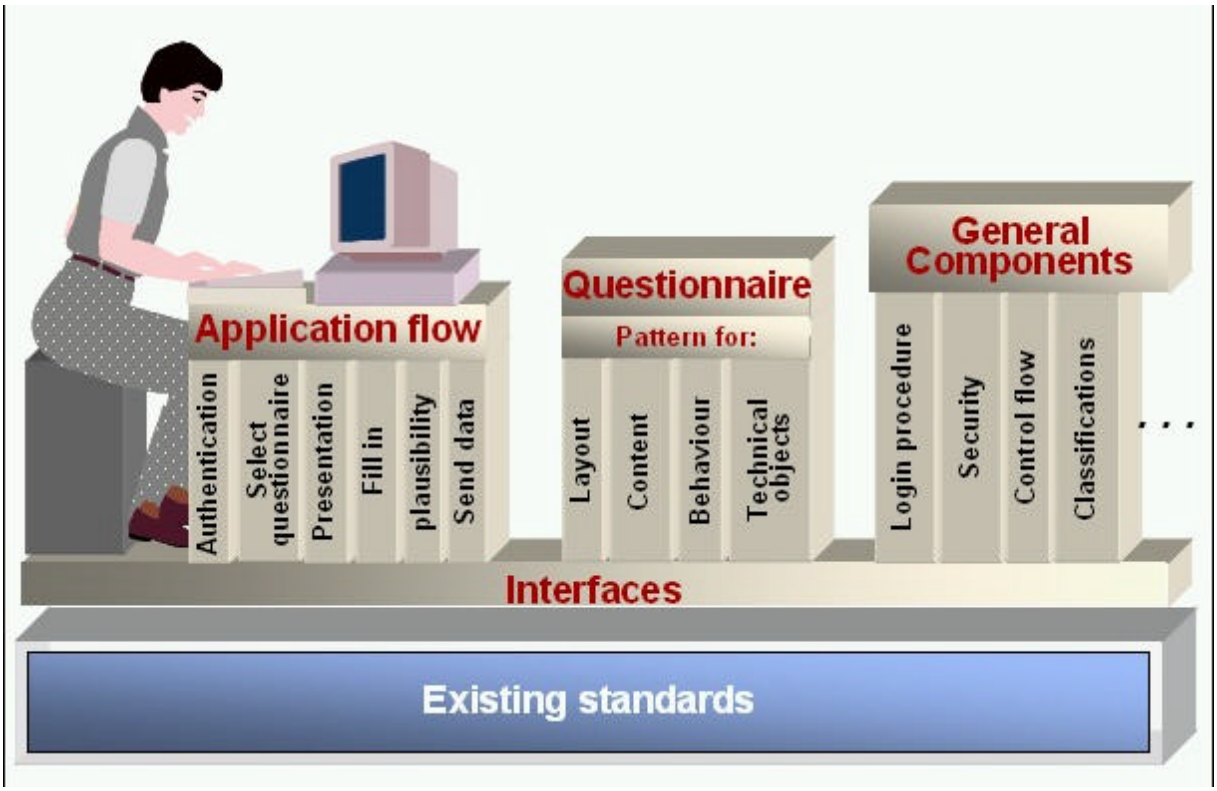


Figure 5: Standardisation

24. We furthermore not only attempted to achieve "re-use" – a generally accepted principle for increased application productivity and quality – for program components which have already been

developed and tested, but also to expand it to include previously input data or information and predefined patterns or structural elements.

25. **Project Phase 2**, including implementation and deployment, was planned for a period of about 8 ½ calender months. The SAG/CFC model and service generators which form the core of WSA were adapted to ST.AT's requirements and system environment. Specifically, the component which in the standard version of WSA expects input from a UML/XMI repository was enhanced and augmented to accept the Metadata Manager's XML output. This allows the definitions from the Metadata manager to be transformed automatically into Java code and connects them with the predefined standard components (authorisation, security mechanisms, questionnaire workflow, etc.).

26. WSA's GUI component was adapted to produce a uniform questionnaire appearance for the Web from the Metadata Manager questionnaire definitions with the use of style sheets tailored to the Austrian e-government standards. The standard persistency mechanisms were adapted to the storage of data input by the respondents. Specific enhancements were also introduced in the communication between Web client and server and for the display of the voluminous classification data (such as NACE) required for certain questionnaires.

27. At present (end of February 2004) the project is undergoing final testing.

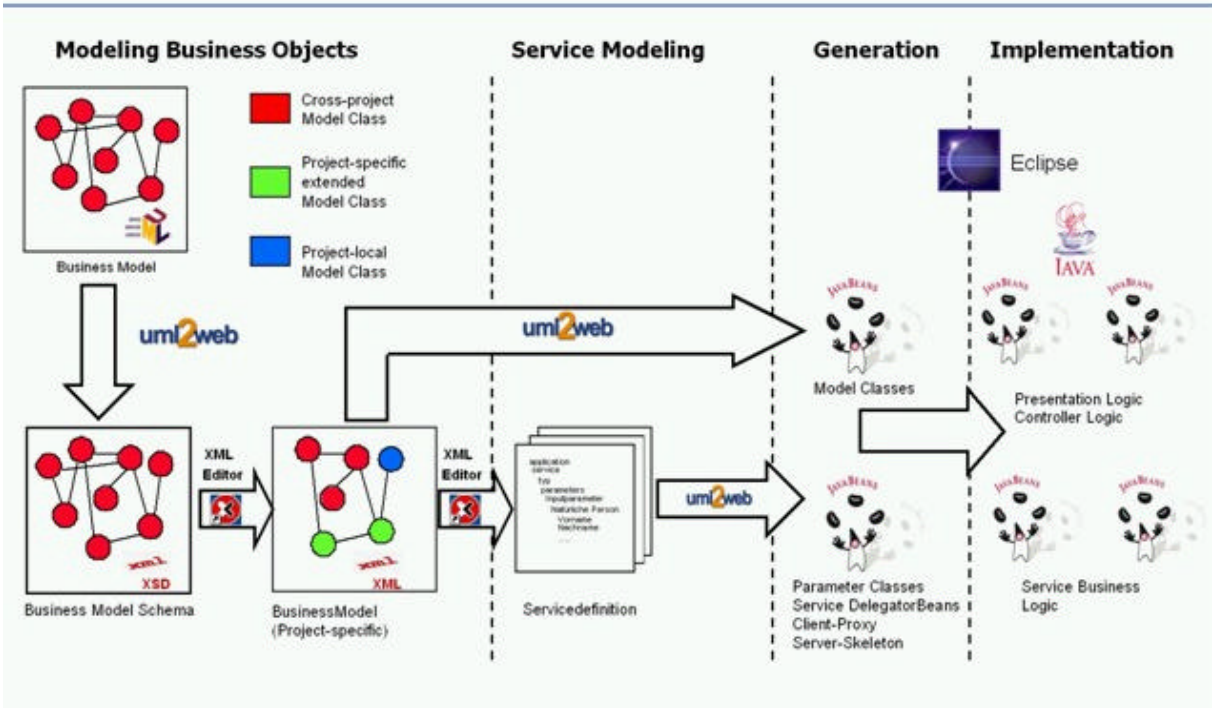
V. APPENDIX

A. WSA (UML2Web)

(*This text has kindly been provided by SAG and CFC Ges. m b H.*)

UML2Web provides an approved method for project managers, software engineers, requirements analysts or software developers to carry out major tasks in the software development cycle automatically and so to produce software much faster. UML2Web also takes care of test and documentation tasks.

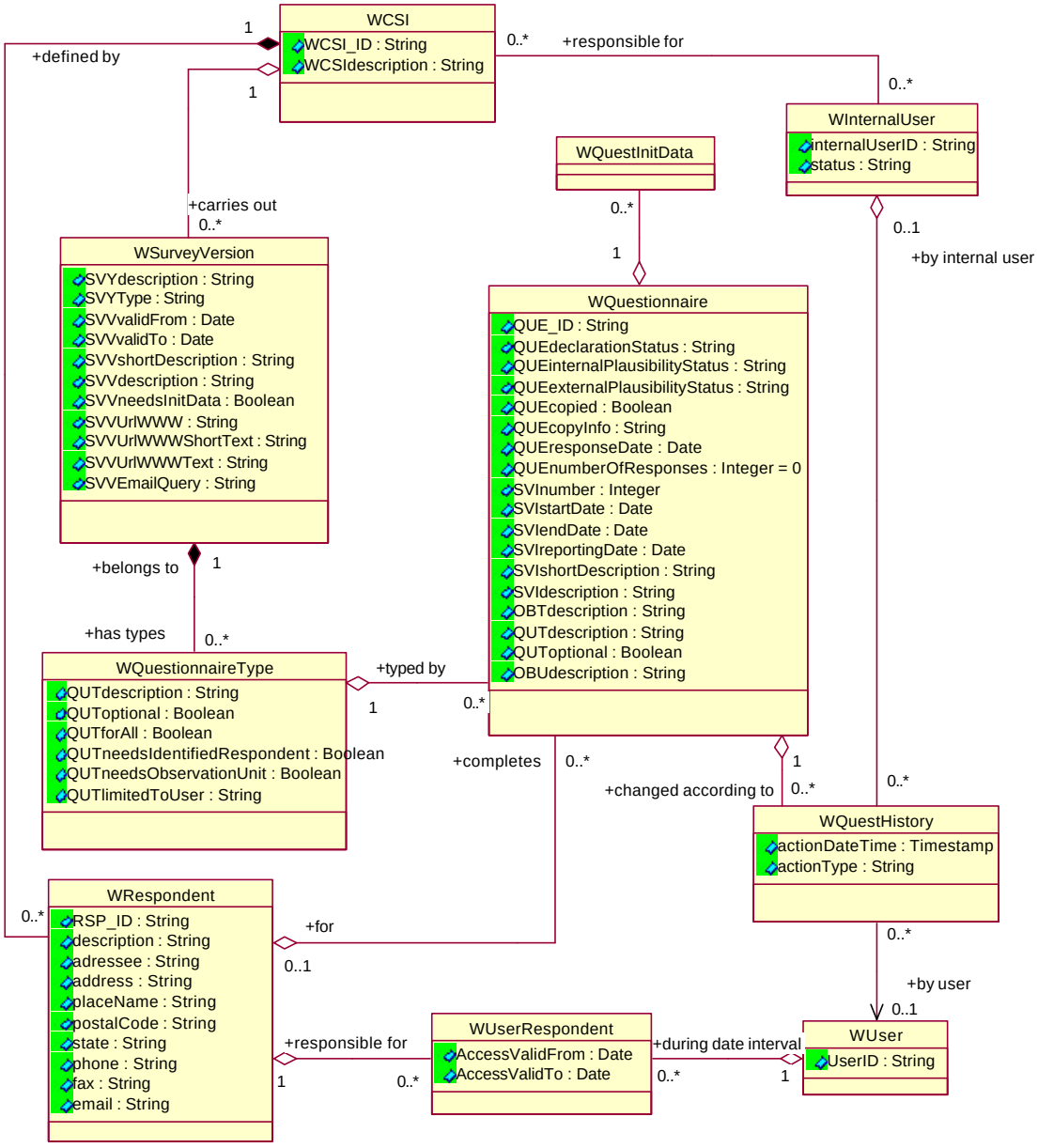
UML2Web includes a development process based on Model Driven Architecture principles on the one hand, and generators and transformers which give optimal support to this process on the other hand.



UML2Web is extensively used in JAVA environments. Applications supporting J2EE architectures with links to host systems as well as client-server architectures with fat clients – even combined architectures with web clients and fat clients – can be produced using UML2Web components.

The UML2Web generators and transformers are platform-independent and can be used in commercial development environments as well as together with Open Source components. UML2Web runs on UNIX / Linux and on Microsoft operating systems.

B. Object model of survey structure for Web questionnaires



C. XML code generated for the example in Figure 1

The following XML code is the result produced by the „e-Quest Metadata Manager“ from the example questionnaire in Figure 1. The questionnaire consists of two pages (<TabDef>), each containing a group (<ItemGroupDef>) with two questions (<ElementaryItemDef>). Three of these questions are displayed as text boxes, the fourth as a combo box.
Note: for better clarity, repeated tags are not shown in their full hierarchical depth. Such tags are marked by a „+“ in front of the line.

```
<?xml version="1.0" encoding="iso-8859-1" ?>
- <QuestionnaireDefinition CSI_ID="ZS" SVY_ID="A" QUT_ID="U" SVV_Version="2001" quName="A
  2001 U" caption="Survey A 2001 - U (monthly)" shortDescription="A 2001 - U"
  eQuestVersion="2">
  <FontDef fontID="FONTDEFAULT" fontName="Arial" fontSize="10" />
  <FontDef fontID="FONTBOLD" fontName="Arial" fontBold="yes" fontSize="12" />
- <TabDef tabID="tabQuestions" caption="Questions">
  - <ItemGroupDef itemName="Name" caption="Name">
    - <ElementaryItemDef itemName="FirstName" caption="First name:"
        valueType="string">
      - <ItemCtlDef>
        - <FieldCtlDef>
          - <CtlDefTextBox ctlType="TextBox" foreColor="COLWINDOWTEXT"
              backColor="COLWINDOWBACKGROUND" borderStyle="1">
              <CtlLayoutDef left="1658" top="452" height="355" width="2335" />
            </CtlDefTextBox>
          </FieldCtlDef>
        - <CaptionCtlDef>
          - <CtlDefLabel ctlType="Label" foreColor="COLBUTTONTEXT"
              backColor="TRANSPARENT">
              <CtlLayoutDef left="308" top="488" height="255" width="1215" />
            </CtlDefLabel>
          </CaptionCtlDef>
        </ItemCtlDef>
      </ElementaryItemDef>
    + <ElementaryItemDef itemName="Surname" caption="Surname:" valueType="string">
    + <GroupCtlDef>
    </ItemGroupDef>
  </TabDef>
- <TabDef tabID="tabFurtherQuestions" caption="Further Questions">
  - <ItemGroupDef itemName="AddQuests" caption="Additional questions">
    + <ElementaryItemDef itemName="Age" caption="Please enter your age:"
        valueType="string">
    - <ElementaryItemDef itemName="Gender" caption="Gender:" valueType="string">
        <EnumValue enumCode="1">male</EnumValue>
        <EnumValue enumCode="2">female</EnumValue>
      + <ItemCtlDef>
      </ElementaryItemDef>
    + <GroupCtlDef>
    </ItemGroupDef>
  </TabDef>
</QuestionnaireDefinition>
```