STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-third plenary session
(Geneva, 13-15 June 2005)

WEB-BASED SURVEYS AT STATISTICS AUSTRIA: ISSUES, DECISIONS AND SOLUTIONS

Invited paper submitted by Statistics Austria∗

INTRODUCTION

1.   The Austrian Federal Statistics Act 2000 explicitly mandates the provision of electronic questionnaires free of charge, so as to lessen respondents' burden. This is only one of the reasons that such a provision has become an important concern for Statistics Austria and has been included in the institution's mission statement.

2.   For Statistics Austria (ST.AT), this implies that, in addition to paper questionnaires, one or more electronic variants are or may be implemented in any survey.

3.   Why several variants? Careful examination of the questions involved and the efforts expended by respondents in replying to surveys, and discussions with business sector representatives on their ideas and wishes, led us to the conclusion that one technology alone cannot satisfy all requirements. None of the various forms of electronic data collection show sufficient advantage to justify implementing this method only; neither can the paper variant be replaced completely, except in special circumstances.

4.   This is confirmed by the current usage quotas in ST.AT. For instance, in the Intrastat survey 2004, 56% of enterprises reported via EDIFACT format, which can be generated by

_____

∗ Prepared by Wolfgang Koller, Frederick Rennert and Guenther Zettl.

the Eurostat program IDEP/CN8 but which is also offered as an output option by several business software products. 29% reported through Web questionnaire and 15% on paper.

5.   The development of electronic questionnaires for a wide spectrum of statistical surveys is, however, expensive and time-consuming. As we have been able to demonstrate, automatising the software production process as far as possible for this form of data collection can contribute to significant savings in the number of personnel involved and thus also reduce costs.

6.   With "e-Quest, the electronic questionnaire manager"[1], a Windows application in production since 2001, Statistics Austria created a flexible and cost-efficient option for satisfying the requirements of those business respondents confronted – due to their type of economic activity or the structure of their enterprises – with many complex and voluminous questionnaires.

7.   In the autumn of 2004, Statistics Austria completed the project "Web Questionnaires". This provides a production system for Internet questionnaires (e-Quest/Web), permitting the cost-effective implementation of Web questionnaires for a broad range of surveys. A complete Internet Web questionnaire application for a survey can be created and implemented automatically, by quasi "pressing the button".

8.   In the following sections, those questions are discussed which appear to the authors (who designed both e-Quest and e-Quest/Web) to be crucial to the development and deployment of Internet questionnaires.

SOME FUNDAMENTAL REMARKS

9.   One of the basic challenges in designing an electronic questionnaire is that the design decisions taken in the run-up to the project itself, concerning such vaguely defined and/or all-encompassing subjects as "lessening the burden on respondents" or "security", lead to measures which a future user may perceive and judge in widely varying ways, depending on his or her individual situation. The results of such decisions thus become crucial factors for the questionnaire's range of application and user acceptance, influencing the degree to which the new reporting medium will be utilized. The users will not be motivated to use an electronic questionnaire if they see no concrete personal advantage in so doing.

10.   In economic statistics, for example, the effort needed to respond adequately varies greatly, even within the same survey. A large enterprise with multiple establishments has to answer far more questions – and in greater depth – than does a small one. The respondents' demands with regard to electronic support and ease of use thus diverge enormously. A web form may completely satisfy all the requirements of small firms while medium-sized and large enterprises demand enhanced support functions and – crucially – satisfactory performance, even with large amounts of data, if the electronic track is to be perceived as useful and not as an additional burden.

11.   Attention must be paid to the respondents' security concerns and to safeguarding their privacy. For example, enterprise management often feels strongly about Internet security, establishing internal security rules and restrictions on browser use (e.g. no scripting, no

cookies) in order to ensure that sensitive enterprise data cannot be spied on. Clients providing such data will not accept Internet questionnaires unless they trust the institution carrying out the survey, and recognize that the system design of the web questionnaire itself does not counteract their security measures.

12.   The financial backers of an electronic questionnaire project attach certain expectations as to its use. As a result of their investment, they anticipate savings on personnel expenses and/or improvements in processing time and quality. In general, however, the electronic questionnaire will not succeed in wholly replacing the paper version, but will form an additional data collection method that must be serviced.

13.   However, the costs incurred must be seen as an investment that should pay off in the medium to long term. In this respect, it is important to observe the whole environment – creation of questionnaires, initial processing of incoming data, data validation, etc. – and to find cost-effective solutions for all the tasks. If we take the Intrastat survey as an example, where electronic questionnaires have been in use for several years, Statistics Austria was able to reduce the number of subject matter persons involved in processing the data from 79 in the year 2000 to 45 at the end of 2003. In the monthly Short Term Survey, the reduction was from 50 persons in 1996 to 30 today.

14.   One aspect, which at first sight offers considerable potential for economy, is the shifting of some responsibility for data editing towards the respondent. This, however, only makes sense to the degree that it does not violate the aforementioned principle of lessening the respondent's burden. Moreover, such a rearrangement also calls for changes to the NSO's internal data editing procedures. Only a strategy which succeeds in making optimal use of the respective reporting method's advantages, while also consolidating the data flows from the various reporting media for combined processing at the earliest possible moment, will result in actual savings. This necessitates additional investment in software and alterations in workflows and administration.

15.   Last but not least, as a consequence of such an intervention in established processing and production systems, the employees concerned – be they subject matter or IT specialists – will have to alter procedures which they have developed over time and which have hitherto proved of value. They must carry out new tasks for which they will also need new skills. In this area, aside from concrete training and other measures, top management's explicit commitment will be required in order to overcome possible resistance.

ISSUES, DECISIONS, SOLUTIONS

<u>"IT-experts should only be used when they are really needed"</u>

16.   Many IT solutions not only incur one-off development expenses but also imply considerable further effort for continuous adaptation to new or modified requirements. In many cases, such changes are part of the work environment and are by no means unexpected – e.g. recurrent statistical surveys are subject to periodic revisions of their content, so that some

questions may vanish while others are added. Even so, system design often neglects such foreseeable alterations, necessitating the services of expensive IT personnel.

17.   In the IT field, many steps in the process require specialist programming skills because suitable tools for use by layman users do not (yet) exist.

18.   These considerations influenced us in the design of e-Quest at an early stage. Accordingly, we developed the "e-Quest Metadata Manager" to support the process of survey design and questionnaire creation. Using this application, subject matter experts can create and configure questionnaires by simply dragging and dropping visual objects, and can specify field and group definitions, control flow and validations. The program function "Save as Web form" converts this information (metadata) into a standardized XML format, which is subsequently used to generate the Internet Web form application automatically.

From piecework towards automatization

19.   Statistics Austria carries out a large number of primary surveys for which electronic forms are eminently suitable. For the purpose of lessening the burden on respondents, the Internet reporting option is offered in all such cases.

20.   In the face of shortening production cycles and shrinking budgets, this demand can no longer be fulfilled in the traditional fashion, i.e. by manual software production. Not only do the expenses of developing individual software add to production costs for the paper questionnaires for each individual survey, but providing a timely electronic alternative for new or modified surveys is an almost impossible task. Even when software development is done in-house, thus avoiding delays due to public calls for tenders and subsequent evaluation periods, developing a user-friendly, secure and thoroughly tested electronic alternative takes much longer than the usual production process for paper forms.

21.   We therefore built an "industrial production line" for the e-Quest/Web forms, which allows a complete Internet web form application to be created by simply pressing a button. Not only is the primary Web application itself (Java classes, database tables and JSP pages) generated, but also a corresponding intranet application including additional validation and processing functions for the subject matter experts within Statistics Austria.

22.   This is achieved by specifying the aforementioned metadata during survey preparation, by standardizing the Web questionnaires themselves and by definition of an Internet application type "Statistics Austria Web Forms" as a general logical schema. This schema is comprised of general rules as to structure, flow, software and questionnaire construction, behaviour, security mechanisms, etc.

23.   What were the most important design criteria?
*   high security of input data;
*   no protracted searches in voluminous production code lists and other classifications, but quick retrieval of the correct code instead, aided by display of the codes in hierarchies, full text and synonym search, and automatic insertion of codes, texts and units of measure in the questionnaire;
*   an extensive multi-layered help system offers context-sensitive help only when and where it

is needed;

- automatic input checks mark incorrect or implausible data with warnings and/or error texts;
- automatic calculations (sums, comparisons, etc.);
- questions can be displayed or hidden dynamically, depending on their relevance in the specific case;
- form completion can be interrupted at any time and the data saved for completion at a later time or date;
- third-party respondents have access to their clients' forms and functions in order to manage them.

Security aspects[2]

24.   **Risk avoidance**: one question discussed extensively during the project was whether the respondents should be provided with data from the previous month as initial values or for the purposes of checking and correcting the current data.

25.   To implement such facilities in the case of Web forms, the previously reported data of all enterprises in the sample would have to be stored and kept accessible via the Internet[3] for the duration of the survey. This appeared to us to carry unacceptable risks. The potential damage a successful hacker could do in such a situation far outweighs the benefits.

26.   We therefore adopted a design that aims to expose the data for as short a time as possible to a "hostile environment". The data from the Web forms – transmitted via secure connection – are stored in a Statistics Austria database but are, as soon as possible, transferred from there to another database which is inaccessible from any Web application.

27.   **Security Audit**: an important feature of e-Quest/Web is its global security architecture. Together with an external software partner, we created a design that focuses on secure and confidential communication with the user over the public and a priori untrustworthy Internet. The completed design was analyzed and validated by external experts with regard to mechanisms and application flows relevant to security. This method of procedure was intended to minimize the risk of design defects and also to ensure the confidence of Statistics Austria and the public in its Web questionnaire system. The results of the expertise were worked into the design before commencement of the implementation phase.

28.   **Taking the respondents' security concerns seriously**: we resolved the security concerns of our respondents and ensured the protection of their privacy. For e-Quest/Web, we decided not to base the application on the use of cookies.[4] The validations do not require Javascript to be enabled. However, if the validations are enabled, they will be made immediately upon data input instead of at the next server interaction.

29.   **Gaining trust**: with respect to protection of their privacy, users often fear that Internet applications will spy on their PC, the software they use, or their "surfing" behaviour, and then analyze or use that information in unknown ways. In order to strengthen respondents' confidence in this area, we guarantee in writing that Statistics Austria only routes onwards for processing the information which they have confirmed by pressing the "Send" button, and that no other information of any kind is kept.

30.   If the user has temporarily stored some data on the server in the course of filling in the form, this data is deleted automatically when the final version is confirmed. Such data is also deleted after a specified time interval if the user afterwards decides against transmitting the data electronically. The temporary data on the server can also be deleted explicitly by the user pressing the "Reset" button.

Respondent-side editing

31.   One aspect that at first glance should improve the quality of data and thus lead to higher efficiency is the shifting of responsibility for data editing to the respondent. It must be noted, however, that this goal does to some extent contradict the expressly formulated principle of "lightening the burden". Any errors the respondent is alerted to while entering data result in an immediate increase in time spent reading the error, determining what is wrong and correcting it (and at that moment, the respondent is not conscious of the time which might otherwise be spent in reacting to enquiries from Statistics Austria about erroneous data). We therefore considered this area carefully during the design phase. The subsequent decisions and their results are described in detail in the following paragraphs.

32.   Single **field checks**: the validation of single data values is controlled wholly by the data field attributes specified during questionnaire definition, such as data type, length restrictions and arithmetic limits. The checks can be classified according to their urgency and importance. Urgency:
- check immediately after the data value was entered,
- check when the user says so,
- check when the whole questionnaire is being validated.

Importance:
- must be corrected in all cases,
- important warning which may, in some cases, be ignored,
- simple warning,
- comment for the subject matter expert.

33.   "Immediately to be checked" validations are only really checked right away if the user has enabled Javascript in the Internet browser. If not, they are carried out on the Web server at the time of the next server interaction and then displayed, complete with error messages and markings to the user for correction.

34.   **Inter-field validations** check rules that describe the relationships of field values to each other. Thus, they always result from a separately specified condition. The electronic version of the Short Term Survey includes about 50 such validation rules, i.e. about half of those checked during subsequent processing. Very few of these are classified as "must be corrected" (which prohibits sending the data before such errors are corrected). In e-Quest, a further error level has been introduced in which the respondent can insist on a value marked with a severe error if he or she adds a comment as to why the value is correct in this special case.

35.   **Inter-questionnaire validations** (checks of the relationships between data values in more than one logically connected questionnaire, such as questionnaires for an enterprise and its separate establishments) were not realized by us in e-Quest/Web, since a Web browser application is not really suitable for displaying and manipulating concurrent forms.

36.   **Abstinence from checks wherever feasible** : as correcting errors causes increased effort for the respondents, type, level and number of validations must be weighed carefully. Some errors may be easily corrected automatically or by subject matter experts and thus could be ignored in the data input phase. From the respondent's point of view, the questionnaire will then be easier to use.[5]

Benefits for the data collector

37.   Significant savings through electronic questionnaires can only be expected for survey data processing, when the relationship of electronic to paper-based responses is relatively high and/or the new data track triggers a climate of change in the traditional methods of processing. An infrastructure encompassing all phases of data collection is needed, and the new processes must be integrated into existing processing systems. As a rule, the latter have developed over time and cannot be simply thrown overboard.

38.   In parallel with the aforementioned e-Quest Metadata Manager for survey preparation and the questionnaire systems e-Quest and e-Quest/Web on the respondents' side, Statistics Austria developed standardized software systems to accept and process the raw data. Incoming e-Quest/Web questionnaires can be directly viewed and edited in the intranet by subject matter experts using the same browser application developed for the respondents – albeit enhanced by specific internal functions. Alternatively, the data can be transmitted to the e-Quest Package Manager (internally nicknamed the "Pot Application") created for e-Quest responses. This is done for the Short Term Survey.

39.   The "Pot Application" is a standardized tool for processing electronic responses. Incoming questionnaires (in the form of XML packages) are automatically distributed into various processing "pots" according to their content, using criteria defined in advance. In the "pots" they can be viewed, edited, corrected and passed on to other subject matter experts for further processing. An enhanced version of e-Quest can be used to view the data, which keeps track of any changes to the data, maintaining a history of modifications and comments attributable to the (internal) user who made them. Questionnaires which have been checked are moved to the "OK pot" and are then transferred via an interface component[6] to other systems.

40.   An efficient integration of electronic questionnaires into the existing processing flows was achieved with the PC correction application "KjeKorr". This is a program that allows data from the monthly Short Term Survey "Production", already residing in the host database, to be viewed, augmented, validated and corrected. It too fulfils all Statistics Austria's requirements for extensive documentation and modification history, which used to be contained in paper form. Without having to consult the paper questionnaires, the subject matter expert can now process the survey data quickly and efficiently and can rely on automatic data editing support:
- automatic calculations (sums, quotas, entry counters, units of measure);
- online search in the classification hierarchies (PRODCOM, NACE);
- modification history, comments and validation checks on the currently active field are highlighted;
- questionnaires that have been "set aside" or deleted can be reactivated;
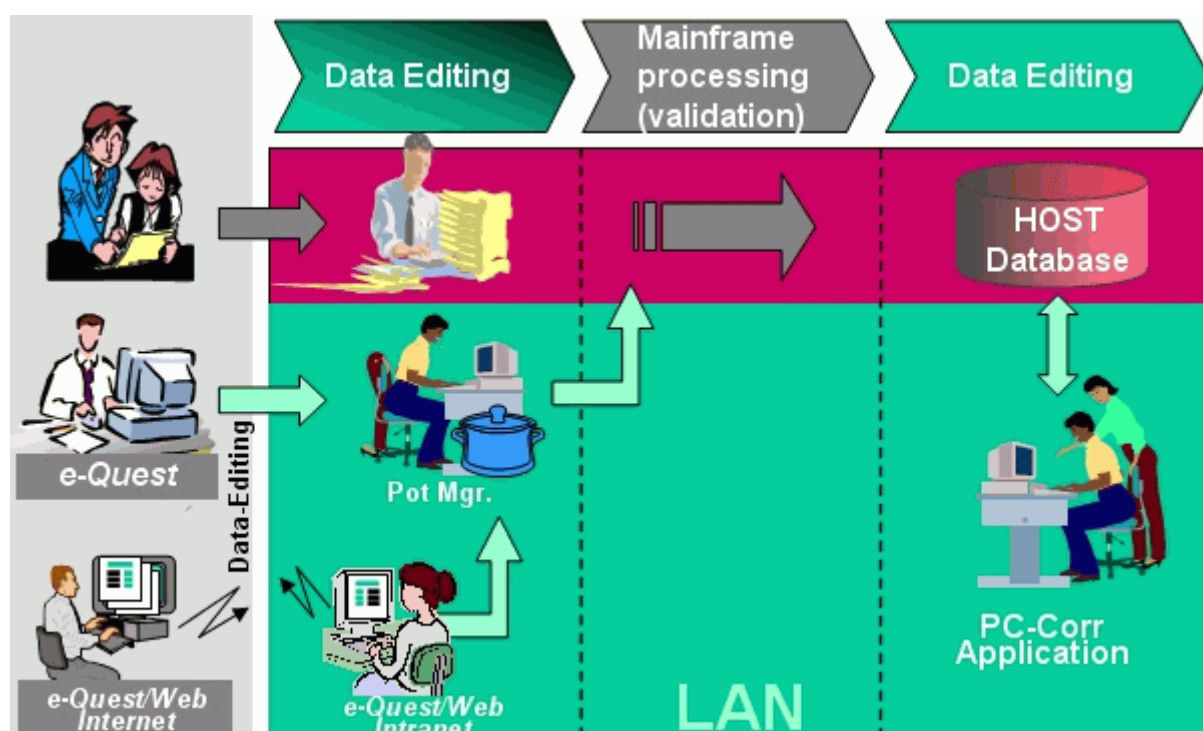- automatic data validation when the data are stored.

Figure 2: Integration

AN EXAMPLE OF USE: THE SHORT TERM SURVEY "PRODUCTION"

41.   The Short Term Survey "Production" delivers the foundation for complying with the EU regulation on short-term economic statistics. In addition, it allows for the calculation of the national production of goods.

42.   Approximately 12,000 questionnaires are sent out monthly. Of these, 10,500 are directed to enterprises with one establishment, 400 to multi-establishment enterprises and about 1,200 to the establishments of the latter. The questionnaire is pre-filled with the respondent's address and some characteristics obtained from Statistics Austria's enterprise register.

43.   In total, ten different variants of the questionnaire are used, depending on the enterprise's size and type. The questionnaire for single-unit enterprises contains about six pages with on average 90 questions, grouped into blocks by topic (employees, hours worked, salaries, orders, turnover and production). Multi-establishment enterprises must fill in three pages for the enterprise as a whole (employees, salaries, turnover) and six pages per establishment. About 25 pages of explanation and clarification accompany the paper questionnaires. Many logical dependencies exist between questions, both within the same block and between blocks (e.g. if some employees only work part-time, the hours worked must be stated separately for full- and part-time employees).

44.   At present, about 2,600 (22%) of the 12,000 monthly questionnaires are reported via e-Quest. The use of e-Quest correlates with the size of the enterprise. While in June 2004 only 12.3% of companies with up to 20 employees made use of this reporting track, the quota for the largest enterprises was 46.6%.

45. Starting with the reporting month January 2005, single-unit enterprises have been offered e-Quest/Web as a new alternative, though without explicit public relations measures, except for a letter included with the paper forms. Of the questionnaires received through to 8 March (altogether about 7,000), 9% made use of this new questionnaire, against 69% on paper and 22% via e Quest. The number of Web respondents is thus roughly equal to the number of e-Quest respondents when that application was first introduced.

46. A first reporting stream analysis, based only on the questionnaires received up to 8 March compared with those received up to October 2004, shows that, of the 1753 e-Quest users in this sample in October 2004, 299 have moved to the Web questionnaire. Of the 5194 paper users in October 2004, 292 used the Web form this time and 88 used e-Quest. Of 64 users new to the survey, 8 used e-Quest and 8 the Web track. These numbers offer only a first impression, but they confirm our assumption that the respondents make deliberate decisions as to what relative importance they attach to ease of use vs. security concerns.

SUMMARY

47. In 1996, 50 employees were entrusted with processing the Short Term Survey from the subject matter point of view. Today, they number only 30, have been equipped with PCs and trained in the use of the new products. All these employees process both paper and electronic reports.

48. At the start of the e-Quest project in 2000, many of our colleagues – in both enterprise statistics and IT departments – questioned the sense of this undertaking and showed considerable resistance. On the one hand, they doubted that an electronic solution for such a complex and voluminous survey would be accepted by respondents. On the other hand, many feared a deterioration in the work environment. It seemed inconceivable that a workflow which required several paper questionnaires to be laid side by side in order to effect the necessary comparisons and corrections, could ever be represented on a comparatively small screen.

49. Both fears proved unfounded. In the second case, the correction application "KjeKorr" for the first time in Statistics Austria shifted data editing from the mainframe to the PC level. It boasts a graphical user interface and its functions are oriented directly to the users' needs and wishes. In consequence, this new processing tool has been widely accepted.

50.   As well as being used for the Short Term Survey "Production", e-Quest has supported the Short term Survey "Services", the Structural Business Survey, the Production Input Survey and the Work Costs Survey for several years.

51.   In the few weeks it has been in production, e-Quest/Web has already been used in additional surveys such as the Livestock Survey and an ad hoc questionnaire for civil registry offices. Further e-Quest/Web questionnaires are under construction. The speed and relatively minor effort with which these questionnaires can now be produced prove that the concept of e-Quest and e-Quest/Web was correct and makes economic sense.

[1] *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 questionnaires and for managing and processing the response data received. 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.

[2] This section does not deal with general measures to ensure data security and confidentiality, but is only concerned with those aspects that seemed of especial importance to the authors during the project's implementation.

[3] Albeit only accessible through the Web application.

[4] Except for a non-permanent session cookie, which the Web server requires in order to associate incoming requests with the correct session.

[5] We must not forget that, in our case, the respondents are required by law to provide the information but are not necessarily happy about it.

[6] The interface component is programmed for the specific survey, but conforms to a standard (COM) interface, and can thus be added to the system with minimal effort.

* * * * *