

Electronic data collection system eSTAT in Statistics Estonia: authentication, security and confidentiality issues

Emilie Abel, Statistics Estonia

Statistics Estonia and statistical production process

The main task of Statistics Estonia is to produce and disseminate objective official statistics for public institutions, business and research spheres, international organisations and individuals. For collecting data, official statistical surveys are conducted. A survey means the whole process of producing statistics of a certain field - the collection, processing and dissemination of statistical data. About 150 surveys are conducted each year; in addition, special orders for foreign users of statistics are fulfilled. Data is collected from respondents by about 200 questionnaires. The number of respondents among economic entities is about 60 000. One respondent has to submit regularly about two-three questionnaires, but mostly one, in very rare cases, about 30.

Electronic collection of data from enterprises and state agencies (economic entities) in Statistics Estonia

The objectives of an electronic data collection are: 1) unified data collection process of different surveys (standardized questionnaires and tools); 2) higher data quality (automatic check of rules); 3) shorter data processing time for both Statistics Estonia and the respondent (accessibility). Another objective is to use more modern means of data collection that will contribute to improving the image of Statistics Estonia.

At present, there are two main channels of electronic data collection:

1. Uploading or sending manually completed questionnaires in Excel format or plain text files by e-mail;
2. Manually completed web questionnaires in eSTAT. This presentation is focused on eSTAT.

In 2007, a third channel will be taken into use - uploading via eSTAT or directing data transfer from XML files (EDICOM project).

There is also a minor amount of data that is received from registers¹.

Introduction of eSTAT

eSTAT is an electronic channel for collecting statistical data from economic entities. eSTAT was developed in 2004-2005. It was implemented in September 2005 for internal use and in February 2006 for public use. In the year 2007, 50% of all questionnaires can be completed in eSTAT². In the future eSTAT will be the main channel of data collection (80% of all questionnaires for economic entities should be collected via eSTAT web forms or XML transmission).

The main functionality of eSTAT for external users (respondents) use is:

1. A list of questionnaires with submission dead-lines;
2. Web questionnaires for data submission;

¹ For example customs declarations

² 15% in 2006; only 2% in 2005.

3. An automatic checking rules during questionnaire completion;
4. Access to previously completed questionnaires (two years backward).

Public users are mainly managers, accountants and personnel executives of economic entities.

The functionality for internal users is:

1. Questionnaire designer;
2. Respondent management: list of respondents and automatic e-mailing to economic entities who have not completed the questionnaires;
3. Query designer to view questionnaire completion statistics by different indicators of economic entities (for instance activity or region);
4. Possibility to view a complete overview of customer contacts made during data collection process of different questionnaires;
5. Data transmission management from raw database to validation and aggregation databases.

Internal users are mainly data collectors and statisticians.

Submitted raw data with associated metadata is saved to a raw database.

Authentication and confidentiality in eSTAT

Internal users need a username and password for authentication. Each user has restricted rights to view or complete different questionnaires. Passwords are held in the user management database and are encrypted with a cryptographic MD5 (Message Digest) algorithm.

External users need one of the following identification documents:

1. Personal ID card;
2. Internet bank card (of one of Estonia's three biggest banks, i.e. Hansabank, SEB Eesti Ühispank or Sampo Bank).

The above authentication methods are widely used for services in different portals in Estonia (tax returns, Internet banks).

Only an individual can be a user of eSTAT. The head of a respondent enterprise or agency delegates the rights to view and complete web questionnaires to an employee of his or her enterprise or agency. The delegation application can only be submitted by persons who's right to representation of the enterprise is registered in Estonia's Commercial Register. That guarantees access only to authorized persons. Application for delegation could be sent by e-mail (with digital signature) or by mail. The right of representation is verified in case of every application. The delegated main user has the rights to include other employees in the list of data suppliers. These rights could be restricted by different questionnaires that users can see and/or complete.

There were about 3200 main users and about 1000 data suppliers delegated by different main users in the beginning of October 2006.

The main problem with authentication:

1. Usage of individual ID cards or Internet banking cards has caused some confusion and displeasure among very few respondents who are mostly afraid

of possible access to their personal information or bank account. Such complaints are caused mostly by confusion and lack of information. In fact, those methods are used only for authentication service and Statistics Estonia has no access at all to any personal data of the users.

Authentication procedure is implemented using iPizza protocol. iPizza is a standard authentication protocol that is used in communication services with Internet banks. In case of successful identification only a authentication identifier (citizen's unique personal code) is transmitted to eSTAT. Communication from Internet bank to eSTAT is encrypted with OpenSSL protocol.

Data security

Data transmission over the Internet or Intranet is via an https protocol. The inner user interface can be accessed only within a restricted range of IP addresses to prevent the access and possible password selection from unauthorized outside users.

Data managers have direct access to the database. Data managers are responsible for special queries, data import and export in case of need, data correction and metadata administration. The number of users is limited (around four) and the procedures are monitored.

Raw data transmission to validation and aggregation databases is managed using the iWay ETL Data Migrator tool. Transmissions have a frequency of one hour to one day depending on the questionnaire.

Data integrity

Database administrators and a limited number of data managers have direct access to the raw database. This has caused some data integrity problems at the beginning of implementation. Personalised database accounts are used at the moment to prevent or to track the causes. Transactions to very important tables are logged and could be roll-backed. All tables have audit columns to track database record lifetime (in most tables traceability level two, where time and user or process responsible for modification are logged).

Backups of the database are made daily at night time. Oracle Recovery Manager enables to rollback backup from the previous day and transactions made after backup in case of the corruption of the database. Backups of the application server are made weekly at night time.

Availability

There have been and still are some availability problems due to different infrastructure problems and a continuous development of new modules. According to service level agreement, availability (maximum reparation time is four hours) is guaranteed weekdays between 8 am and 6 pm. During that time support helpline is available by phone and e-mail.

There has not been any huge overload of the system yet. A version of software that runs in a clustered environment is already under development and will be implemented in the near future when the number of users will probably have increased. Availability problems close to submission deadlines could affect data collection productivity indicators and indirectly even data aggregation and dissemination deadlines. Inner and external users should be informed at least 48 hours before of any planned interrupt of the system.

Additional information

eSTAT was declared the best innovation in Estonian public administration in 2006 in September 2006 by the Secretary of State. There has been a positive feedback from Estonian accountants. According to them, eSTAT has simplified data submission to Statistics Estonia.

Technical details of the system

Application Server: Linux server, BEA Web logic 8.2;

DBMS: Oracle Database 9i;

Questionnaire designer: a client-server program in Java;

Questionnaire presentation: Java Server Page files, checking rules in xslt files.