



**Economic and Social
Council**

Distr.
GENERAL

CES/AC.71/2001/29
31 January 2001

ENGLISH ONLY

STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

COMMISSION OF THE EUROPEAN
COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE/Eurostat Meeting on the Management of Statistical Information Technology
(Geneva, Switzerland, 14-16 February 2001)

Topic (iv): Integration of statistical (survey) data with registers (administrative) data

THE KOSTRA¹ PROJECT

Submitted by Statistics Norway²

CONTRIBUTED PAPER

I. INTRODUCTION

1. The Kostra project was launched in 1994, as a response to a Parliament Report discussing the relationship between the local and central government administrations. In 1995 a project group led by Statistics Norway was established to carry out the first pilot project within the framework of Kostra. Statistics Norway has since then had the leading role in the project, which is sponsored by the Ministry of Local Government and Regional Development.

2. The project objectives are:

¹ KOSTRA is a Norwegian acronym for "Direct reporting of data from Local to Central Government Administrations".

² Prepared by Rune Gloersen.

- ◆ to enhance the quality and consistency of data and statistics on the use of resources in the local government administrations (i.e. the municipalities) in Norway, and to ensure better comparability between information from different local administrations;
- ◆ to collect, compile and disseminate statistical information in this field with less resource consumption in terms of money and manpower.

3. To achieve this, the following activities are carried out through a comprehensive development project involving all municipalities and several governmental administrations. The work within some of the activities has been redefined due to the experiences gained over the past years. The listing below is currently to be read as a summary of ongoing and finished activities within the project:

- ◆ to redefine the structure of the accounting system in the municipalities, including classifications and coding, in order to make better relationship between the accounts and the services performed by the municipalities;
- ◆ to provide an IT network defining rules and formats for the collection of data, supported by third party software to be installed in the municipalities;
- ◆ to derive accounting data directly from the different accounting systems;
- ◆ to develop and distribute electronic questionnaires to be filled in for collection of data on services;
- ◆ to develop a database to store and administer the data received, acting also as the master file for all data during decentralised data editing within Statistics Norway;
- ◆ to establish methods and procedures for data editing both horizontally and longitudinally, across different sectors or statistical domains and between different subject matter units;
- ◆ to disseminate key figures once the data are received, and to disseminate final results not longer than five and a half months after the end of the reporting year.

II. THE PARTICIPANTS IN THE PROJECT

4. There are currently 434 municipalities in Norway, clustered in 19 counties. For the reporting year of 1996, the Kostra project comprised 4 municipalities as pilots for testing possible technology and architecture of the future system. The most important task from the beginning has been the necessity to change the structure and contents of the local accounting systems.

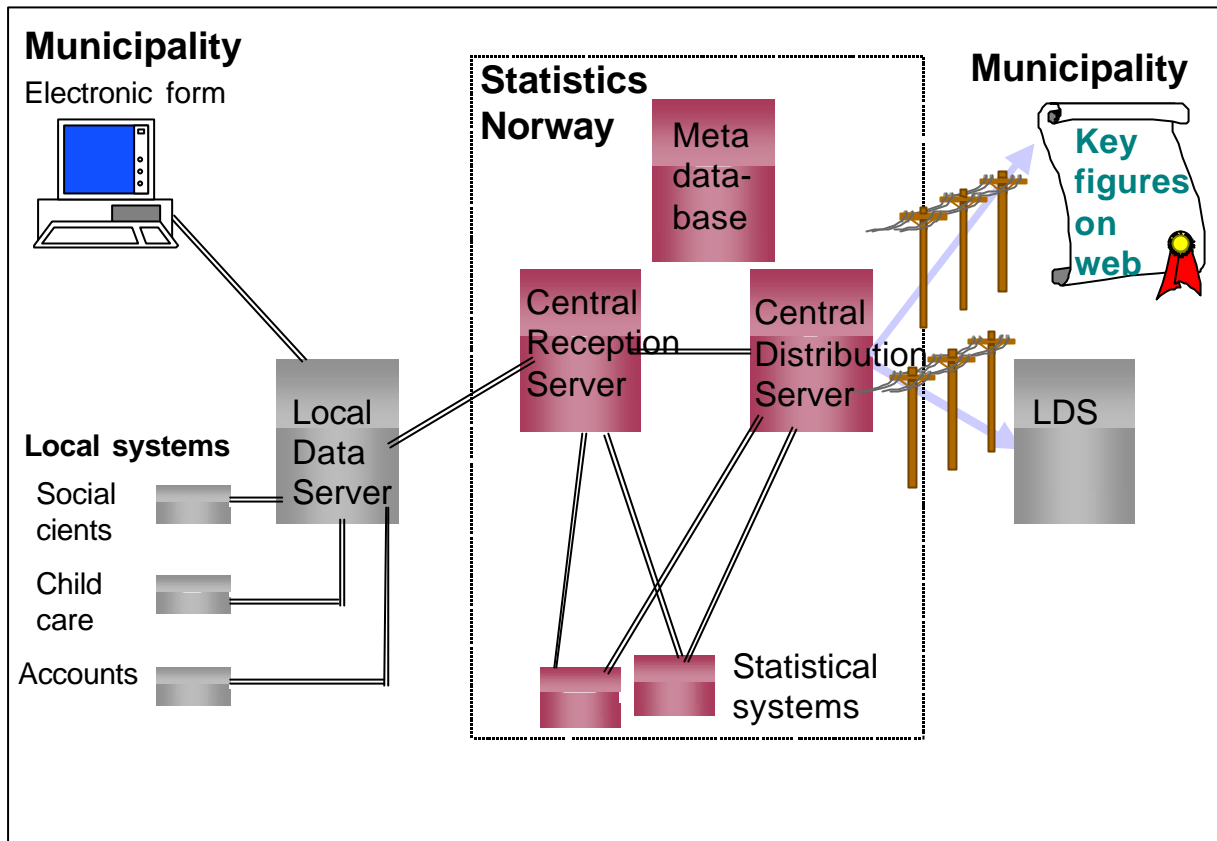
5. For the reporting year of 1998, 25 municipalities and 1 county took part in the reporting. For 1999 the number of participants was increased to 108 municipalities and 5 counties, and for 2000 the numbers are 220 municipalities and 7 counties. From the reporting year of 2001 all municipalities and counties will be reporting due to the new accounting rules, and by electronic means.

6. The users of Kostra data are the municipalities themselves, the ministries, the county level administrations and other bodies involved in this sphere. Not to mention any public user interested in key figures on municipal performance.

III. THE KOSTRA INFRASTRUCTURE

7. During the period of development, the infrastructure of Kostra, that is the use of electronic questionnaires, the direct reporting from local administrative systems and the systems for receiving and administering data will become a common platform for electronic data collection within Statistics Norway.

III.1 Data reception



7. Data are collected using electronic questionnaires and as file extracts from different administrative computer systems within the local administrations. Ideally, data should be loaded into a local database (the Local Data Server "LDS") in each municipality. So far, not many municipalities have this LDS. At respondents' side, data are stored in EDIFACT format using the RDRMES messages directly from the electronic form or from extracts from the computer systems. The messages are then sent to Statistics Norway as attachments to E-mail using EDIFACT converter software. Sensitive data on social clients and child-care are encrypted before they are transmitted.

8. The correct versions of the different questionnaires can be downloaded from Internet. Statistics Norway has specified these questionnaires, while the task of questionnaire design has been outsourced. The metadatabase contains all questionnaires and all variables. The long-term objective is that the questionnaires should be automatically generated from the metadatabase (see chapter 4).

9. The local data should be submitted to the central server at Statistics Norway, the Central Reception Server (CRS) before February 15th every year.

10. Arriving at the CRS, data is checked for errors and then stored in an Oracle database. Receipts are returned to sender and web pages showing status for received data from all municipalities are updated. These pages also show error messages if something is wrong.

III.2 Support and maintenance

11. During the reporting period, a help desk is established serving the respondents on both technical and subject matter issues. There are systems developed to keep logs of all incoming messages and for analysing them. There are also systems developed for following up all incoming reports regarding status, error messages, missing reports and so on.

12. Today metadata maintenance is done centrally by IT-persons. This is time consuming, especially during the period of designing forms. A more user-friendly interface for the metadatabase will be developed, in order to decentralise work to the subject matter units. When Kostra reporting is put into operation in 2002, this will be increasingly important.

III.3 Data editing

13. All data received from local administrations are, as mentioned above, stored in one Oracle database in Statistics Norway. The collected data covers a number of statistical domains. The data providers can change these data by sending corrected data when errors are detected either by Statistics Norway or by themselves.

14. All data received before March 1st are processed automatically, and published as preliminary figures March 15th year t+1.

15. At an early stage, Statistics Norway defined a clear goal to establish a common architecture for data editing, enabling more comprehensive cross sectional checks, and remaining the micro data in one database as a master file. It has been quite a challenge for Statistics Norway to establish rules and systems allowing parallel editing across subject matter unit boundaries.

16. Each variable within the Kostra system has been designated an owner in terms of responsible subject matter unit. For each questionnaire (in fact for each data flow), there has been developed a targeted editing system, based on an overall architecture. The data for a specific questionnaire are loaded into a (temporary) database, where a set of predefined checks are carried out in a batch system, labelling records that contains errors. Experts in statistical methods together with subject matter specialists have designed checks and imputation rules. Since Kostra is a full-scale survey, it is relatively easy to detect errors or cross sectional inconsistency, while it is more difficult to impute correctly. When data editing is finished, the cleaned data are transferred back to the CRS.

17. Based on cleaned micro data, final figures are published 15th of June, year t+1.

III.4 Publishing

18. The data ready for dissemination are loaded into the Central Distribution Server (CDS). Via a web interface data is published as key figures (for the moment only in Norwegian).

0403 Hamar: Barnehager - nivå 3
Versjon av 07.07.2000

Barnehager - nivå 3
[\[Sett startkommune\]](#) [\[Velg egne kommuner\]](#)

	Hamar	Gruppe 10	Hedmark	Landet ut/Oslo
Funksjon 201 Opphold og stimulering av førskolebarn				
Tjenester: Barnehager, basert på grunnbemanning (basistilbud), ordinære driftsutgifter, barneparker				
Ressursbruk				
Brutto driftsutgifter (i 1000 kroner)	56 003	42 182	26 151	21 678
Korrigerte brutto driftsutgifter (i 1000 kroner)	44 005	29 906	21 081	15 920
Brutto driftsinntekter (i 1000 kroner)	45 353	34 635	20 435	17 047
Inntekter fra ordinært statsliskudd (i 1000 kroner)	25 122	19 108	9 478	8 707
Inntekter fra foreldrebetaling (i 1000 kroner)	19 302	12 881	8 936	6 343
Overføring tilskudd til private barnehager (i 1000 kroner)	10 276	10 865	10 276	6 120
Kjøp av private barnehageplasser (i 1000 kroner)	1 642	1 748	845	2 316
Antall årsverk knyttet til basisvirksomheten i alle barnehagene				
Styrere	20,45	24,62	13,49	11,50
Styrerassistenter	0,40	2,69	0,40	2,32
Pedagogiske ledere	50,65	45,85	24,78	24,00
Assisterende	151,48	122,56	68,30	60,44
Antall årsverk knyttet til basisvirksomheten i kommunale barnehager				
Styrere	19,00	9,80	8,00	5,44

19. The web site has got functionality for selecting municipality and one or more municipalities to be compared with (a group of municipalities). You can also select area of statistics among several predefined tables. Data can be downloaded in Excel format. This year we will also provide functionality for more dynamic queries and downloading using PC-Axis software (Px-Web).

20. As mentioned above, preliminary figures will be published March 15th. No manual checks will be carried out on these data. To some extent, obvious errors in the figures will be flagged. Final figures will be published June 15th. This is approximately a year faster than before the Kostra project.

III.5 Locally - in the municipality

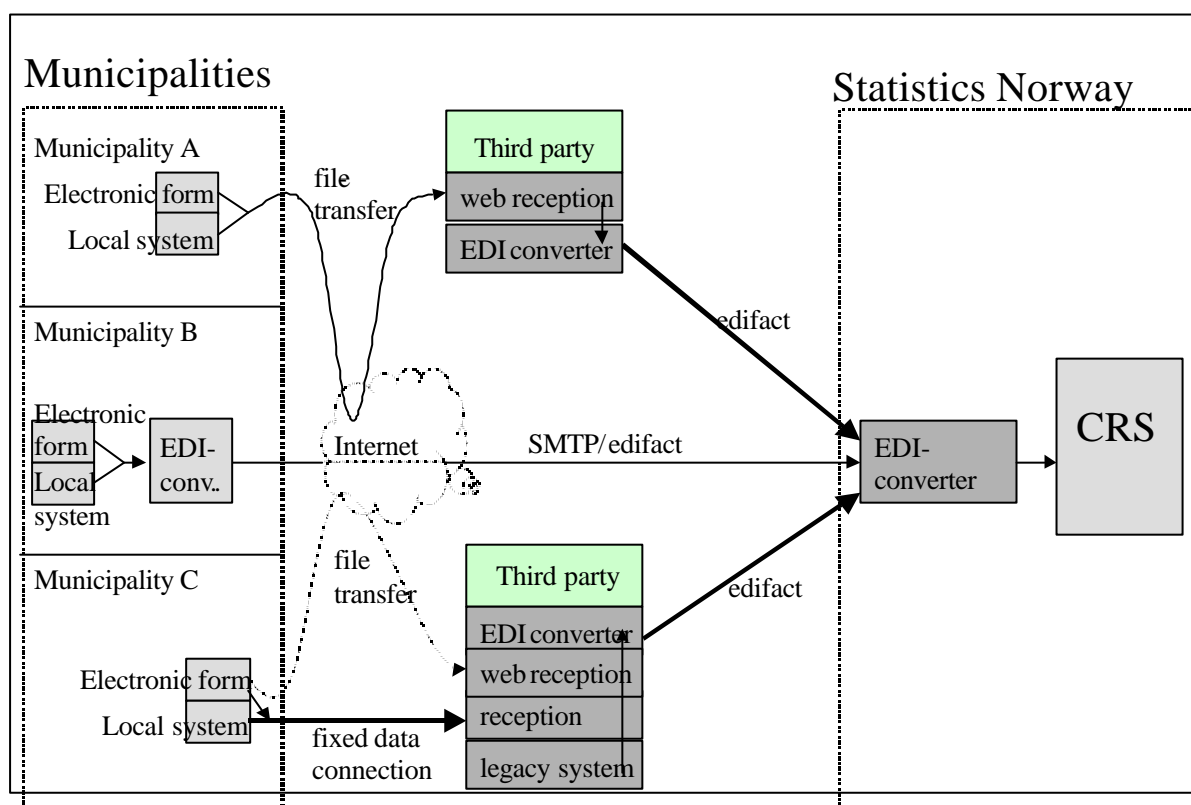
21. Data extracts from different sources in the municipalities are converted into EDIFACT RDRMES format and sent via SMTP to Statistics Norway. At the moment we only receive file extracts as general lodger from accounting systems, client data from system for child care and client data from system for social support.

22. For those areas that are not covered by file extracts, we use electronic forms (questionnaires). These forms are filled in locally in each subject department and sent to a central Kostra administrator in each municipality. For 1999 and 2000 we have used Form Flow to develop the electronic forms.

23. The Kostra administrator in each municipality is responsible for collecting all information both from file extracts and from the electronic forms and transfer data to Statistics Norway.

24. There are two fundamentally different ways of transferring data:

- ◆ one is via a locally installed EDI-package which handles both the conversion to EDIFACT and file transportation to Statistics Norway (shown as municipality B in the figure);
- ◆ the other is to buy this service from a third party. In this case the third party will do the EDIFACT conversion and the data interchange with Statistics Norway (alternative A). A lot of municipalities have since many years been running some of their legacy systems at a service bureau. For those municipalities it will be convenient to use intermediaries also to provide EDI conversion (shown as alternative C in the figure).



IV. FUTURE DEVELOPMENT

25. During the project we have experienced that any system that requires installation of software at the respondents, cause problems. The willingness to pay is also rather low. Furthermore, if the software requires IT personnel to be installed or operated, this expertise can not always be expected to be present. The EDIFACT solution, which is based on widely accepted standards with regards to format and concepts, is not supported by widely known or commonly used software. A lot of efforts have been spent until now on installation, training and support of the technology. Moreover, since the reporting is done only once a year, quite naturally problems seem to reoccur because of loss of memory.

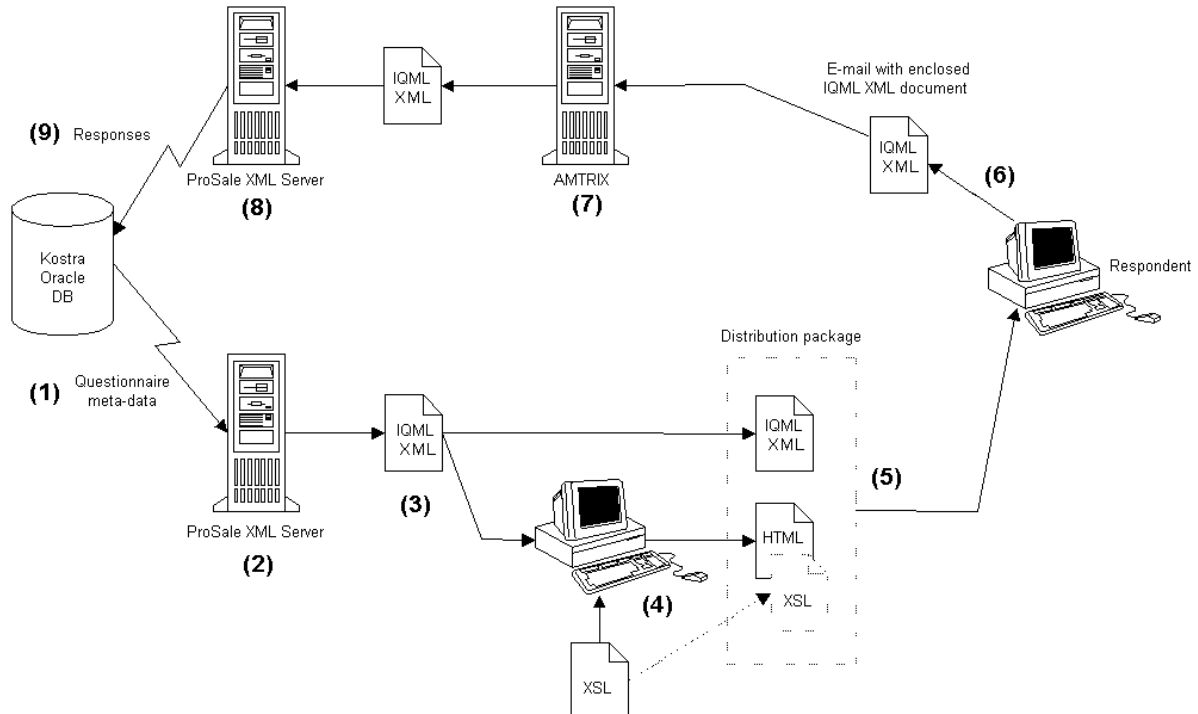
26. The maintenance cost of questionnaires is somewhat too high. When questionnaires are changed the content and layout is changed using Form Flow. Then the questionnaires have to be tested, and the flow of questionnaire data through each component of the system must be tested if variables have been added or deleted.

27. New technology under development seems to be able to solve some of the problems we are currently facing. The objectives of our future developments are that:

- ◆ installation of third party software should not be necessary or at least be kept to a minimum;
- ◆ installation and use of electronic questionnaires must not require any kind of IT expertise, just the skills of an average user of IT equipment and software;
- ◆ the software needed at respondent side should be offered free of charges;
- ◆ the cost of software distribution should not be subject to third party license fees for the collector;
- ◆ questionnaires should be generated from a metadatabase, to ease the process of creation or modification.

28. We are about to start a pilot project where we will use software developed under the framework the IQML (Intelligent Questionnaire Markup Language) project. This is a project co-financed from the 5th Framework for Research and Development in the EU. Statistics Norway is one of the participants.

29. The figure below gives an overview of the reporting cycle as intended in the pilot project. The numbers refers to components of activities shortly described in the following. Stage 1 to 5 is the creation and distribution of electronic questionnaires, while stage 6 to 9 is the handling of response and upload of data.



30. Explanation of the figure:

- (1) This is the metadatabase designed and developed in the Kostra project. The database contains all information needed to design questionnaires, including rules on checks, calculations etc.
- (2) The ProSale XML server connects to the database and extract data relevant to each questionnaire. This is then transferred into an (IQML) XML document.
- (3) The IQML XML document follows schema rules defined in the IQML project.
- (4) An XSL style sheet has been developed to visualise the questionnaire. To use (view and respond) the questionnaire in a browser, the style sheet can be used directly with the XML file, or the style sheet can be used to produce an HTML file. Which way to use it depends on the type of browser that is to be used. In the pilot, all questionnaires will be distributed as HTML files.
- (5) The XML file, the HTML file and some additional software are distributed to the users, for instance via e-mail, as one package.
- (6) The respondent saves the distribution package. Opening the corresponding HTML file in Internet Explorer activates the questionnaire. If there is data to be pre-filled in, this is done when opening the questionnaire. During fill-in, checks and calculations are done according to the respondents navigation throughout the questionnaire. The completed questionnaire is saved to disk, and supplemented with the necessary administrative data required when received in Statistics Norway.
- (7) Statistics Norway will receive the questionnaires as attachments to e-mail sent by the respondent. We use the same software to receive and store the attached XML questionnaires as we use for the EDIFACT messages.
- (8) ProSale XML Server monitors the folder of incoming questionnaires, and uploads them into the Central Reception Server (CRS).
- (9) Data are verified in the CRS and then distributed to the subject matter databases.

31. The same approach and the same software can be used to run the questionnaires on a web server in Statistics Norway, leaving no need for the respondent to install any kind of software on his own PC.

32. The next phase of development will be to provide tools for the respondent to connect the questionnaires to his own databases, using a Database Interrogation tool. By doing this, he can for instance connect dynamically to a specific monthly report from his accounting system, calculating directly the figures that should be filled in and sent back to Statistics Norway. This would bring us close to a fully automated procedure for data collection, dramatically lowering the burden on the respondents.

33. If we succeed, we will have created a basic infrastructure for on-line electronic data collection using electronic questionnaires on web. Furthermore we can also provide off-line use of electronic questionnaires where the main purpose of distinguishing between these two methods is that in the latter case the questionnaire should be connected to internal data sources of the municipality, or the company. Most important being the fact that both methods are managed and controlled from one source for generation of questionnaires, and by one receiving point for incoming data.