Statistics Norway

Statistisk sentralbyrå

Hans Viggo Sæbø, Rune Gløersen and Dagfinn Sve

Documents

Electronic Data Collection in Statistics Norway

CONTENTS

Summary 3

1. Electronic data collection in an overall strategic setting 5

2. Main data collection developments in Statistics Norway 6

    2.1 KOSTRA 6

    2.2 IDUN 8

    2.3 Other 9

3. Response burden and electronic data collection 9

4. Metadata 10

5. Data security 11

6. Organisational aspects 11

7. Future prospects 12

References 13

Recent publications in the series Documents 14

This paper has been written as a contribution for the International Association of Official Statistics (IAOS) Conference in London, August 27 - 29, 2002

Summary

Electronic data and information exchange is a major issue in Statistics Norway's new strategic plan. The purpose of electronic data collection is to minimise the response burden and to improve quality of data and efficiency of data collection in the institution. Statistics Norway is already involved in several large projects in this field. Data from all municipalities and counties are reported electronically from this year, and work is underway to establish an infrastructure that will enable us to offer all companies the possibility of reporting by the Internet. During the last years the Internet has become the main channel for disseminating statistics, and since data providers also are users of statistics, communication and feedback have come into focus. A relative large part of the statistics in the Nordic countries is based on administrative registers, and the work on developing electronic data collection also includes direct transfer from data systems of the data providers (such as accounting systems and technical systems). Other important issues in our work are how data suppliers perceive electronic questionnaires, the handling of metadata in data collection, data security and organisational aspects of data collection (for example how to treat data collected in different ways effectively). Considerations on more technical aspects such as data formats are included in the paper, but emphasis is put on strategic issues.

1. Electronic data collection in an overall strategic setting

In the strategic plan it is stated that Statistics Norway shall change the content of statistics and production methods in line with the development of the society. Administrative registers are the dominating data source for official statistics in Norway today, and increasing the use of these is an important part of the strategy for data collection. But electronic data and information exchange is a major issue in the strategic plan as well. It affects all quality aspects of statistics. In particular, we expect it to have visible and positive implications on:

- Response burden (costs for society)
- Efficiency (costs for Statistics Norway)
- Timeliness of statistics

Box 1 shows some key figures on data collection in Statistics Norway.

Box 1. Key figures on data collection in Statistics Norway[1]

Sources
– 60 administrative registers
– 350 000 questionnaires (per year)
– 150-200 000 persons interviewed (per year)
Response burden (2001)
- 165 manyear
- Of which 90 for businesses which is about 2 percent of the total response burden for businesses
Response rates (2001)
- Compulsory surveys: 95 percent
- Voluntary surveys: 72 percent

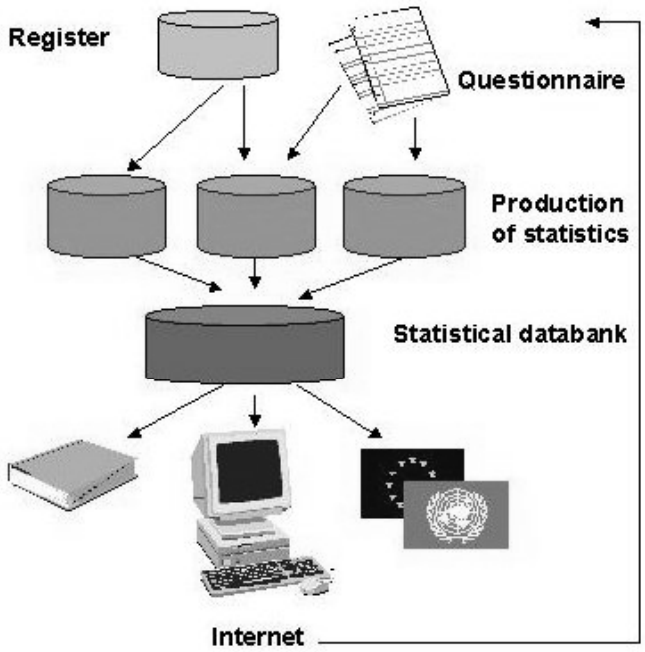
The strategic plan states that: "In total the resources used on reporting to Statistics Norway are modest, but the response burden is unequally distributed and may be perceived as a nuisance by many. The aim is to reduce this by the use of administrative registers instead of direct data collection, motivation of the respondents, improvement of questionnaires, co-ordination of samples and by offering the possibility to report electronically. Feedback where one's own data are compared with other statistics is an example that adds value to statistics and motivates data suppliers."

It also states that "it is an objective that as much of the data reporting as possible shall be done by direct extraction from the data systems of the data providers. There shall be an opportunity for everyone to report over the Internet."

With the Internet the dissemination of statistics has undergone a revolution during the last 5 - 6 years. During the next few years it is expected that data collection will undergo corresponding changes. Electronic data collection should promote a reduction of both absolute (measured in time as in box 1) and perceived response burden, by enabling automatic data collection from administrative data systems within the companies and by motivating the data providers by for example providing them with relevant statistics. Figure 1 is a simplified model of the data flow through a typical national statistical institute. The arrow between the final product and the users of statistics and the data providers illustrates the feedback.

[1] Excluding the Population and Housing Census 2001

Figure 1. Strategy for data flow in Statistics Norway



2. Main data collection developments in Statistics Norway

Statistics Norway works with data collection development projects both in the public and the private sector. The institution is already using computer assisted interviewing (CAI), with electronic transfer of data from its interviewers spread over the country to the main office.

Development work in this area is mainly linked to two major projects comprising electronic data reporting from external agencies such as municipalities and companies. These two projects are called KOSTRA[2] and IDUN[3]. The paper will give a short overview of data collection in both these projects. The technical aspects of the projects are considered in some more detail in Hopland (2002). See also Ljones and Svinnset (2001) for a description of all parts of KOSTRA. Bergstrøm (2002) has also described our approaches to data exchange.

2.1 KOSTRA

The KOSTRA project started in 1994 and aims at a co-ordinated and improved reporting chain between Norwegian municipalities and counties and the Norwegian government. The project encompasses a harmonisation of the content of the data to be reported, a better organisation of the reporting chain (by reporting mainly from one point in each municipality or county to one point in Statistics Norway), and implementing a solution to replace the existing paper based reporting forms with electronic forms. The final objective is to:

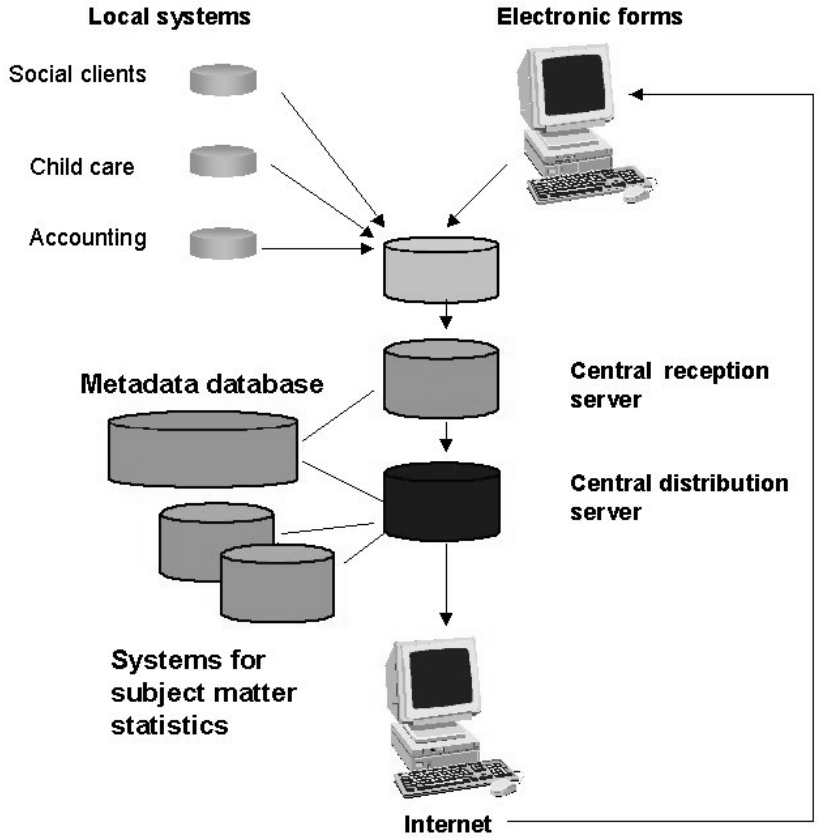
- Enhance the quality of data and statistics on services and the use of resources in local government administrations in Norway, by improving relevance, accuracy, timeliness, comparability between local administrative units and accessibility
- Collect, compile and disseminate statistical information in this field with fewer resources in terms of money and manpower

[2] KOSTRA= Kommune - Stat - Rapportering = Municipality - State - Reporting

[3] IDUN = Informasjon og Datautveksling med Næringslivet = Information and Data Exchange with Businesses

The KOSTRA project is currently in the end of a pilot phase. The number of reporting units has been increased gradually over the years, and for the first time all municipalities (435) and counties (19) have reported their data from the year 2001 electronically following KOSTRA concepts and definitions. Electronic reporting is now compulsory for these administrative units. Most of the respondents have reported in EDIFACT format, but this year 24 of the municipalities and all the counties participated in a test reporting in XML format. KOSTRA collects data both by electronic questionnaires and by direct extraction of files from administrative systems. Some data have to be encrypted. Figure 2 shows the KOSTRA reporting chain.

Figure 2. KOSTRA reporting chain



KOSTRA represents a revolution in our way of producing statistics for the public sector. So far it has affected in particular the timeliness of statistics. All figures for the year 2001 were to be reported by 15. February 2002. We were able to publish data 15. March, though without in-house editing (only automatic controls linked to the questionnaires have prevented or warned for possible erroneous data at reporting). Hence, these data contain some errors, but since they are published the respondents can participate in the editing process. Final figures are published 15. June 2002. Even this date represents a considerable improvement in timeliness for most of the statistics.

Until this year electronic reporting and the traditional routines with paper questionnaires have been going on in parallel, and extra resources have had to be put into development. Hence, an expected gain in costs has not materialised in KOSTRA yet. Development will go on, but improved efficiency will be an important issue from the reporting year 2002.

A gradual change from EDIFACT to XML format will contribute to improvements, both for the respondents and in Statistics Norway. XML is supported by all major software vendors, and is less cumbersome to install and to use than the different software products needed for the EDIFACT solutions. Furthermore, the XML solution will be given to the respondents free of charge (the only requirement will be IE 5.0+), and training courses should be unnecessary. Another advantage of reporting using XML is that it probably will give less errors since all questionnaires are produced from a metadata database. Metadata stored here control the questionnaires through the entire reporting chain.

2.2 IDUN

The IDUN project started in 2000 and has as its main goal to develop a scaleable, general solution for web-based data collection from business companies. The purpose is to enhance the quality and speed of the exchange of information with business companies in order to reduce businesses workload. Co-ordination and co-operation with other authorities collecting data, such as the tax authorities, is also an important objective to avoid double reporting and to reduce the total response burden in the society. By developing an electronic report-chain system, Statistics Norway also wants to give each client:

- Feed-back on their own information
- Simple statistical analysis of the clients' data
- Links to relevant statistics or market information
- Basic information about the enterprise or business from the Enterprise and Business Register, and an opportunity to update information located there

Today, companies have several possibilities to report data electronically to Statistics Norway:

- Transfer of files by diskettes and E-mail from each business company, or from the parent company of chain stores
- Use of local software installed in the computer of the data provider
- Linkage to Statistics Norway's Internet home page

Statistics on earnings are based on transfer of files that has been extracted from the companies' administrative systems. These files have been submitted by diskettes, but we are now able to offer an encrypted solution for transfer via the Internet. Data covering chain stores are transferred jointly and electronically from the parent company by diskettes or e-mail. This solution is offered for data to be used for the monthly retail trade and the consumer price indices.

The solution with locally installed software at the respondents contains electronic questionnaires including electronic controls, encryption functions and automatic routines for submitting the data to Statistics Norway. The same solution can be used to send data to tax authorities (self declaration) and accounts to the "The Brønnøysund Register Centre[4]". The solution is operable for collecting structural data for industries and for trade. So far, the possibility to report by linking to Statistics Norway's web service is offered to the data providers for the monthly retail trade index survey and for the quarterly investment statistics.

IDUN is still in a pilot phase, and so far few companies are reporting to Statistics Norway by the help of the Internet. The number is assumed to increase rapidly, and the objective is to offer all companies the possibility of reporting electronically and probably over the Internet by the end of 2003. However, unlike KOSTRA, there are no plans to make electronic reporting compulsory for the companies in the foreseeable future. Hence, we have to keep an infrastructure that enables us to receive information in different ways and different formats.

[4] "The Brønnøysund Register Centre" is the Norwegian registration department, and consists of among others "The Register of Business Enterprises", "The Register of Company Accounts", see: http://www.brreg.no/english

The co-operation with the Norwegian Tax Administration and the Brønnøysund Register Centre to ensure that all agencies can benefit from each other's experience, and in some cases on common solutions, is central in Statistics Norway's strategy and in the IDUN project. No business company should need to deliver the same information more than once. This collaboration has a future goal to develop a common web-solution for all the information the three agencies receive over the year.

In contrast to KOSTRA, which so far is based on offline reporting, IDUN is based on online reporting directly to the Internet service at Statistics Norway

2.3 Other
The Population and Housing Census 2001 was entirely based on administrative registers for data on persons and families. However, data on houses and households have been collected by questionnaires. These data will complete a Housing register that will make new direct data collection unnecessary. The housing questionnaire was available as an Internet questionnaire (two versions, one simple and one more advanced) in addition to the paper version. About 10 percent of the households used the Internet questionnaires. All data from the Census have not been analysed yet, but the results from the Internet survey look promising with regard to data quality, see Haraldsen and Hendriks (2002).

Statistics Norway produces statistics on hunting of wild game on commission for the Directorate for Nature Management. From this year, all hunters (that have paid hunting license) were given the possibility to report on Internet by linking to Statistic Norway's web service.

An internal staff survey carried out recently has been based on Internet (Intranet in this case) reporting and XML format.

3. Response burden and electronic data collection

Most of the efforts on reducing the response burden have been focusing on the *time* it takes to provide answers to the survey questions. This is an important aspect, and a major strategy has been to replace direct data collection with indirect data collection from administrative registers. As we have seen, electronic data collection also makes it possible to collect data automatically from the data systems of the data providers, and there is certainly a potential for further reducing the time respondents use to provide data for statistical purposes. However, even if the time used by respondents for reporting to Statistics Norway is relatively low compared to the time used for reporting to tax authorities and other governmental agencies, providing data to Statistics Norway is often mentioned as a nuisance. It has been noted by Haraldsen et al (2002) that the *perceived* response burden is not determined by response time alone. Statistics Norway has used cognitive interviewing carried out at respondents' place of work to test electronic questionnaires and study the perceived response burden.

First we have learned that response burden is a personal feeling that may not coincide with the time it takes to answer the questions. Secondly the tests have revealed that the cognitive problems the respondents encounter in front of the PC-screen is a mixture of general business survey problems, and specific problems that have to do with the Internet format of the surveys. Both kinds of problems must be addressed in order to reduce the response burden. Third, the tests clearly show that the weight we put on the respondents' shoulder can be counterbalanced by tailored question flow and wordings, automatic help and quality control while the answers are filled in, and by different kinds of electronic reports after the questionnaire is delivered. The test participants gave us many ideas about what kind of functionality and feedback they expect and appreciate in an Internet-survey. Haraldsen et al (2002) have also noted that it is not only the respondent's work with the questionnaire that is relevant for the response burden, but also the administering of the questionnaire. A paper questionnaire normally requires more work connected to reception and submission of mail than an electronic (on line) questionnaire.

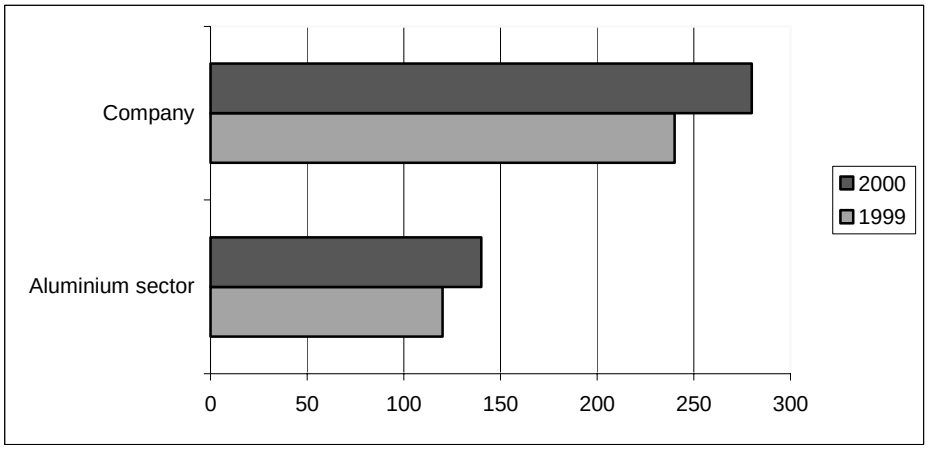
To sum up, electronic data collection can reduce perceived response burden by:

- Reducing response time by making data collection more automatic
- Providing a simpler and (for some) a more interesting and quick way of answering
- Motivating the respondents by communication and giving something back

These possible effects should also contribute to improve data quality.

An example of a motivating measure is to give the respondent statistics where their own data are compared to statistics for other respondents (expressed as statistics for groups to ensure protection of privacy), for example for benchmarking purposes. Such statistics must be worked out specifically for each data provider, but it can be done in a standardised way. Figure 3 shows an example where emission of CO_2 of one company in the aluminium sector is compared to the average for the sector.

Figure 3. Emission of CO_2 per produced unit of aluminium. Kg/ton



I addition to direct reporting of such comparable indicators, the companies are referred or linked to Statistics Norway's web-pages where relevant official statistics can be found. The industrial sector will in general demand quite detailed statistics, for benchmarking and market analyses. These needs have to be balanced against protection of privacy and data quality.

4. Metadata

Electronic and automatic data collection requires precise definitions of concepts and classifications, and good explanations and guidance both with regard to the content of questionnaires and the technical solutions. In both the KOSTRA and IDUN projects electronic questionnaires are derived from metadata stored in databases.

The aim of the metadata database is to store and maintain all information necessary to automatically generate questionnaires. This includes the questions, the formatted fields for the answers, checks on validity, cross-checking, calculations and response based routing. This information can also be used by software vendors to prepare interfaces to their own systems for direct reporting for instance from accounting systems. Also draft paper questionnaires should be generated from the database using templates or style sheets.

With regards to the statisticians, the main objective is to provide them with an interface to the metadata database that enables them to design and maintain their own questionnaires. Since many data will be collected by different means, one single source for the definition and maintenance of questionnaires will be increasingly important.

This kind of questionnaire repository must be linked to other basic metadata components like variable definitions, classifications etc., contributing to the design of an overall and integrated metadata system throughout the institute. This is important in order to be able to link the questions asked to the statistical results finally presented. A link that often vanishes when turning into the use of electronic questionnaires.

5. Data security

If Statistics Norway either collects data directly or derives statistics from administrative registers, protection of information about persons and companies/enterprises is crucial for its operations. The Personal Data Act and guidelines from the Data Protection Directorate give rules and procedures for collection and treatment of individual data with very strict routines for sensitive data on persons. Some of the information collected in KOSTRA concerns persons. Information sent to Statistics Norway from companies does usually not contain personal data and is thus not considered sensitive in the framework of the Personal Data Act. However, as these data may contain market sensitive information, a very high degree of data security is necessary also here. Hence, solutions to protect data are being developed for all our projects on electronic data collection.

The security solutions can be split into two components; one is the security at the level of transactions and the other is the security within Statistics Norway. We have a long experience in protecting our data from any kind of forced entries, both with regards to physical barriers and widely use of the need-to-know principle. However, the solutions on data security at the level of transactions are not yet mature. A number of solutions are available, both for encryption and for trusted third party certificates and acknowledgements. Like in most countries we are still waiting for common standards to evolve within this field. It makes no sense if different public bodies implement different security solutions for their reporting. Currently, large efforts are put into development and decision upon a common infrastructure of security for electronic reporting to the public authorities.

6. Organisational aspects

Over the last years, technology has led to a stronger co-ordination of the dissemination activities in Statistics Norway. This is due to Internet technology and the need for a common profile and common routines for release of statistics on http://www.ssb.no/. A new and common technology for data collection through the Internet calls for a stronger co-ordination of data collection and a co-ordination between data collection and dissemination. A common reception for electronic data has already been established in the Division for data collection. This division will be responsible for all direct data collection in Statistics Norway except for data from persons collected by the Division for sample surveys.

One of the great challenges for the Division for data collection is to handle and combine data collected from different sources and in different ways. As mentioned, the business companies will have possibilities to report both electronically and on paper for many years. In pilot surveys in the IDUN-project about 25 percent of the respondents announced that they would try the Internet-version. While it will take time to reduce the costs caused by electronic reporting for the society, the costs for Statistics Norway may increase in the short run. Developing a new technological platform and IT-systems and operating several systems for data-collection in parallel, require both extra financial and human resources.

In the Division for data collection data from different sources and collected in different ways (for example data from administrative systems, paper questionnaires that are scanned and Internet questionnaires), are merged before they are transferred into databases for editing etc. Research indicates that there can be differences not only in data quality but also in the way respondents interpret

and answer the questions on paper and on Internet. Hence, collecting data in different ways gives methodological challenges in addition to challenges linked to technology and costs.

Administrative challenges of electronic data collection in Statistics Norway are addressed in a project with the objective to establish good and well-organised routines and systems to handle the different data flows from the respondents into the productions systems. Organising routines for assuring data security (for example administering usernames and passwords) is an important part of the project. It also includes routines for reminding respondents that do not answer (using E-mail as an alternative to ordinary mail), and new routines for fining respondents not answering in compulsory surveys. Consideration of the legislation rules that have been made for paper mail is an issue. Methods to organise a help-desk and administer phone-numbers and e-mail addresses are also considered.

All the new challenges put requirements on how a data collection has to be organised and what kind of knowledge that is needed. It seems obvious that IT-knowledge becomes of more and more importance. The Norwegian experiences are nevertheless that the data collection unit shall not be an IT-unit. But a close and good cooperation with IT-staff is a precondition for success. The need for knowledge changes when data are received in a new way. Resources are to be allocated from scanning etc. to other areas, such as operating new IT-routines and quality control. Human resource development is therefore a key issue when introducing electronic data collection in an organisation.

The overall organisation of Statistics Norway is considered on the basis of the new strategic plan. The need for a co-ordination of data collection is underlined. This has already resulted in a common reception for electronic data in the Division for data collection. However, among other proposals it is also being considered if a larger data collection unit (department) should be formed, possibly including both the Division for data collection, Division for sample surveys, Division for business register and expertise in questionnaire methodology.

7. Future prospects

The paper describes solutions for electronic data collection that are being developed in several areas in Statistics Norway, in collaboration with different external agencies. The long term strategic goals of improving quality and reducing costs both for the society and in Statistics Norway concern all areas of statistics, and the solutions developed in projects such as KOSTRA and IDUN will be extended and used also in other areas. In the first round we will develop a common web portal for both projects and other data collection activities. Electronic questionnaires and the technology they are based on will be improved and harmonised, using our knowledge on how respondents perceive such questionnaires. We foresee XML as the preferred format, common security solutions and a common metadata database for reporting both from municipalities, businesses and other data providers. A further build-up of expertise in data collection activities and technologies is an important measure.

References

Bergstrøm, Yngve (2002): *New approaches to data exchange.* Paper to UNECE/Eurostat Work Session on Electronic Data Reporting, Geneva, February 2002

Haraldsen, Gustav and Hendriks, Coen (2002): *Census going online.* Sigma, the bulletin of European statistics. No 1/2002

Haraldsen et al (2002): *Identifying and reducing the Response Burden in Internet Business Surveys.* Paper to the International Conference on Questionnaire Development, Evaluation and Testing Methods, Charleston, South Carolina, November 2002.

Hopland, Magne (2002): *Electronic data reporting from municipalities and businesses. Some experiences from two projects in Norway.* Paper to UNECE/Eurostat Seminar on Integrated Statistical Information Systems and Related Matters (ISIS 2002), Geneva, April 2002

Ljones, Olav and Svinnset, Anne Britt (2001): *KOSTRA - A Model for Official Statistics Built on a Partnership between Local and Central Government.* Paper to the 53th ISI session, Seoul, August 2001

Recent publications in the series Documents

2000/16 J. Thori Lind: The Use of Household Welfare Functions to Estimate Equivalence Scales

2001/1 T. Bye, R Choudhury, M. Harlarson and P. Hardarson: The ISM model: A CGE model for the Icelandic Economy

2001/2 K.Ø. Sørensen, J.L. Hass, H. Sjølie, P. Tønjum and K. Erlandsen: Norwegian Economic and Environment Accounts (NOREEA) Phase 2

2001/3 G. Haakonsen, K. Rypdal, P. Schøning and S.E. Stave: Towards a National Indicator for Noise Exposure and Annoyance: Part I: Building a Model for Traffic Noise Emissions and Exposure

2001/4 T. Åvitsland: User Costs of Real Capital

2001/5 H.V. Sæbø: Strategic planning and management in Institutio Nacional de Estatística, Mozambique: Short Term Mission Report, November 2 - 9, 2000

2001/6 H.V. Sæbø, J. Byfuglien and R. Johannessen: Quality issues in Statistics Norway

2001/7 J. Byfuglien: Comparability of income data: Project report

2001/8 A.S. Abrahamsen,.M.Q. Andersen and R.Ragnarsøn: Project: Quality of Manufacturing Statistics and Evaluation of Methods for Producing Preliminary and Final Figures. Final Report

2001/9 T. Skoglund: Employment in the Norwegian National Accounts

2001/10 O.C. Moen: Nordic Economic Growth in Light of New TheoryL: Overoptimism about R&D and Human Capital?

2001/11 H.V. Sæbø: Strategic Planning and Management in Instituto Nacional de Estatistica, Mozambique: Short Term Mission Report, August 27 - September 6, 2001

2001/12 B. Hoem: Environmental Pressure Information System (EPIS) for the household sector in Norway

2001/13 H. Brunborg, I. Bowler, A.Y. Choudhury and M. Nasreen: Appraisal of the Birth and Death Registration Project in Bangladesh

2001/14 K. Rypdal: CO_2 Emission Estimates for Norway. Methodological Difficulties

2001/15 E. Røed Larsen: Bridging the Gap between Micro and Macro: Interdependence, Contagious Beliefs and Consumer Confidence

2001/16 L. Rogstad: GIS-projects in Statistics Norway 2000/2001

2002/1 B. Hoem, K. Erlandsen og T. Smith: Comparisions between two Calculation Methods: LCA using EPIS-data and Input-Output Analysis using Norway' s NAMEA-Air Data

2002/2 R. Bjørnstad: The Major Debates in Macroeconomic Thought - a Historical Outline

2002/3 J. L. Hass and T. Smith: Methodology Work for Environmental Protection Investment and Current Expenditures in the Manufacturing Industry. Final Report to Eurostat.

2002/4 R. Bjørnstad, Å. Cappelen, I. Holm and T. Skjerpen: Past and Future Changes in the Structure of Wages and Skills

2002/5 P. Boug, Å. Cappelen and A. Rygh Swensen: Expectations and Regime Robustness in Price Formation: Evidence from VAR Models and Recursive Methods

2002/6 B.J. Eriksson, A.B. Dahle, R. Haugan, L. E. Legernes, J. Myklebust and E. Skauen: Price Indices for Capital Goods. Part 2 - A Status Report

2002/7 R. Kjeldstad and M. Rønsen: Welfare, Rules, Business Cycles and the Employment of Single Parents

2002/8 B.K. Wold, I.T. Olsen and S. Opdahl: Basic Social Policy Data. Basic Data to Monitor Status & Intended Policy Effects with Focus on Social Sectors incorporating Millennium Development Goals and Indicators

2002/9 T.A. Bye: Climate Change and Energy Consequenses.

2002/10 B. Halvorsen: Philosophical Issues Concerning Applied Cost-Benefit Analysis

2002/11 E. Røed Larsen: An Introductory Guide to the Economics of Sustainable Tourism

2002/12 B. Halvorsen and R. Nesbakken: Distributional Effects of Household Electricity Taxation

2002/13 H. Hungnes: Private Investments in Norway and the User Cost of Capital

2002/14 H. Hungnes: Causality of Macroeconomics: Identifying Causal Relationships from Policy Instruments to Target Variables

2002/15 J.L. Hass, K.Ø. Sørensen and K. Erlandsen: Norwegian Economic and Environment Accounts (NOREEA) Project Report -2001