Proceedings of the Joint IAOS/IASS Meeting
Electronic dissemination: Supply and Demand Issues, September 1998

SERVING DIVERSE USER NEEDS

Lars Thygesen[1]
Gösta Guteland[2]
Hallgrímur Snorrason[3]

Abstract

The paper deals with the problem of the wide diversity of needs of different types of dissemination among users of statistics. It describes the range of preferences regarding presentation mode (on demand, push), media and formats, language and terminology (colloquial or professional language, foreign language), level of detail, quality descriptions and other metadata, analysis, search mechanisms. This diversity speaks in favour of a similar diversity in products and services, ultimately one special product for each user. How can this be managed, how can coherence be preserved?
The paper seeks to describe work in three NSIs on uncovering the diversity, and on the segmentation of the statistical user population. It describes strategies for efficient, user oriented dissemination, with a special view to coherence of presentation formats. How the NSIs monitor user satisfaction. What are the budgetary implications of this user oriented dissemination approach, experience of resource requirements, creation of revenues. It describes how sharing of methods and of work across borders can help the NSIs reach their goal.
.
KEY WORDS: User Orientation, Dissemination Media, Dissemination Strategy, Output Database, Metadatabase, Segmentation, User Satisfaction Surveys.

1. The importance of dissemination

It might be argued that dissemination is only of secondary importance to the work in official statistics: The first and most crucial task of a central statistical office is to produce high quality statistical information on phenomena in society that are important to the government as well as for business and the public at large. But our work is no good if this information is not really received by anyone. Dissemination and even marketing is necessary in order to make the result of our work used by those for whom it was intended. Without good dissemination, our work is useless. The test of official statistics is whether it is being used - and of course if it is being used as it should be used!

2. The diversity of user needs

If the statistics are to appeal to the users, they must be tailored to the needs. It is necessary to use a wide variety of dissemination channels and to aim at satisfying the needs of many different groups of users. This means that statistics should be published with different levels of detail, e.g. at the geographical level. For instance, the press and the general public need a broad overview while business analysts need more details, the Government and Parliament need even more detailed

[1] Lars Thygesen, Statistics Denmark, Sejrøgade 11, DK-2100 Copenhagen Ø (Denmark), Phone +45-39173931, lth@dst.dk

[2] Gösta Guteland, Statistics Sweden, Karlavägen 100, S-11581 Stockholm (Sweden), Phone +46-87834101, G.Guteland@scb.se

[3] Hallgrímur Snorrason, Statistics Iceland, Skuggasund 3, IS-Reykjavík 150 (Iceland), Phone +354-5609804, hallgrimur.snorrason@statice.is

statistics for law preparation, and also researchers often need to make their analysis based on very detailed data. Some users need more analysis than others, some like to do the analysis themselves. And consequently, the users want different media of dissemination: Paper, diskettes, CD-ROMs, online service, Internet.

There is no doubt that demands for diversity of statistical products is increasing steeply as a result of the revolution of electronic dissemination means. The users as well as the producers are aware that it has become so much easier to customize the products, to add new facilities, and this stimulates demand. The development of IT has a bearing not only on the electronic media, but also on the paper publications. Demand shifts away from large volumes of detailed tables ("databases on paper") towards differentiated products with various levels of analysis and explanations. And it is easier to produce such products. A small or medium size database may often be supplied as a supplement to the books. "Print on demand" of publications is a rather new possibility.

2.1 Segments of users

When setting up a strategy for the diversity of dissemination, it is important to establish a segmentation or classification of the users. Until we have defined categories of users, each category with some degree of homogeneity in needs, it is difficult to discuss the organisation of our products. Of course we may find out during discussions with users that our segmentation has been wrong - in which case we will have to revise it.

In many studies users are divided into three main groups; the public, decision makers (including politicians), and researchers. However, it may be more pragmatic to work with a more detailed division into six or even more groups: The public administration, research institutions, the business sector, media, the educational system, and households, perhaps adding international institutions as a separate group.

In addition to such groupings it is often very useful to distinguish between frequent and rare users.

2.2 Measuring the needs

How can we uncover users' needs? We can *observe* their behaviour: Which products do they actually order or buy, which Internet-pages do they access. How many retrievals from the databases are made by different user segments.


When the new statistical databases were launched in Sweden in the beginning of 1997 they were free of charge. The interest was very large from the beginning. In March 1997 4,000 statistical tables were retrieved by the users. From the beginning of March a fee was introduced. In 1998 the figures are above the corresponding figures for 1997 in spite of the fact that a fee now has to be paid.

The observation method has the shortcoming that there may be strong needs that cannot be met by existing products. So maybe it is better simply to ask the users, listen to them. This process can take several different shapes.

The most common way is to conduct *discussions* with the customers or customers-to-be. Sometimes these discussions occur in user committees or in special seminars with users and

producers. The information service of the statistical institute may also be a valuable source of information on user needs and priorities, both as regards content and the service medium preferred by the users.

Measurement of users' needs can also take the shape of systematic *user surveys* based on traditional statistical methods: questionnaires or more intensive interviews. The questions may be adapted to the different user segments.

(S) At Statistics Sweden, committees with users have been created for discussion of needs.
As an example the newly created Board on Databases can be mentioned. This board consists of users from a ministry, two agencies and representatives for municipalities, research institutes and private companies. The board gives advice to Statistics Sweden on how to develop the databases. The members act also as test-pilots when new data bases from Statistics Sweden are presented.
A lot of opinions have been presented at the meetings with the Board on Databases. They can be summarised in the following way:

a. The databases have to be very user-friendly
b. They have to be cheap
c. They have to include metadata
d. They have to include both macro and micro data

A distinction must be made between those users who use the databases very frequently and those who use them more seldom. It may be necessary to develop two versions of the databases - or two access methods to one database.
Priority must be given to availability even if it is also important to include as much as possible in the databases.

(DK) As part of the preparation of a new dissemination strategy in Denmark, three surveys were carried out in order to uncover user needs[4]:

1. The first survey was conducted in November 1996. 490 questionnaires were sent to subscribers who receive publications free of charge, response rate 83 per cent.
2. In May 1997, a survey based on intensive interviews was conducted. 11 important users of statistics from different sectors of society were interviewed.
3. In October 1997, a second questionnaire-based survey was conducted. 1,200 paying subscribers were approached, response rate 59 per cent. In addition to the subscribers, 23 hand-picked journalists also received a questionnaire, with a response rate of 78 per cent.

The surveys revealed differences between users' preferences, e.g. that young users will be satisfied with a strengthening of web dissemination while older persons are much less interested in this. They gave important information on users' views on, *inter alia*, the balance between paper and electronic information, the relative importance of elements like documentation of data quality, and comments on trends. The results play an important role as the basis for the new strategy.

(IS) Statistics Iceland is such a small institution that it is hardly feasible to carry on user contact in formal committees or to gauge user needs frequently through systematic user surveys. Instead, emphasis is placed on utilizing the information service of the institution for systematically observing user demand. This relates to all relevant factors such as the field of statistics, the level of detail, the geographic breakdown, and the requested medium for dissemination. Information on this observed demand is then transmitted, both directly and indirectly, to the different production units together with concrete requests. The units are obliged to tailor their production to these needs, if that seems feasible, or make their output available in such a form that the demand may be more readily satisfied.

[4] Hornum (1998)

2.3 How to react to user needs

Once information on user demand has been collected or user needs have been measured or assessed in some way, some kind of systematic analysis should be made. The basic task is to compare the observed needs with the products that are available or we plan to make available. This may for instance be done by creating a profile for each segment or group of users.

One possible approach is to draw up a matrix with one of the axes showing the different fields of statistics, broken down by type or level or whatever may be relevant, with the other axis showing the different user groups. Such a matrix may be a useful tool in analysing the present and future demand for statistics and the present and eventual supply of products and services.

Furthermore, it can be quite useful to examine a total matrix of user needs where the user segments are cross-tabulated against a range of different information products, some of which may be at the planning stage.

(DK) Statistics Denmark set up a strategy for dissemination in December 1997. An important element was an analysis of a segmentation of the users, supplemented by a matrix like the following:

Example: Matrix of user demand for information products

	Interest-ed citizen	**The press**	**Libra-ries**	**Govern ment-**	**Muni-cipali-ties**	**Finan-cial sector**	**Re-search**	**Edu-cation**	**Inter-natio-nal**
A1. Overview publications									
Work Plan		X		X		X	X		
Catalogue	X	X	X	X	X	X	X	X	X
etc.									
A2. General statistical publications									
News Release		X		X	(X)	(X)		(X)	
Statistical Yearbook	X	X	X	X	X	X	X	X	X
etc.									
A3. Thematic publications									
Living conditions	X	X	X	X	X		X	X	X
etc.									
B1. Web site	X	X	X	X	X	X	X	X	X
B2. Online Data banks		X	X	X	X	X	X	X	
B3. Nordic Statistics on cd-rom	X	X	X		X	X	X	X	X
B4. Meta databases									
Product description	X	X	X	X	X	X	X		X
Classifications database			X	X		X	X		X
etc.									

The matrix was supplemented with a series of “scenarios” for each group of users: How do we believe that this group can best be served? It can be used as a basis for discussion with users of their real needs: Do you really use these products, and how can they support one another. It revealed blank spots where services were missing.

2.4 Educating users

Educating users may be quite a worthy task of a statistical institute and a very practical one as well. Frequently, users are not able to state their wishes clearly. They may have unrealistic, preconceived ideas of what they need or state their demand in terms of specific variables, tables or time series that may not fit the use they are going to put the statistics to. They may also misunderstand the data, be badly acquainted with definitions or classifications in use or be unaware of the limitations of the material they request. Misconceptions of this kind may be detected at the outset when the users state their original demand or after the event when the damage has been done, such as in the media or when requests for updates are received. Hence, it is very important that whenever there is doubt as to how sensible the request for data may be, that the prospective use of the data is discussed with the user. One thing that ought to be borne in mind is that the user will normally not have full knowledge of the supply of statistics and the availability of products or services.

For these reasons it is important to provide the users with knowledge of the availability of statistics as well as their usefulness and limitations. The users should also be made aware of the available metadata. Educating the users in this respect can be done in several different ways, by continuous user contact, by producing and transmitting to them brochures or even user magazines (and making similar information available on the web-site of the institution) and by conducting special courses for specific user groups such as journalists. The main aim is to make them better and more demanding users thereby enhancing the quality and quantity of the statistical services and their utilization.

3. Diversity versus coherence of dissemination

User-orientation does not equal trying to satisfy every whim of the users. If we were to do that there would be a great risk that we would end up in chaos, both with respect to production and dissemination. Thus, it is necessary to distinguish between whims and needs and it is important that all needs that fall within our realm of competence and feasibility be met. If this is to be achieved we must ensure that what we disseminate is strictly coherent, both as regards *content* and *format*. This demands a high degree of *standardization* and a clear *dissemination strategy*.

First of all, this is important because of the users. They need to know that when we speak of one concept, one definition or classification we always mean the same thing. They need to compare one field of statistics with another in a meaningful way. Hence, it is evident that if dissemination is to be satisfactory it needs to be *standardize*d and coherent as regards content. Concepts, definitions and classifications must be the same across the different areas of statistics. There has always been the need for that, only it is becoming even more pressing as the volume of disseminated statistics grows.

But maintaining strong coherence is no less important to the statistical office in order to keep up an efficient production. All products should be made in one process without duplication of work, meaning that publications are based on the databases, instead of first producing the books, and afterwards the database.

In order to accomplish all this, it is important to use one common software platform for all dissemination, be it electronic or on paper. This software should support a *statistical database* (or *data warehouse)* from which all publications and disseminated products be drawn. Furthermore, it is expedient to have one common “place” for storing the data and to use common metadata.

There is also a need for *standardiz*ation of formats and systems. The new policies stress the need for using the same formats for storing data, including the metadata. They also emphasise that it should be made possible for the user to access all of our statistics using just one system or user interface. Choosing formats and systems poses a difficult problem as there is no widely accepted standard format.

Finally, it is important for the NSI that it can be recognised by the users across the whole spectre of products. This means that there must be a corporate identity, or a common look-and-feel.

4. Basic versus value added services

In most countries the Government will pay the bulk of the expenses for the production of statistics. Since the tax-payers have already paid for the statistical content, it is evident that the public and the different user segments have a right of expecting that these data will be put at their disposal. In some format and level of detail the statistics should be free of charge, or nearly free. It may be necessary to have some fee, covering the bare marginal cost of making the statistics available. A minimum requirement is as a rule that the statistics that are paid by the Government are available free of charge at libraries.

Which statistics shall then be available at this low cost? How can a border be drawn to other statistics? In the Nordic countries the concept "Official statistics" is used for this delimitation.

Ⓢ In Sweden, the Act of Parliament concerning official statistics states that official statistics are statistics utilised for planning, research and public information purposes and for international reporting, produced by a government agency in accordance with provisions notified by the Government. Under the Act, official statistics are to be objective and available to the public, and they are to be produced and published taking protection of the individual into consideration.
A decision is taken by Statistics Sweden that all official statistics shall be available in databases at the end of 1999. For the moment the most frequently used official statistics are available in this new media.

When we proceed to fulfil the diversity of users' needs, it is evident that the possibilities of spending resources are unlimited. The Government has not allowed for this extravagance in their often very scarce allowances. This is the main reason for demanding at least some fee for the access to databases. The revenues can be used for personal service to the users, for the maintenance of the databases and for investments. We are aware of the fact that the situation is not the same in all countries. NSIs in some countries are not allowed to keep revenues from selling[5]. In such cases special economic agreements have to be reached.

If users want more sophisticated products or access methods, these are additional services which must be paid for. There are many possibilities of introducing such products.

Ⓢ In Sweden a new service will be introduced during the autumn where the user can click on a button in the databases for a personal service where demanded figures will be sent to the user by e-mail as soon as new data are available. The user will be charged for this service.

[5] TACIS (1998)

It is of course possible to combine the electronic format with paper format. A fax can be sent at the same time as new data appear on the screen. The new medium gives new degrees of freedoms. Basic statistics can be combined with value added services in an unlimited number of ways.

5. Strategy for dissemination

All of the elements we have touched upon so far have to be integrated in a dissemination strategy. This strategy will naturally comprise paper publications as an essential part, as well as electronic media. As the strategy must aim at maximum coherence and efficiency, it is necessary that these media are produced in one process. Therefore, the central element in the strategy will be the *Statistical Database*[6] with aggregate figures. The publications will have to be produced from that database.

To-day it is natural that every statistical office will use the Web as perhaps the most important channel of dissemination. The Web site should offer services for advanced users as well as for the ordinary citizen. There should be access to interactive databases, overall descriptions of the range of products and services, as well as comprehensive metadata

5.1 The national approach

Setting up of a strategy for dissemination is of course an important task for every national statistical office, as well as for the international statistical agencies. The strategy will have to be adapted to the conditions in the country. The strategy needs current updating to meet users expectations, including technical updating. For instance, it is no good sticking to a browser that has been abandoned by everybody else.

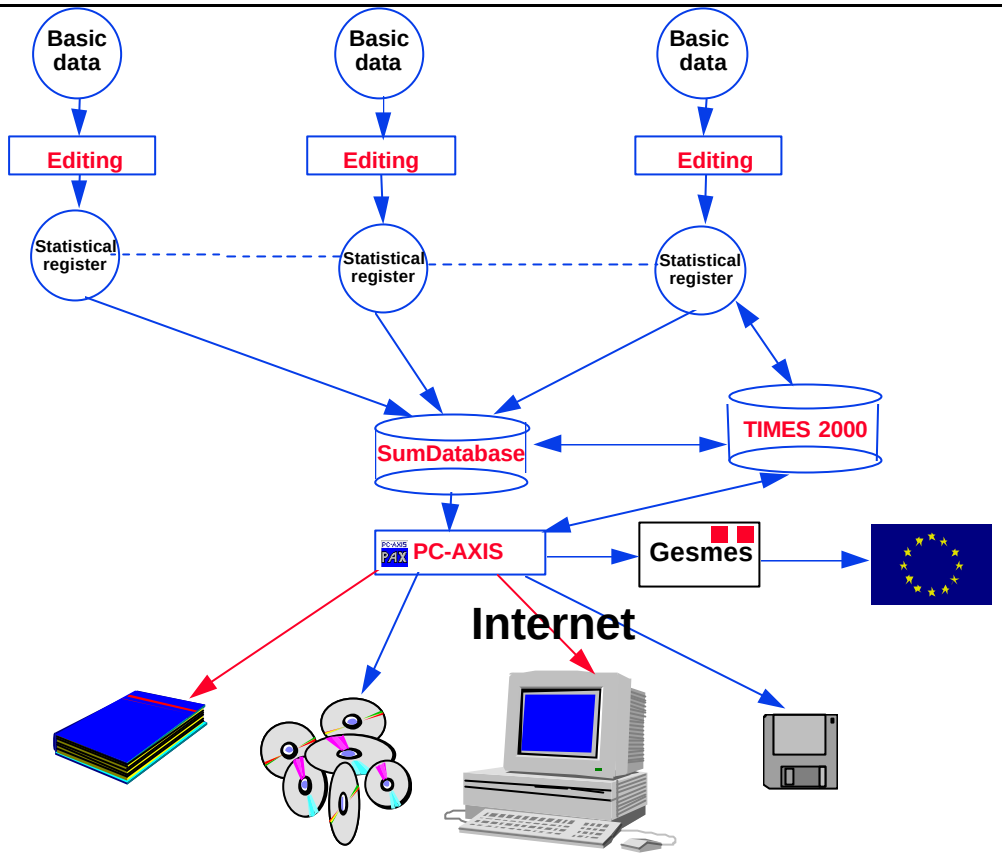
Keeping the strategy up to date and always adapting to new needs is a heavy burden for the NSI. At the same time, the NSI also has to deal with the conservatism of many users who could not bother to change anything. This means that in many cases we will have to maintain the old systems as well.

DK **The SumDatabase and TIMES 2000**

In Statistics Denmark, a Technology Policy lays down that there will be one common output database, the *SumDatabase*, where all statistics to be disseminated should be stored, whether the dissemination is to take place via paper or on-line or off-line electronic media. This database will be connected to a general documentation database, *TIMES 2000*. Both databases will form the basis for an on-line service on the Internet by January 1999.

The overall picture will be something like below.

[6] This concept is known under different names: The reference database, the output database.



A short explanation, starting from the top: The statistical production process results in a set of *Statistical registers* with clean data at the most disaggregate level (micro data). They constitute the treasure of the NSI. They are described by metadata in the database *TIMES 2000*. From these registers are produced a lot of aggregates (macro data) that are put, in a standardized SQL format, in the *SumDatabase* which is also well documented in the metadatabase. In the lower part, different outlets or services are shown, using the PC-AXIS software. Standard EDI messages in GESMES format are sent to Eurostat (and other international agencies).

The new strategy for dissemination

In Sweden a new strategy for dissemination and publication has been adopted. As is said above all official statistics and also some other statistics have to be available in the new databases at the end of 1999. From that time the following moments can be separated in the processes of dissemination:

1. Official statistics and metadata are transformed from the production processes to the Swedish Statistical Databases.
2. Press releases and metadata are produced in a standardized way based on data and metadata from the databases.
3. The electronic document is converted to a PDF-format which gives guarantees a correct printing independently of the printer chosen. The document is also converted to a HTLM-format for publishing on the Internet.
4. A user can reach the information in different ways:

a) The user looks at the web-page and, if desired, prints it partly or as a whole
b) The user fetches the corresponding PDF-file and prints it.
c) The user fetches a PX-file with corresponding content and continues in PC-AXIS
d) The user phones or sends an e-mail to somebody at Statistics Sweden who prints the PDF-file at the office and sends it back by mail, fax or e-mail.
e) The user wants a tailor-made product and asks somebody at Statistics Sweden to produce it.
f) The user calls the service telephone that is open day and night and asks for figures in some pre-announced fields like the latest figures on Consumer Price Index. It is also possible to order books or other statistical documents by this service telephone
g) The user can just as before buy a book or a statistical message from Statistics Sweden or a bookshop.

5.2 The need for an international approach

It is a rule almost without exceptions that every statistical database and every off-line product has its own access method or at least its own software (browser). In some instances, even the same bureau will offer a number of different systems. This is not helping the customer, even if the differences may be due to special wishes from different kinds of customers - which is not always the case.

This problem becomes evident in the work of our statistical libraries who have to serve customers with statistics that go across many systems. We may of course see this as a business opportunity: When users can't manage themselves to learn to use all these systems, there is money to be earned in helping them. And why shouldn't we earn them? But even our own librarians have great difficulties in finding their way. This is not a trivial or easy problem to solve as long as there are no common standards.

On the other hand, it ought to be possible to agree on such standards if we take a pragmatic approach. In the European Union, work has been going on for several years within the IDA project (Interchange of Data between Administrations) with the goal of creating a distributed statistical system, allowing users to retrieve statistics from different databases in different countries without noticing the boundaries. This would result in a coherent system of basic database services. Unfortunately, progress is very slow and it seems that a more result-oriented and pragmatic approach is needed.

Another good reason for international co-operation is the need to be resource efficient. The rapid development of IT, as well as the adoption of IT in all sectors of society has a great impact on the behaviour of the NSIs concerning the dissemination of statistics. The cuts in budgets for the NSIs and the demand for easier access to the statistics will force the management of the NSIs to find the most rational and streamlined ways to reach the users. Pre-conditions for such an approach is to use standards to establish cost-effective and flexible processes for electronic as well as paper based dissemination efforts. The limited resources will also force the NSIs to a greater extent to co-operate. Not only talking to each other on conferences but also co-operation on a practical level.

We would like to reach a situation where the national and international statistical offices, within their own countries as well as between them, agree on a set of common database and metadata formats. This should allow a statistical office to give service to their users across boundaries, using one user interface (or browser). The standards should be sufficiently rich to allow for inclusion of important statistical content and metadata, e.g. footnotes, or standard symbols for "not applicable". The standard formats should allow for use of different browsers, as there may be a wish to use "one's own brand" of browser in each country.

The desired situation, as described above, is very far away to-day.

DK IS S **Co-operation between the Nordic countries**

There is a strong tradition for co-operation and exchange of experience between the Nordic statistical offices (Denmark, Finland, Iceland, Norway, and Sweden), going back more than 100 years. During the past 12 years, a more operational co-operation has developed in the field of electronic dissemination. This involves increased use of tools developed in the other bureaux, and also common development where the offices contribute with components of a common product.

Since 1986 Statistics Denmark has gained very much by developing its own database services using the Swedish AXIS statistical database management system[7]. In 1992, all five Nordic countries joined forces in publishing a common yearly CD-ROM with Nordic statistics containing a rich statistical database for each country as well as a selection of harmonised statistics. The CD is based on the statistical database system PC-AXIS, developed by Statistics Sweden. The other countries contribute economically to the financial basis for the further development of this product. Increasingly, new modules and facilities are developed in co-operation. For instance, Statistics Denmark participated together with Statistics Sweden in the development of the PC-AXIS Windows version in 1995, and Statistics Denmark has also developed some of the commonly used modules.

The fruitful co-operation made it possible with very short notice to develop a small viewer program (PX-MINI) which was urgently needed in Denmark for presenting small electronic tables to be delivered together with a paper publication.

Another result of the co-operation is the newly introduced possibility of using the Nordic CD-ROM in combination with Internet. The user can start with the CD, look at some figures on the CD, and then continue to a database in some Nordic country by using a link that is immediately available at the CD. A precondition is of course that the user is connected to the Internet.

From 1997, the five countries together started the development of a new Nordic Statistical Yearbook with comparable Nordic statistics, based on a common PC-AXIS database.

When Statistics Sweden launched its new SQL databases on the Internet in 1996 with a PC-AXIS client as the main user interface, Denmark decided to follow basically the same model. This has made it possible to work together to a degree never seen before. Knowledge, models, software and pieces of software can be exchanged. One office does not have to carry alone the heavy burden of keeping abreast with users demands. Practical results have already proved to be very concrete. Development plans are discussed and decided together. The development work is pooled and can be used by both offices, as well as other users of the PC-AXIS family of tools, including UN/ECE Geneva, Spain, and the Philippines.

Statistics Iceland is now determined to follow the same model of database development on the Internet even though it is a tough task for a small office. Also Finland and Norway have shown interest in joining this line, and it is hoped that in this way we can create a really strong team behind a common effort.

REFERENCES

Hornum, H. (1998). *Shifting to electronic products - keeping strategies consistent with user preferences.* UN Geneva.
www.unece.org/stats/documents/1998.06.infmedia.htm (paper no. 3)

TACIS Task Force on Dissemination Policy (1998). *Dissemination Policy - Problems and Recommendations.* Strategic document for the *High Level TACIS Seminar in Statistics* in Almaty October 1998. Luxembourg, June 1998

[7] Mainframe based