Policies and Tools to make OECD Statistics more Visible and Accessible

Enrico Giovannini,
OECD Statistics Directorate
2 rue André Pascal
75016 Paris, France
enrico.giovannini@oecd.org

Russell Penlington
OECD Statistics Directorate
2 rue André Pascal
75016 Paris, France

Lars Thygesen
OECD Statistics Directorate
2 rue André Pascal
75016 Paris, France

1. Why is visibility important?

Statistics are at the heart of OECD's work, supporting analysis and policy making in a large number of policy areas. Creating riches databases of internationally comparable statistical information, gathered from national sources, is thus a strong point of the Organisation. This is by no means a small or simple task.

Once the appropriate statistics are collected, OECD must make sure they are used efficiently for informed decision making in all relevant fora, such as the OECD policy committees, as well as by other bodies and users outside the Organisation. Therefore, the statistics must be well recognised and widely accessible. There is a clear interrelationship between the visibility on the one hand, and the quality and recognition of the statistics on the other. Accessibility – itself one of the quality dimensions of the OECD Quality Framework[1] - underpins the role of statistics and enhances the possibility of devoting efforts to improvement of other quality dimensions.

The policies and tools employed by OECD to enhance visibility of statistics aim at very diverse audiences, thus calling for a number of different measures and media. Firstly, the statistics should be readily available and easy to locate for experts working with the organisation and national experts in committees. Secondly, the gold mine of information should be visible to stakeholders, such as governments of member states. The international community of research should also be guided to use high quality international data in their work. And lastly, private citizens with an interest for societal matters, including students, should not be forgotten.

2. Policies for OECD statistics

In 2002, OECD set out a Statistics Strategy, aiming at developing a corporate Statistical Information System (SIS) which would increase efficiency and at the same time enhance quality of its services, including visibility and accessibility. The first version of this SIS is now in production.

[1] www.oecd.org/statistics/qualityframework

In 2003, the OECD Quality Framework was added, making more concrete the measures to be taken to strengthen visibility. The framework builds on UN Fundamental Principles of Official Statistics and on best practices among statistical organisations. As a part of the framework, all activities have to be reviewed every 4-5 years with a view to all quality aspects, including accessibility.

Particularly important to enhance interpretability of statistics is the role of metadata. Therefore, the OECD is developing a handbook on how to present data and metadata. In addition, in 2004 the OECD has established a set of common metadata management principles as well as a common structure for metadata content, to be used in all OECD statistics. The basis for the common metadata structure has drawn on the terminological work undertaken for the Metadata Common Vocabulary (MCV) of the SDMX[2]. In setting up the structure, OECD has worked with Eurostat and IMF to make sure that structures were compatible, allowing for easy exchange of metadata with other international organisations in future. The importance of well structured metadata is emphasised by key objectives to improve two aspects of statistics at the OECD: Quality and Visibility.

Some of the principles identified in the Quality Framework (mentioned above) can be addressed by the improvement in the structure and content of statistical metadata. The common metadata structure across OECD datasets enhances the coherence and interpretability of the data, to which it is attached, for end users. Furthermore, the organisation of metadata under common and clearly defined headings and definitions makes it more consistent for end users.

The OECD dissemination policy stresses the need for good and clear communication, be it in printed books or on electronic media. The policy also demands a common look and feel of OECD statistical products and services. The current dissemination policy involves charging for OECD information, including detailed statistics which are thus not seen as a public good. The aim is to recuperate costs of the dissemination. For statistics, in 2002 it has been decided that, in addition to specific databases disseminated for free according to the agreements made by member countries (for example, the database on the Programme for International Student Assessment), 10 per cent of all other databases can be made freely available on the Internet. However, the OECD policy of charging for statistics has been recently criticised, especially since many other statistical organisations have recently abandoned such a policy. During the November 2004 OECD World Forum on Key indicators, many important users called for a change.

In the area of communication and dissemination of statistics several initiatives have been undertaken since 2001. In particular, five main projects have been developed:

a) the publication of statistical press releases containing estimates for key monthly and quarterly economic indicators;
b) the publication of the “OECD Statistical Newsletter”;
c) the publication of “Statistics Brief”, a new occasional statistical publication on measurement issues elevant for policy making;
d) the creation of a “statistical portal”[3] in the framework of the new OECD web site
e) the improvement of the presentation of statistics in SourceOECD, the electronic gateway to all OECD products[4].

In conclusion, a lot of work has been done over the last few years to improve visibility, accessibility and interpretability of OECD data and users widely recognised and appreciated these

[2] www.sdmx.org
[3] www.oecd.org/statistics
[4] More information on these tools are available in the 2004 OECD Statistical Programme of work, available at www.oecd.org/statistics.

improvements. Nevertheless, there are still some outstanding issues that need to be addressed in the near future. Some of such issues are discussed in the following sections.

3. Search Engines and Visibility

While some expert users may be well aware of existing OECD statistics and how to get it, most potential users are not. Therefore the search engines on the World Wide Web are important channels for visibility, and the OECD must try to ensure that users are guided to the relevant OECD sources when they look for comparable statistics of the types provided by OECD databases. Certain key words asked should show OECD high in the search results.

Well structured and content rich metadata serve to make OECD data, metadata, and analytical material more visible online through improved indexing by major search engines. Major search engines such as Google, Yahoo, and MSN provide results by indexing text found through the inherent 'linked' nature of the internet. The two main aspects which dictate how a web page is ranked for such search engines are *content* and *link popularity*.

a) content: search engine algorithms take into account several different components of a web page's text in order to determine ranking, such as:

- Metatags: the title, description, and keywords specified in the 'head' HTML tag of a web page play a dominant role in overall ranking. If metadata is well structured, these tags can be populated with relevant content, focusing on important aspects of the OECD data available, to assist the indexing and ranking of desired search terms and phrases. For example, titles of datasets in OECD's statistical database should be sure to mention the most important concepts that can be retrieved, and keywords should further underpin this.
- Page Content: the actual words found within a web page must focus on the key concepts, increasing the "keyword density" and helping search engines rank keywords and phrases. Clearly defined common metadata item titles, actual metadata text, and descriptions of the dimensions and dimension members of the statistical tables can serve to optimize this ranking.

b) link popularity: search engines attribute each web page a popularity ranking by analysing the content and respective popularity rankings of other web pages which link to that page: The more pages point to OECD pages with statistics, the higher they are ranked. Links within the OECD web site also contributes. As metadata is attached to a defined multi-dimensional cube structure of the observation data, some metadata relating to a whole dataset, others to a time series for a country, yet others to one or a set of observations within a time series, automatic links are provided to related observation data at higher and lower levels. These links add to the link popularity of all pages where metadata is displayed. This also ensures that all metadata is found and indexed by the search engine. The optimised link popularity of the metadata page adds to the link popularity of both the metadata page and data observations page.

Therefore, to improve the visibility of OECD statistics the structure of metadata and keywords used on the Statistics Portal has to be carefully reviewed according to the metadata principles already established. On the other hand, it would be important to increase the number of links to the Portal, especially from the web sites of national statistical agencies and international bodies interested in the promotion of the visibility of statistics.

4. Tools

In addition to the set of metadata management standards, the OECD has developed a customised system to assist statisticians in applying these standards to metadata management. The system, named "MetaStore", connects to each database in a decentralised data production environment and inherits the underlying data structures and user rights. Metadata can then be imported to the metadata database and managed using the common standards. Metadata managed

and stored in this system can be easily published to the central data warehouse OECD.Stat, the basis for all publishing, online database access as well as in other dissemination media.

Both the MetaStore metadata management system and the online dissemination interface of OECD.Stat adhere to the predefined structure of common metadata items and utilise the same system of metadata attachment to data coordinates. The integrated Statistical Information System allows the seamless management and dissemination of metadata content to optimize the online visibility of OECD data and metadata.

5. Accessing Data

The online dissemination interface of OECD.Stat has been developed to facilitate the accessing of data for end users. The design aims to allow users, once they have reached the database, to easily find the data that they want and to understand the meaning and usefulness of what they have found.

Datasets have been categorised within a hierarchy of themes to facilitate navigation to desired subject matter topics. Specific data queries can then be built and saved for individual users to be reused at a later date. Metadata is displayed where applicable as users select data coordinates, starting at the dataset level and drilling down to the individual cell. These features, together with predefined data queries presented as "Hot links" giving a default look at some part of the contents of each dataset, support the indexing of structural and reference metadata by search engines to optimize the visibility of data and metadata on the World Wide Web.

Furthermore, the OECD has developed a customised site search engine that enables users to find and access data based on the dimension structure of metadata and data tables, as well as the reference metadata text descriptions which are linked to all levels of the data. This search engine applies different weightings to different metadata attachment levels, giving top priority to titles, less to labels of dimension members, etc. The response from the engine returns the specific dimension members of interest, allowing users to gain rapid access to data relevant to their search queries.

6. Challenges for the future

Although much progress has been made lately, there is still a long way to go before OECD quality statistics are used to the extent they deserve. They are not as visible to all potential users as they ought to be, and their quality and usefulness is not as obvious to everyone as one would like.

Especially, there is insufficient awareness among non-professional users, such as the student or the general public. In order to reach these audiences, it may be necessary to develop facilities for answering questions asked in natural language, using the client's own terms. It is also necessary to work more with presenting the statistics, once they have been located by the user, in an easily understandable form, revealing all its interesting secrets and messages.

The other side of the coin, when speaking for increased visibility and use, is the risk that users may find the statistics but may not be able to understand limitations of their use, or the frame of reference in which they are created. Specific information to prevent people from misinterpreting or overly use the information beyond its limits has to be developed.

RESUMÉ

Dans le contexte du Cadre de qualité pour les statistiques de l'OCDE, l'OCDE fournit des statistiques comparables nécessaires à l'analyse politique et la prise de décisions internationales. Afin de s'assurer que ces statistiques soient connues des utilisateurs potentiels, la visibilité est accrue grâce à une gamme de moyens : des publications mettant l'accent sur des développements importants dans les statistiques ; des principes communs de métadonnées, une structure commune de métadonnées à employer dans toutes les statistiques, et un outil, MetaStore, pour aider à la gestion des métadonnées ; un portail statistique et une base de données commune de statistiques, OECD.Stat, avec des métadonnées riches sur Internet ; structurer les données sur le Web de façon à les rendre facilement accessibles aux moteurs de recherche mondiaux, tel que Google. Davantage de travail est encore nécessaire sur la façon de présenter les données et métadonnées statistiques.