Statistics Denmark 23 January 2014

Annegrete Wulff, awu@dst.dk AWU/-

Dissemination policy of Statistics Denmark

The dissemination goals and objectives of Statistics Denmark are widely based upon the [European Statistics Code of Practice](http://www.dst.dk/~/media/Kontorer/16%20-%20Formidlingscenter/pkh/Code-of-Practice-european-statistics.pdf) – as a minimum. **Code of Practice** is designed to boost confidence in the European statistical system and to ensure that all producers of official statistics in the EU apply the best international statistical principles and methods.

In our interpretation of the codex we are always taking two things into account:

* We strive to fulfill the user needs and still treat all users equally.
* We need to organize our work in an efficient way so the users get the best products using the always scarce resources

Within the area of dissemination The Code of Practice includes the following principles:

* All users have equal access at the same time (6.7)
* Timely and punctual releases (13)
* Accessible, documented and comprehensible statistics (15)
* Coherence and comparability over time and across subjects (14)
* The products shall be relevant to the users(11)
* Securing confidentiality (5)
* The dissemination set-up shall be cost effective (10)

The policy describes the different actions Statistics Denmark has taken to fulfill these goals.

### Equal access for all users

Users

We intend to treat all users equally. Doing this we have to realise that users are different: they have different needs, different interests and different competences and knowledge in IT as well as statistics. When developing the products and services we usually do it within a frame of user types: the novice, the experienced and the expert or in the Corporate Information Factory terminology The Tourist, the Farmer and the Miner: the *“tourist”* who has a basic and broad interest in statistics, likes small applications giving a quick impression and easy understanding of the figures. There is the *“farmer”* who comes at a regular basis to “harvest” our web site and the StatBank for the same information every month. He needs efficient ways of connecting the always updated information to his own environment. Finally there is the *“miner”,* the specialist who wants to dig deep in our data – using the StatBank, the research access and the important metadata. Taking different user types into account helps developing products and services more suitable to the users.

Functionality and needs

What we call “the general web site” dst.dk is considered to be the primary source and “tool” for the novices. Around 50% of the users of dst.dk are first comers, not necessarily novices but novices on our site. They are offered an overview and access to statistical information without being too familiar with statistics and statistical terminology. They may tap in an “every-day” search term and are led to a topic page containing statistics where the search term belongs. It is essential, however, that these users still have the possibility to get access to detailed information - in case they did not find what they needed, or if they want to go more in detail than what is shown in tables on dst.dk. This will be done via links to StatBank tables and to publications or by filtering the search. Coherence across the dissemination channels is thus important.

For the experienced user quick retrievals and alerts are valued services: e.g. saved queries, Excel web queries, data shoots, time series retrieval, xml output and API. They worship deep links to get directly to the statistics they know so we need to be aware when things are changed. Both the experienced user and the experts are in need of proper documentation and good, standardized metadata. Some expert users have a need to analyse developments in certain statistics. It is crucial for them to know what was published months ago even though the data has been revised. We shall establish a public revision database for that purpose.

### Timely and punctual

All releases are announced in advance in a calendar. First date is set a year in advance. This can be changed later, but only if there is a good reason for it. However, 9 days before the release date the date is fixed. Any change after that will have to be approved by the Director General and to be explained on the website calendar.

All releases take place at 9:00:00 am. This regards the News releases as well as StatBank tables. The punctual release is handled via a database system where one internet database is updated during the night, while another database is public on all the internet servers with “yesterday’s news”. At exactly 9:00:00 am a switch will give the public access to the updated version of the StatBank.

Timely data are important when statistics are used for planning purposes. Statistics Denmark requires monthly statistics being released within 62 days, quarterly statistics within 92 days and annual statistics within one year after the observation time or period. The directors get monthly reports on timeliness, which is regulated through written contracts with the divisions.

### Accissible and comprehensivble

Improved accessibility

One important aspect of accessibility is *findability*. It is never enough for the statistical office to claim “we have put the information on the website so it is there for you to use”. It should be structured in a logical way using understandable lables. Statistics Denmark has recently secured that all statistical products (the web site, the StatBank, the News releases, publications and the metadata) follow the same structure of topics. This makes it possible to link statistics at different levels of detail. Search is another aspect of findability and relevant indexing of the content is a priority task. All publications are available in pdf on the web site.

A [*disclaimer*](http://www.dst.dk/en/OmDS/omweb.aspx) specifies the right to re-use the statistics. All official data can be accessed free of charge from our website and the StatBank. Data can be downloaded in different formats (e.g. Excel, csv, pc-axis, xml) and queries can be saved for easy re-use. Downloaded data can be re-used commercially as well as non-commercially on the condition of mentioning the source.

To encourage the public to use our statistics even more an API is being developed in 2013. This makes communication between machines possible and allows for using our data in other systems and with another user interface: A version of the StatBank will be accessible from smartphones and tablets. Moreover an interest among some users to get and comment on the News Releases has let Statistics Denmark to also use Twitter in the dissemination.

Visualization

Visualizing statistics creates often a better understanding and overview of the data than can be shown in a table. At least many users think so. At the same time a principle should be only to use visualization where it brings information: not making bars “jump” on the axis just to create movement. And do not use multiple layers of curves just because it creates a beautiful “landscape”. Statistics Denmark has a set of rules and guidelines to be followed when creating graphics. Maps are widely used in presentation of geographical statistics. More interactivity has been introduced: web-TV, [interactive maps (Statistics eXplorer](http://www.dst.dk/da/Statistik/bagtal/2011-07-011-unge-forlader-provinsen.aspx)), and interactive, animated graphs. The NCVA Statistics eXplorer is a tool for the experienced and expert users and provides good possibilities for the making of analyses.

Documentation/ metadata

The quality and clarity of the disseminated statistics depends to a large extent on metadata: i.e., descriptions of quality, sources, definitions etc. All tables in the StatBank and all News Releases are linked to such a [*quality declaration*](http://www.dst.dk/en/Statistik/dokumentation/Declarations.aspx). Quality declarations, Classifications, Variables and concepts are all accessible from the website and are to be connected in one system based on the DDI standard.

Metadata are updated on each round of a statistics, and domain managers have to declare that their metadata is up to date before results can be published.

Error handling

There are written procedures for error handling. Errors in the statistics are corrected as soon as possible, and it is announced on the website - either directly or with the following release. Subscribers to saved queries in the StatBank are informed directly by e-mail. Access to the erroneous periods of the table is closed while the extents of the error are being examined and corrected.

The following procedure is followed (here illustrated regarding an error in the News Release):

1. The author of the article informs the Communication Centre and the relevant head of the subject division
2. The head of division together with the head of Communication and if necessary the Director decide whether the error is a minor or a major one. In case of major errors subscribers and the press will be informed.
3. In cooperation between the subject matter division and the Communication Centre a corrected edition of the News Release is produced. And the error is reported to the internal monitoring system.

In case of major errors the News release (and StatBank table) is no longer accessible on the web site and information is sent to the press and relevant internal and external users as well as information is put on the web. A complete new version of the News Release is produced. A message about the error will be in the introduction. If the corrections have any effect on the conclusions this will be mentioned too.

Date and number of the first release will be kept. However, a date and a time for the corrected version will be added on the release and in the calendar on the web site. Only the corrected version will be available.

In case of minor errors the corrected News Release will get the remark: *Corrected in relation to the original published version*. The press maybe informed

In case of serious misinterpretations of our statistics in the media Statistics Denmark takes contact to get it corrected. We subscribe to media monitoring covering printed and electronic media.

### Coherence

Common metadata are essential to creating coherence across time and contents in surveys and topics. The StatBank that contains all official statistics uses a common set of metadata across the topics. International classifications are used where they exist.

Change of methodology or definitions may unfortunately create breaks that prevent full comparability over time. To minimize such disadvantages it has been agreed producing the statistics for an overlapping period: the same period presented with old and with new method, if possible, even re-calculating the whole series back in time. As a minimum the break will be explained in the quality declaration connected to the statistics. Another solution used for keeping longer time series is creating tables that eliminate the dimension that causes the break (e.g. longer series without a geographical dimension if this dimension has been changed over time).

### Relevance of services and products

User needs and user competences have to be taken into account when disseminating statistics. User committees representing the main users and data suppliers meet on regular basis and discuss initiatives and needs.

Regular usability tests and user satisfactions surveys are carried out and consequences and actions are planned to be taken. Moreover a survey of citizens’ view on our organisation and our products is carried out every second year, among other things to measure users’ trust in the organisation and the statistics.

Through web analytics we are studying users’ behaviour on the website and we take their search terms into account when updating the search keywords and indices.

Users are different: when we want to know about specific users’ needs (e.g., access to proper metadata or data revisions) we conduct focus group meetings. Web analytics are used as a supplement in our study of user behaviour. We follow the media use of our published statistics, collect media citations and deliver results at a daily and monthly base to all in-house interested. This is done to give the statisticians feed back on the use and interpretation of their releases.

### Confidentiality

All statistical results are published immediately after end of the production process including quality assurance. If there are quality problems limiting the usefulness, these are explained in the publishing and in the related quality declaration. Statistical materials that are so detailed that confidentiality would be at risk if published must not be published. The limits of confidentiality are defined in the Data Confidentiality Policy <http://www.dst.dk/ext/292786082/0/formid/Datafortrolighedspolitik-i-Danmarks-Statistik--pdf>

Only non-confidential data are accessible from the website. The StatBank contains aggregated, however, very detailed statistics.

Micro data are under strict control and only accessible to acknowledged and authorised researchers holding a contract with Statistics Denmark. ([*For Researchers*](http://www.dst.dk/en/TilSalg/Forskningsservice.aspx)). Contracts have been signed since 2003

### Effective dissemination

The Single Source Principle

Cost effeciency

The Single Source Principle setup is not only a principle of cost efficiency but more to secure consistency across the dissemination channels and products as well as punctual releases.

The principle of a single source for all dissemination across the media channels has been a working principle for the last 8-10 years in Statistics Denmark. All official statistics (i.e. all statistics we are obliged to produce as paid by the Budget) will be disseminated through the public and free databank [www.StatBank.dk](http://www.StatBank.dk), which will also be the source for any tables, graphs and saved queries displayed on the general web site [www.dst.dk](http://www.dst.dk). Even publications can be output from this set-up.



*The Single source principle in dissemination*

The StatBank covers all subject areas like Population, Living conditions, Foreign Trade, National Accounts: 12 topics in total. The detail of the geographic level goes down to the municipality. It consists of 3,500 multidimensional cubes that are linked together by common metadata. Data in the StatBank will be stored in the internal database SumDatabase until release time when they are replicated to the StatBank. Some data may not be replicated but made accessible for other dissemination purposes e.g. publications or international reporting. All statistics in the topic pages under [*Find statistics*](http://www.dst.dk/en/Statistik.aspx) on the website are created from the StatBank. A page showing [*Short term trends*](http://www.dst.dk/en/Statistik/emner/Konjunkturindikatorer.aspx) is another example of a page based on saved queries from the StatBank. Every single indicator is updated at the time when the “mother table” in the StatBank is updated.