

Distr.
GENERAL

ECE/CES/SEM.54/12
11 May 2006

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD Seminar on the Management of Statistical Information Systems (MSIS)
Sofia, Bulgaria, 21-23 June 2006

Topic (ii): Dissemination and client relations

HOSTING A DATA BANK

Invited Paper prepared by Annegrete Wulff, Statistics Denmark

Summary

I. Background

1. Statistics Denmark is the central authority for Danish statistics. The corporate strategy 2010 claims that all official statistics produced by Statistics Denmark shall be available free of charge on www.StatBank.dk by the end of 2007. Up till now, several billions of data are available, covering around 90 per cent of the official statistics.
2. However, statistics are produced in other institutions as well. That goes for local authorities, some ministries, public agencies, organisations etc.
3. Several of these institutions have, as well as international colleges, now and then contacted Statistics Denmark in order to get inspiration and advice in building up a database – more or less like StatBank. Or they wanted to copy the StatBank directly. We spent quite a long time with these organisations and in the end they found some commercial product or paid external consultants for carrying out their own solution. And still they were not happy.
4. As we consider ourselves among the specialists in electronic dissemination we took the opportunity to start a project where our expertise could be reused under external conditions. It was initiated late 2004 - 2005 and will be extended in 2006 and forth. The concept is *StatHost*.

II. StatBank in Statistics Denmark

5. The StatBank in Statistics Denmark is a flexible, interactive database containing macro data. No confidential data are accessible there.

Data are stored in more than 1,500 multidimensional cubes in an internal Sumdatabase.

6. Loading the data the producer is prompted for adding a release date. At that date data are copied to an external database, the StatBank that is accessed via the Internet. All released data will be instantly accessible at 9:30 AM. This is possible because we have two parallel versions of the StatBank: one is open and in use. The other is under update with all data to be released the following day. At 9:30 AM we switch between the two databases. The newly updated base will be accessible while the other one will be ready for new updates during the coming night.

7. In the StatBank users can save queries for later re-use in several standard formats. They can be alerted of updates by e-mail (Data shoot) or use the Excel web queries facility directly on the StatBank. Saved Queries can even be accessed directly from the url where you can specify whether you want it shown as a graph, Excel file or in xml (the url should then include GIF, xls or xml respectively). An example would be: www.statbank.dk/gif/58829, www.statbank.dk/xls/58829, www.statbank.dk/xml/58829

8. Connected with every table you will find a contact person. Moreover link to declaration of contents as well as links to related publications. A table can have a specific footnote which also can include links to further documentation, i.e. classifications. Navigation and functionality is explained further in a help function.

9. Retrieved data can be presented in tables, different file formats, graphs or in a map.

III. StatHost

10. In short, StatHost is built of two main elements:

- A metadata model, which relates metadata and data. Shared metadata are used across the statistical themes and topics
- A user interface accessible via the Internet

The system is built with a separation of the code layer and the presentation layer. This gives the necessary flexibility to set up different presentations based on the same code. This is necessary when the customer wants the databank to have the same look and feel as their homepage layout.

11. Around these core elements “circle” some other important systems like:

- user administration systems
- upload applications
- access right control systems
- release time control
- alert services
- web services
- etc.

12. All in all a very complicated setup that would be overkill to establish and maintain for most minor administrations. However, it is required in order to fulfil the wishes of the distributor

of statistics: These systems secure reliability and timeliness, metadata access and documentation, detailed statistics on use and users, restricted or free access etc.

13. In all, the customer can skip all technical trouble and instead blame Statistics Denmark for all inadequacy that might arise.
14. The StatHost idea is a division of labour between the customer and the host (i.e. Statistics Denmark).
15. The business model consists of:
 - a design agreement;
 - a functionality agreement;
 - a metadata and data agreement;
 - a payment that covers our work in development and maintenance.
16. The customer takes usually responsibility for the contents:
 - collects the data and transfers a flat file to Statistics Denmark
 - provides Statistics Denmark with the necessary metadata – including a contact person. Structural metadata are mandatory while reference metadata are optional. We recommend delivering reference metadata, but they are not “required” by the system.
17. For each cube or table the customer provides Statistics Denmark with several files: a flat file including data for all combinations of codes in the variables. For each variable they produce a file with codes and the attached text to be presented. The files are sent to Statistics Denmark by e-mail. In some cases the customer might buy data from Statistics Denmark which for some reason are not published. In such a case Statistics Denmark will also be responsible for delivering metadata.
18. Statistics Denmark takes responsibility of maintaining the system and carries out the following tasks:
 - loading data and metadata into the database
 - advising in construction and presentation of the multi-dimensional cubes
 - storing data
 - publishing as planned at 9:30 am on the day of release
 - tailoring the site according to design wishes
 - adding and cutting functionality
 - informing about planned maintenance workAll software and hardware are present in Statistics Denmark’s premises.
19. StatHost metadata are stored in the same Oracle database as our ordinary StatBank metadata, though separated by different naming convention. Each StatHost site gets a unique prefix attached to the value sets.
20. The StatHost databank is presented in a style sheet with logo, colours, font etc decided by the customer. It varies from site to site though, the kernel is still our StatBank.

21. The functionality in a StatHost can be exactly as in StatBank, or it is possible to deselect elements like:

- registration / password protection;
- user profile settings;
- accounting system;
- statistics on use;
- information and documentation;
- tabs;
- bi-lingual;
- search function;
- help;
- newsletter;
- listing and aggregation facility.

22. The resulting database is presented on the customer's homepage, with its own url. As they are all on Statistics Denmark's server the name for all will be xxxxx.statistikbank.dk, where the "xxxxx" is decided by the customer. So far we have examples from six different institutions. Negotiations are going on with another three institutions. Some examples are shown in annex 1.

23. A contract is set up covering the obligations, deadlines, agreements on functionality etc. Usually Statistics Denmark will offer a StatHost site on the Internet within three weeks after the customer has signed the contract.

24. The customer will pay according to a price list. Prices are set in relation to the work load experienced - and foreseen for development as well as maintenance. The tasks include are:

(a) Setting up a StatHost site

- Setting up a site on the StatBank web server (internet address)
- Adding design (style sheet and graphics)
- Setting options and functionality for the site

(b) Defining metadata and loading data

- Defining topics/themes and sub-themes
- Defining value sets (typically common for a number of cubes)
- Linking value sets to variables
- Defining datasets and loading data for the cubes
- Assigning subject areas to the site
- Adding descriptive metadata

(c) Maintenance

- Loading – updating - data
- Changing metadata, if necessary
- The technical part of hosting the site

25. The current prices¹ for these tasks are:

One-off payment	DDK	Euro	USD
Setting up a new site: <i>(Including hosting expense for the first year)</i>	35,000	4,613	5,579
Adding a cube <i>(including time dimension)</i> :	1,000	132	159
Adding one value sets to a cube:	750	99	120
Adding existing StatBank tables to a site:		500	66 80
There will be a discount of 30% per cube, if more than 10 cubes.			

Maintenance

Yearly hosting expense:		25,000	3,295	3,985
Loading data for a cube				
<i>Monthly data:</i>	250	33	40	
<i>Quarterly data:</i>		500	66	80
<i>Annual data:</i>		750	99	120

26. Thus, a data cube with two dimensions (e.g. sex and age) and a time dimension will cost DKK 2,500 for the setup and DKK 750 for annual updating. Thus, the overall costs for a StatHost solution depend mostly on the number of data cubes and the number of variables that needs to be included in the cubes.

27. Hosting an external database is not one of Statistics Denmark's primary tasks -and certainly not a task that is covered by the government grant. That is why prices are set to cover all expenses. To cover *one person* working with StatHost the revenues have to be at least 100,000 Euros a year.

28. A similar model is planned for handling and presenting customer services and statistics for sale for different divisions of Statistics Denmark itself. A StatHost bank will in that case be established in parallel to the public, free of charge StatBank. It is intended for substituting deliveries in large Excel sheets or the invention of a variety of different service systems in the statistical divisions.

29. The model builds on the idea of centralized standard tools. That means a better service to the customers and more efficient production in the divisions: They can spend the time collecting and analysing data and leave the technical presentation to a centralized standard database.

30. How does selling data base systems on the market and storing external data on our servers fit with Statistics Denmark being a serious data producing institution? We have to balance that. Statistics Denmark has come up with a solution to that challenge. We keep different principles regarding data from public and private customers respectively. The decision agreed is reflected in this table. (The PX-WEB method mentioned in the table is an Internet user interface built on the PX-file format. In this way it can run totally independent of Statistics Denmark's systems.)

¹ Exchange rates: 100 DKK = 13,18 Euro = 15,94 US Dollar

	Statistics Denmark (DST ²)		Public customer		Private customer	
	Data stored in DST		Data stored in DST	Data stored at the customer	Data stored in DST	Data stored at the customer
Data collected by DST	StatBank	StatHost model	StatHost	PX-WEB	StatHost	PX-WEB
Example	StatBank with all official statistics free of charge	"Paid service databank" (not yet implemented) e.g.	<ul style="list-style-type: none"> Labour market agency 	<ul style="list-style-type: none"> Øresund databank 	<ul style="list-style-type: none"> Estate agency Tourist organisation 	<ul style="list-style-type: none"> Estate organisation Denmark
Data collected by the customer	StatHost	PX-WEB	PX-WEB	PX-WEB
Exemple	<ul style="list-style-type: none"> The Media Secretariat Social Security Appeal Tribunal National Agency for Enterprise and Construction 	-	-	-

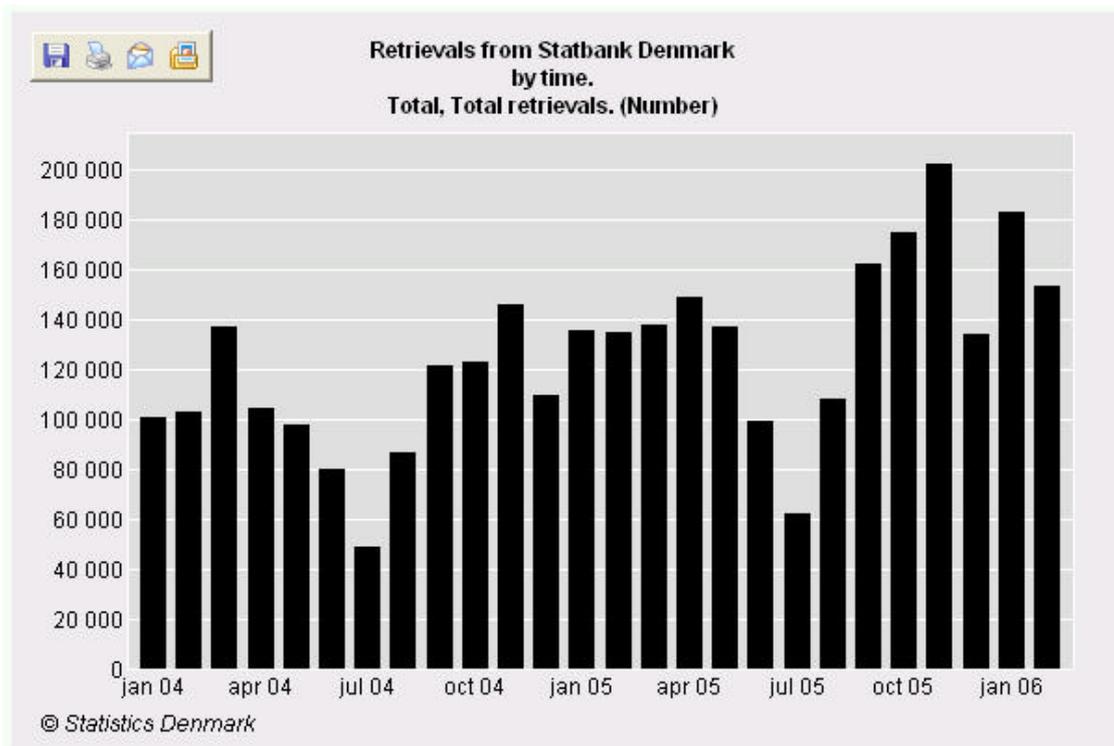
31. Monthly information on *which tables are retrieved - how often and by who* - is registered in a database. The result is made accessible to all internal users in several StatBank tables. The different possibilities of combination are shown below. This is much appreciated by the producers of statistics as a feed-back. Moreover measures on use are part of the Management by Objectives and contract with our Director and the Ministry.

32. Similar systems will be made accessible to the customers of StatHost.

```

Use of StatBank and Web (dst.dk)
├── Use of StatBank
│   ├── STAT1A: Retrievals from Statbank Denmark by producing division, file format, user group and user category. (2001M06-2006M02)
│   ├── STAT2A: Extracts from Statbank Denmark by table, file format, user group and user category. (2001M06-2006M02)
│   ├── STAT3: Stored and possible numbers in tables by producing division, table, status and type. (2002M02-2006M03)
│   ├── STAT4: Number of registered users of Statbank Denmark by points in time and user category. (2001M07-2006M03)
│   ├── STAT5: Tables by producing division and status (2002M06-2006M03)
│   └── STAT6: Burden of updating by producing office, frequency and number of variables. (2003M02-2005M09)
└── Use of Web (dst.dk)
    ├── STAT1SC: Hits on dst.dk by part of website and user (2003M12-2006M02)
    ├── STAT3SC: Hits on dst.dk (detailed) by part of website and user (2003M12-2006M02)
    ├── STAT2SC: Usage of homepage by counting unit. (2003M01-2006M02)
    └── STAT4SC: External hits by site and country (2005M01-2005M07)
  
```

² DST : Statistics Denmark



References:

Lars Knudsen: StatHost - A new concept of renting a Databank, IMAODBC conference 2005.

Annegrete Wulff: StatBank Denmark 2003: Current development and future plans. Statistical Journal, 2003.

Annegrete Wulff: One source-serving multiple needs – Easing data collection of recurrent users. OECD expert group on SDMX, 2004

Annex 1 Examples of StatHost Sites:

<http://www.statbank.dk>

1. This is the official Statistics Denmark databank web interface for statistical dissemination in Danish and English. All data are for free, and user registration is possible for saving queries etc.

The screenshot shows the StatBank Denmark website. The header includes the logo and navigation links for 'London', 'DK', and 'Help'. A search bar is located in the top right. The main content area is titled 'Welcome to StatBank (Denmark)' and contains a list of subjects such as 'Population and elections', 'Education and culture', and 'Labour market'. There is a 'Log on' section with fields for 'Username' and 'Password', and a 'Register as new user' link. A sidebar on the right lists 'As registered user you can...' with options like 'retrieve larger tables' and 'personalize in accordance with your preferences'.

<http://mediesekretariatet.statistikbank.dk>

2. The site is in Danish only.

The screenshot shows the Medie-Statistikbanken website. The header features the logo and a search bar. The main content area is titled 'Medie-Statistikbanken' and contains a list of tables such as 'Forbrug', 'Udsendelser', 'Apparater', and 'Distribution'. There is a 'Find resultat i tabellerne' section with options like 'Færdige tabeller' and 'Forklaring til tabellerne'. A sidebar on the right lists 'Seneste opdateringer' with options like 'FOR410: TV-stationernes andels af udsendelse efter kanal og millimeter' and 'FOR401: TV-stationernes andel af TV-seertiden efter tv-kanal'.

<http://amsbank.statistikbank.dk>

3. This site contains data collected in Statistics Denmark but bought by the customer. The site is in Danish.

Målesystem for beskæftigelsesindsatsen i AF og kommuner Databank - Befolkning og arbejdsmarked

Erhvervs- og beskæftigelsesfrekvens for 1.1 fordelt efter Erhvervsfrekvens / beskæftigelsesfrekvens, bopælsområde, køn, alder og højeste fuldførte uddannelse

Erhvervsfrekvens / Beskæftigelsesfrekvens	Bopælsområde	Køn
Erhvervsfrekvens / Beskæftigelsesfrekvens	Hele landet Beskæftigelsesregion Hovedstaden Beskæftigelsesregion Syddanmark Beskæftigelsesregion Midtjylland Beskæftigelsesregion Nordjylland AF-STORIKØBENHAVN & AF-BOR	Køn i alt Kvinder Mænd

Alder	Højeste Fuldførte Uddannelse	År
16-66 årige i alt 16-19 år 20-24 år 25-29 år 30-34 år 35-39 år	UDDANNELSER I ALT GRUNDSKOLE ALMENE GYMNASIAL UDDANNELSE ERHVERVSGYMNASIAL UDDANNELSE ERHVERVSFAGLIGE PRAKTISK UDDANNELSE Pædagogiske (EF)	2004 2003 2002 2001 2000 1999

Antal valgte tal til tabellen: 1 (Vælg max. 10000) [Vis tabel](#)

<http://dembase.statistikbank.dk>

4. This site is password protected.

DE

Ejendomsalg

Tabeller

- EJ51 Ejendomsalg efter område (kommuner), ejendomskategori, nedlæst og overtagne lejligheder (kvartal) (1992-1 - 2009-3)
- EJ52 Ejendomsalg efter område (kommuner), ejendomskategori, nedlæst og overtagne lejligheder (år) (1992 - 2004)
- EJ53 Ejendomsalg efter område (postnummer), ejendomskategori, nedlæst og overtagne lejligheder (kvartal) (1992-1 - 2009-3)
- EJ54 Ejendomsalg efter område (postnummer), ejendomskategori, nedlæst og overtagne lejligheder (år) (1992 - 2004)