

Platforms for aggregated output databases

Presentation originally developed by Lars Knudsen
(lak@dst.dk)

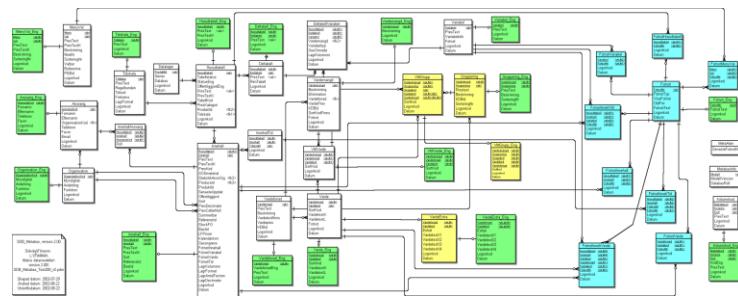


What is an output database?

- A contextual definition:
a collection of statistical information on the Internet, that is organized so that it can easily be accessed, managed, and updated.
- An example: [EuroStat](#)
- Micro or **Macro** data?
- Datawarehouse – Terminology?
- Most are aggregated!



A relational data model (example)



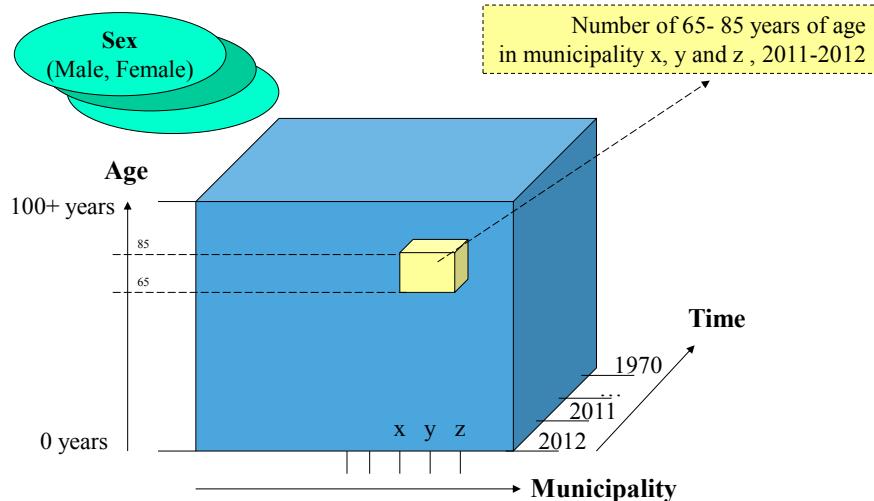
Related Dataset (example)

Kode "M"
for "Male"

MUNICIPALITY	SEX	AGE	CIVILSTATUS	TIME	CONTENT
000	K	16-24	OFF	2000K1	33
000	M	16-24	OFF	2000K1	57
000	TOT	16-24	OFF	2000K1	90
000	K	25-29	OFF	2000K1	204
000	M	25-29	OFF	2000K1	110
000	TOT	25-29	OFF	2000K1	314
000	K	30-34	OFF	2000K1	416
000	M	30-34	OFF	2000K1	183
000	TOT	30-34	OFF	2000K1	599
000	K	35-39	OFF	2000K1	752
000	M	35-39	OFF	2000K1	378
000	TOT	35-39	OFF	2000K1	1130
000	K	40-44	OFF	2000K1	1104
000	M	40-44	OFF	2000K1	477



Slice - and - dice

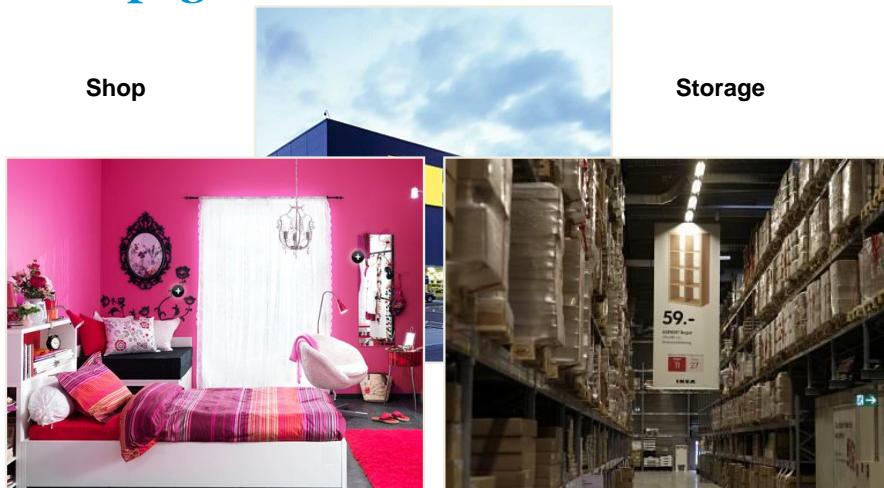


Why use an output database?

- Web is recognized as the main dissemination channel
- Output databases offer tailor-made outputs to versatile users
- No paper limits (size of tables)
- Can offer long time series
- Can offer detailed groupings of geography, industries, commodities etc.
- Feed your publications
- Feed your web site
- Offer API's – programmatic access



Web pages vs. Database



Consistency with publications, web etc.

- One source serving multiple needs
- Prepare publications using (unpublished) data from dissemination system
- Easy updating (demo)?
- Error checking
- Consistent metadata?

Demo: From database to web pages

Example:

- [Statbank](#) ->
- [Web](#) / News release / Publications etc...



Some products on the market

- PX-Web
- Beyond 2020
- SuperWeb
- OECD.Stat
- StatLine (Netherlands)
- TGM (EuroStat)
- Other tailor-made solutions



About the web user interface

- Should not require software installation by users
- Should be user-friendly without additional learning (cognitive perception)
- Support versatile ways of data extraction
- Provide different download formats



Common flow ...

- Define or redefine selection
- Rearrange stub and heading
- Choose multiple output formats
- Save tables / extractions for later use
- Notification of changes / updates
- Additional documentation- included / external links

Some considerations ...

- Should it be free or not?
- Should it contain all NSI data?
- Cater for more than one system?
- Relationship between micro and macro data
- In more than one language?
- Data Cubes versus Time series
- etc...



Demo / examples – Different but same

- <http://epp.eurostat.ec.europa.eu> ...
- <http://stats.oecd.org>
- <http://www.ine.es>
- <http://www.statbank.dk>
- <http://statline.cbs.nl/statweb/?LA=en>
- <http://www.stat.si/pxweb/>



Predefined table before selection?

STATISTICS DENMARK

FOLKI - Population at the first day
Citizenship
Unit Number

Select Advanced selection | More options

REGION (185)
March 12 2014 | more info

All Denmark
Region Hovedstaden
Region Midtjylland
Region Frederiksborg
Region Tønder

MARITAL STATUS (?)
More options.
Married and
Married/separated
Widowed
Divorced
Registered partnership
Surrogate partner

CITIZENSHIP (237)
More options.
Denmark
Albania
Andorra
Angola
Bosnia and Herzegovina
Bulgaria

Number of selected data cells for the table

Producer Prices 2010=100
March 12 2014 | more info

SIM 2008
C Manufacturing

Demographic balance and crude rates
Last update: 11-03-2014
Table Generation: 2014-03-11 10:45:20

geo	time	2004	2005	2006	2007	2008	2009	2010	2011	2012
European Union (27)		421,700,350	426,774,059	426,433,573	426,405,541	406,423,527	392,326,405	381,452,162	386,995,413	384,802,504
Euro area (16)		488,403,375	490,443,738	492,320,886	494,098,517	496,118,936	497,294,605	499,105,158	500,709,793	501,308,522
Euro area (18)		223,612,463	224,890,029	226,502,500	228,042,368	230,016,060	231,402,563	232,212,232	233,078,426	233,244,170
Euro area (18 countries)		319,569,455	321,281,455	322,821,916	324,510,608	326,485,810	327,803,991	328,758,538	329,674,361	329,875,540
Belgium		10,296,411	10,445,652	10,511,382	10,584,334	10,666,864	10,750,060	10,839,953	11,006,438	11,194,452
Bulgaria		7,445,649	7,548,649	7,650,230	7,759,230	7,862,502	7,967,230	8,072,230	8,182,230	8,292,230
Czech Republic		10,398,347	10,198,888	10,233,977	10,294,233	10,345,423	10,410,763	10,482,888	10,488,731	10,503,445
Denmark		5,207,440	5,411,405	5,427,459	5,447,084	5,475,791	5,510,451	5,524,708	5,562,428	5,585,314
Germany (including Bremen)		83,531,471	82,500,849	82,437,171	82,370,905	82,314,904	82,217,837	82,102,398	81,802,257	81,791,402
Iceland		1,000,209	1,000,209	1,000,209	1,000,209	1,000,209	1,000,209	1,000,209	1,000,209	1,000,209
Ireland		4,638,471	4,111,872	4,208,156	4,241,113	4,457,763	4,521,522	4,549,458	4,575,881	4,670,752
Greece		11,037,743	11,073,713	11,112,113	11,143,780	11,182,224	11,196,654	11,193,516	11,123,359	11,123,639
Spain		42,202,228	42,202,228	42,202,228	42,202,228	42,202,228	42,202,228	42,202,228	42,202,228	42,202,228
France (metropolitan)		63,292,241	62,772,870	63,229,833	63,645,065	64,057,393	64,390,238	64,658,858	64,978,731	65,287,861
Portugal		60,925,411	60,963,264	61,399,733	61,795,238	62,200,864	62,405,709	62,765,223	63,076,344	63,378,547
Latvia		37,495,960	37,874,753	38,064,214	38,223,744	38,652,873	39,036,598	39,190,143	39,364,690	39,394,220
Italy		722,859	732,067	744,012	757,918	776,321	786,850	813,146	829,751	862,011
Lithuania		2,205,440	2,205,440	2,205,440	2,205,440	2,205,440	2,205,440	2,205,440	2,205,440	2,205,440
Luxembourg		3,998,919	3,355,202	3,389,835	3,349,963	3,312,408	3,355,856	3,314,876	3,052,988	3,003,644
Hungary		15,159,342	15,457,249	15,676,391	15,686,138	15,695,401	15,695,873	15,624,834	15,144,860	15,012,914

Available flags:
b: Incomplete series c: confidential d: definition differs, see metadata
e: estimated f: forecast g: see metadata
h: not available i: not applicable
j: preliminary estimate k: low quality l: not applicable

Special value:
e: not available

Source of data: Eurostat

Data and MetaData

- What is a Dataset?
 - Only codes and figures
- What is MetaData?
 - Everything else related to data
 - Title, texts, footnotes etc.

Example



Public and private metadata

- Devide access to metadata depending on the nature
- Public metadata:
Centralize adm.
- Private metadata:
Decentralize adm.

	2004	2003	2002	2001
33 years				
Women				
All Denmark	37 267	36 865	38 065	41 068
Copenhagen	4 818	4 547	4 602	4 972
Frederiksberg	878	804	919	939

Before 1 January 2003 Bornholm includes Christiansø.
Unit : Number
Contact : Jeanne Rasmussen jso@dst.dk [+45 39173308](tel:+4539173308)

Need for coordinated metadata

- How to organise database administration
- Common- and table specific metadata (Public and private)
- Coherence between tables and valuesets
- Policies are important for success



Organisational structures

- How to organise database administration
- Common- and table specific metadata (Public and private)
- Coherence between tables and valuesets
- Policies are important for success



Key word:

1 000 DKK	1 000 DKK.	DKK 1 000	Dkk. 1 000	Thousand DKK.
1.000 DKK	1.000 DKK.	DKK 1.000	Dkk. 1.000	Thousand DKK
1000 DKK	1000 DKK.	DKK 1000	Dkk. 1000	Thousand Dkk
1 000 Dkk	1 000 dkk	DKK. 1 000	dkk 1 000	Thousand Dkk.
1.000 Dkk	1.000 dkk	DKK. 1.000	dkk 1.000	thousand DKK
1000 Dkk	1000 dkk	DKK. 1000	dkk 1000	thousand DKK.
1 000 Dkk.	1 000 dkk.	DKK 1 000	Dkk. 1 000	thousand Dkk
1.000 Dkk.	1.000 dkk.	DKK 1.000	Dkk. 1.000	thousand Dkk.
1000 Dkk.	1000 dkk.	DKK 1000	Dkk 1000	thousand dkk
				thousand dkk.

Standardization



Centralization

The screenshot shows a web-based statistical reporting system. At the top left is the Statistics Netherlands logo. Below it is a search bar labeled "StatLine" and "Search term". To the right of the search bar are buttons for "theme" and "map". The main content area has a title "Bankruptcies; monthly figures by legal form and activity SIC" and a subtitle "March 12 2014 | more info". A table titled "Bankruptcies by legal form" displays data for various months in 2013. The table includes columns for "Total pronounced bankruptcies", "Single-owner comp., trading partnership bankruptcies", "Single-owner companies", "Trading partnerships", and "Natur Total bankr". The data shows values such as 1 041 for June 2013, 2 464 for July 2013, and 873 for February 2014. On the right side of the table, there is a vertical sidebar with sections for "Subjects" and "Periods", listing months from "2013 juni**" to "2013 februar*". The bottom of the page features a blue footer with the Statistics Denmark logo and the number "21".

Periods	number	Total pronounced bankruptcies	Single-owner comp., trading partnership bankruptcies	Single-owner companies	Trading partnerships	Natur Total bankr
2013 juni**		1 041	763	88	675	
2013 2e kwartaal**		3 247	2 464	297	2 167	
2013 juli**		1 146	897	117	780	
2013 augustus**		866	687	79	608	
2013 september**		875	682	86	596	
2013 3e kwartaal**		2 887	2 266	282	1 984	
2013 oktobe**		1 116	844	100	744	
2013 november**		903	687	105	582	
2013 december**		887	671	76	595	
2013 4e kwartaal**		2 906	2 202	281	1 921	
2013**		12 306	9 456	1 181	8 275	
2014 januari**		948	707	112	595	
2014 februar*		873	684	78	606	

Interfaces to output databases

- Graphical User Interfaces (GUI's)
 - web and other software user interfaces
 - Platform challenges

- Application Programming Interfaces (API's)
- Web Services (SOAP, Rest)
- E-mail services
- SMS services
- Etc ...



Devices- I

The left screenshot shows the Statistics Denmark mobile site for 'INDAMP01: All population by region, group of persons, key figures'. It displays a hierarchical menu for 'REGION (116)' and 'GROUP OF PERSONS (10)'. The right screenshot shows the Statbank DK mobile site for 'The Mobile Statbank - Live births by age of mother, live-birth order and time'. It features a search bar for 'Age of mother' with options from 10 to 45 years, and a bar chart at the bottom.

Devices II

The left screenshot shows the CBS StatLine mobile site for 'Bankruptcies; monthly figures by legal form and activity SIC'. It displays a table of bankruptcies by legal form for various months in 2013 and 2014. The right screenshot shows the CBS Netherlands mobile site for 'Birth; age mother (on 31 December)'. It displays a summary table with counts for 2012, 2011, 2010, and 2009, along with a bar chart at the bottom.