

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

UNECE Work Session on Statistical Dissemination and Communication
(Geneva, 13-15 May 2008)

Topic: Managing communication and dissemination

**IMPROVEMENTS TO THE GOVERNANCE AND DISSEMINATION PROCESS OF STATLINE,
AT STATISTICS NETHERLANDS**

Submitted by Statistics Netherlands¹

I. INTRODUCTION

1. All statistics published by Statistics Netherlands can be found in what has become one of the largest public databases of the Netherlands, the StatLine database. This database comprises more than 4 billion data cells containing around 400 million statistical facts. It can be accessed free of charge through the Internet. However, because of the size of the database and the complexity of statistical classifications used, people were struggling with accessibility. Users could not find the data they needed. The text - or metadata - of the statistics was not easily comprehensible. The internal dissemination process was inefficient. The process and the software used for the process reached their limits, as the number of statistics and data now far exceed the capacity for which the original process was designed. Also the maintainability of some of the software - which dates from the 1990s - was poor.
2. In the face of these challenges, and given that a more ambitious earlier project (StatLine 4) was unsuccessful, Statistics Netherlands set up a programme called StatLine CoOP, in which *CoOP* stands for cooperation but is also an abbreviation for the three projects it comprises:
 - Achieving Coordination level 1, in which the comprehensibility and uniformity of the statistics have been improved;
 - Improving the Output (dissemination) process, in which the governance of the communication and the dissemination process has been redesigned, the software has been and will be upgraded, a preview to a new or updated table is presented, and several other improvements in the dissemination process have been made;
 - Improving the Presentation of statistics, in which the layer around the database has been completely revised (StatWeb 5) to make it more user-friendly.
3. These projects will be discussed in the paper and presentation. A demonstration of the new presentation software will be given. Future developments within statistics Netherlands and the lessons in project management approach will be presented.

II. ACHIEVING COORDINATION LEVEL 1

4. Why are some of the figures produced by Statistics Netherlands hard to understand? Why is it that when searching for a word you cannot determine which of the results found to choose?
5. A part of the explanation can be found in the contents: the text or meta information of the table is unclear or not easy to understand; another part can be found in the search algorithm. The first part was investigated more thoroughly. The answers were:

¹ Prepared by Ran van den Boom (cbmm@cbs.nl), Roel Canisius (rcis@cbs.nl) and Michiel Eijkhout (mekt@cbs.nl), with review contributions of Olav ten Bosch (obos@cbs.nl) and Edwin de Jonge (ejne@cbs.nl).

- Abbreviations are used in the title of the table, in the explanation of the table and in explanation of the labels. Most of these abbreviations are unnecessary, incomprehensible or should have been explained. These include “MAO” (mestafzetovereenkomsten, fertiliser disposal agreements), “GDP” (Gross Domestic Product), “HOMALS” (homogeneity analysis by means of alternating least squares).
 - Officialesse, difficult, solemn and old-fashioned words such as “plausible”, “inherent”, “substantial”, “vacant”, “multifarious”, which could easily be left out or replaced. Note that the tables translated into English did not suffer from this.
 - Technical words such as paulianeus, cargadoor (ship-broker), effluent, polypatriden, entrepot (bonded warehouse), which did not have an explanation. Perhaps only an expert knows what they mean. For some of them it is even unclear whether they are existing Dutch words, as they could not be found in the dictionary.
6. What is a comprehensible text? How can you define the level of understanding? What to aim at when rewriting texts? Together with the statistical departments we decided the level of understanding should be at the level of viewers of the Dutch television news programs, which is about fourth grade in secondary school [havo-4]. In pilot projects, a number of tables were investigated and the estimations indicated that one table would take an average of 8 hours to review, rewrite and replace.
7. In parallel, a review was held during Summer 2007 on a thousand of the active tables by students at havo-4. “Active” means tables that are not discontinued and not incidental. This led to various observations for each of the tables, and a list of words that were incomprehensible to students at the havo-4 level.
8. To prevent the pitfalls from the past, we chose a bottom-up approach. Representatives of each of the seven statistical departments, together with a representative of the dissemination department, were placed in a board (“Points of support”) in which decisions were made on subjects such as guidelines, templates to be used, and the text for classification categories and topics. The project team enabled this process, and was the single point of contact for questions and issues. To facilitate the re-use of text (reference text), the “Level 1 Tool” was developed that enabled using reference text in a table. The Editorial Board was put in place to make the final decisions and for escalation purposes and management support. It consisted of the managers of the departments. Issues raised in the field were answered by the project team. If the answer required discussion with the points of support, the issue was raised in the weekly meetings with the “Points of support”. This guaranteed that each decision was supported by all departments; it also made the process of decision making rather long. As noted, pilots revealed that an average of 8 hours would be required to modify a table. To limit the scope of the project, only tables that were active in the Dutch language would be modified: tables in English, in the archive, or occasional tables would not have to be improved. This reduced the number of tables to improve from 1700 to 1000. Even so, the departments had to make a major effort since the work was not foreseen in the plans for 2007.
9. December 2007 was set as the completion date. However, the guidelines were not stable at the beginning, tooling was not ready, and it took the departments some time to get things moving. From November 2007 onwards the pace increased and each week we presented the group that had made the best effort in improving one or more tables with a pie. As a result, by the end of March, 900 tables had been reviewed, rewritten and replaced. Two departments are still working. The main results of the project were:
- 900 tables have been improved.
 - The number of abbreviations has been reduced to a minimum.
 - The explanation of tables and labels has been improved.
 - The awareness of the quality of StatLine tables has significantly improved. Also management is now aware that quality is a major issue.
 - New guidelines, supported by the departments, have been made, along with a checklist to use for updated tables and several courses.
 - Reference texts have been developed for 15 of the most used classifications such as SIC (Standard Industrial Classification), Region, Gender, Marital status, and Political Parties.
 - Colleagues can preview the resulting table on StatLine before the table is actually published.

- A tool has been developed to get an overview of most of the StatLine meta.
10. There were also goals that we did not meet:
- The explanation in some cases is beyond “havo-4” level and some explanations are too lengthy.
 - The guidelines were seldom completely applied. For instance titles in most cases do not adhere to the guideline the main subject of the table should be named first.
 - Inconsistencies still remain since data are stored in multiple places. For instance, the update frequency is located in the summary text of the table and in the general explanation. The description of the methods used in calculating the data sometimes is not consistent with other information.
 - The tools were less successful than planned.
11. For the future, we will keep improving the quality of the tables, the guidelines, courses and explanation on the guidelines and tools. The initiatives that originate from within the departments will be facilitated. The most important issue for the future is not striving for higher coordination levels, per se, but to embed the awareness and continuous improvement of table quality within the departments.

III. IMPROVING THE OUTPUT (DISSEMINATION) PROCESS

12. The dissemination process was a spaghetti process, containing many laborious steps. There was no integrated software solution to support the dissemination process. For instance, to report an upcoming table, one had to download a template, fill it in, and fax it. The dissemination department typed the information on the print in a software system, printed the information, handed it over to a person controlling the table, et cetera et cetera. Moreover, the table information was stored in several places, the table data was stored in more than ten places and it was manually copied from one place to another. This was time-consuming, inefficient, and led to inconsistencies and errors. More than 20 software systems were used for this process. The IT department could not guarantee that some of these systems would work in the near future, sometimes the problems could be solved but the cause could not be determined. Time for improvements!

Figure 1: Current dissemination process

Statistical department

- Announce upcoming table
- Design and build table

Dissemination administration

- Print announcement of the table
- Check process metadata, type over
- Contact statistical department in case of problems
- Copy files

Dissemination editors

- Check files
- Check against guidelines
- Contact statistical department in case of problems
- Check
- Copy files
- Update management reports
- Copy files

Dissemination administration

- Check
- Copy files
- Register data
- Copy files for publication
- Update publication tree
- Batches for updating and copying
- Administrate
- Archive

13. Our first step was an analysis of the most irritating problems. These were tackled by quick-wins. Some of the examples are:
- One template that was hard to fill in, was simplified and could be sent by e-mail.
 - Previewing the table not yet published was made possible.
 - Copying the tables to the Internet was troublesome and took a long time due to the huge amount of data. This was adapted so that the deltas were copied instead of all the data.
14. Second, a more thorough analysis showed which long-term improvements were required. A new dissemination process -- a new governance model -- was developed. Main pillars for this model were:

- A statistical department, not the dissemination department, should be responsible for the quality of its tables.
- The dissemination department is responsible for facilitating the process, checking whether the process is correctly followed and is responsible for improving the process.
- Roles in the process should be clearly defined, and sometimes these roles can be combined
- Checking the meta and figures should be done as early in the process as possible, not at the end.
- Data should be stored in one place, so that there are no inconsistencies.
- The software systems should be maintainable for at least 2012, and the number of systems should be reduced to the minimum.

15. This new governance model led to the development of StatFlow, a workflow system that enables storing the table data and meta: both process meta such as the creation date and the approval date, as well as conceptual meta such as the title, frequency and population. Second, the system allows previewing all tables in the database, and third, publishing the data on the Internet will become much simpler. The development of this system started in April 2008. Much effort has been put into getting the statistical departments involved, to make sure that they will use the system and that they will improve the data quality. Figure 2 shows the new process.

Figure 2: New dissemination process

This process implies more responsibility for statistical departments. The state of the table is set after each modification and check.

Statistical department and dissemination editors

Design and build meta of the table

Statistical department

Announce table and meta

Build the table and store

Preview, check, report ready for next step

Dissemination editors

Check against guidelines

Contact statistical department in case of problems

Statistical department

Final preview and check, report ready

Publish automatically

Dissemination administration

Administrate

16. Different governance models of communication and dissemination have been tried out at Statistics Netherlands. It is all about roles that can or cannot be combined, and roles that should be played by people from one or more departments. Who is responsible for the initial quality? Who is checking it? Should that be someone of the same group or department? Who has the knowledge to develop tables? Should that knowledge be centralised? In the previous dissemination process the dissemination department performed the final quality check. They were seen as the annoying “StatLine police”. In the new process, they play a consulting role. One of the first results is the improving relationship with the statistical departments.

The importance of communication cannot be underestimated in the process. Due to the large number of persons involved (more than 200), there should be continuous meetings at all levels in the organisation: from the work floor to high-level management of Statistics Netherlands, not only during the project, but throughout the regular work as well. As a result, the dissemination department will now have monthly meetings with representatives of the statistical department.

IV. IMPROVING THE PRESENTATION OF STATISTICS

17. For the last 12 years the StatLine database has been disclosed with the in-house-made software, StatWeb. The previous version StatWeb 3 was in production from 2001 and was successful. The Statistics Netherlands (SN) website contains many links to StatLine directing to table selections made by StatLine editors.

18. StatWeb 3, however, was out-dated and had user interaction problems. People visiting StatLine suffered a number of problems:

- StatLine was not part of the SN website. StatLine was experienced as a different website and not as part of the overall website of Statistics Netherlands.

- The look and feel of StatLine was outdated
- Users got lost when dropped in StatLine. Especially for inexperienced users the user interface could be improved significantly. 85% of the users enter StatLine through a hyperlink, arriving in a screen showing a table without much explanation on what to do next.
- Users could not adjust a StatLine table due to technical problems. A Java applet was required for showing the tables that could be selected and the variables selection tree. This applet was difficult to download, or was sometimes even prohibited. The pop-up screen (the so-called 'webselector') did not pop-up, disappearing behind the main screen and proved to be buggy.



19. The following solutions were chosen:

- StatWeb 5 has the same top-navigation as the SN Website and is now a part of the overall website.
- In StatWeb 5 the icons, colours and layout were modified to have a modern looking web application.
- StatWeb 5 improves navigation and adds more explanations for inexperienced users.
- StatWeb 5 replaced the Java applet with a more commonly available JavaScript/HTML solution.

20. The StatWeb 3 layer that presented the StatLine 3 data had to be adapted. Research showed the StatWeb 4 presentation layer (originating from the StatLine 4 project, which was not implemented) could be mounted on the StatLine 3 with a minimum of effort. Rewriting the layer and migrating the StatLine 3 data format proved too expensive, time-consuming and troublesome.

21. However, while programming, it turned out that more modifications were required. The StatWeb 4 engine was modified to speed up performance, and the usability suggestions from three usability reports were applied to make StatWeb even more user-friendly. The updated and improved software was called StatWeb 5.

22. A live demonstration of the new StatWeb 5 can be given. Here follows a short tour in case a user enters the main page of StatLine.

First select a table:

StatLine database > Tables by theme

Select a table 

Browse through themes and choose a table select

Themes	Population	Marriages and marria...
Education	Births and deaths	Marriages; key figures
Enterprises	Marriages and marriage dissolutions	Marriage dissolutions; key figures
Financial and business services	Migration and migrants	
Government and politics	Population and household	
Health and welfare		
Income and spending		
International trade		
Labour and social security		
Leisure and culture		
Macro-economics		
Manufacturing and energy		
Nature and environment		
Population		
Prices		
Security and justice		

Second, a user can select which topics and classifications are to be shown in the table:

StatLine database > Tables by theme > Select data

Key figures marriages and partnership registrations 

Select the desired data for every tab. show data

Subjects (1/48) Periods (8/57)

Subjects	Marriages
Marriages	<input checked="" type="checkbox"/> Marriages total
Marriage partners	<input type="checkbox"/> Marriages per 1,000 inhabitants
Average age at marriage	<input type="checkbox"/> Marriages between a man and a woman
Average number of marriages per person	<input type="checkbox"/> Marriages between two men
Partnership registrations	<input type="checkbox"/> Marriages between two women

After confirming the selected variables, the table is shown:

StatLine database > Tables by theme > Select data > Show data

Key figures marriages and partnership registrations

Table explanation Changed on 27 August 2007 Frequency: yearly

Search: | Theme



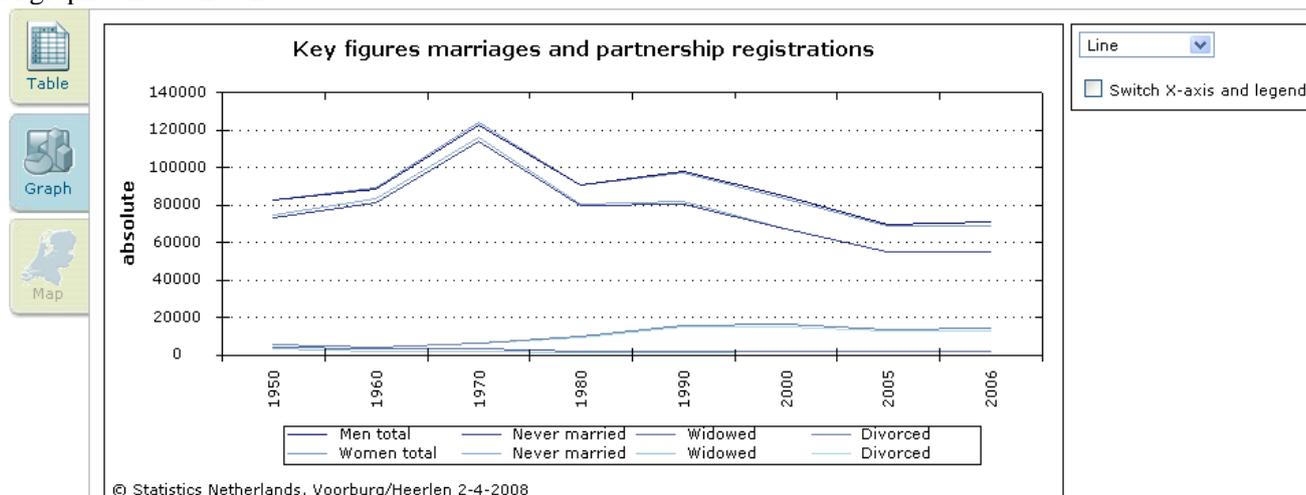




Subjects				Periods		1950	1960	1970	1980	1990	2000	2005	2006
Marriage partners	Absolute	Men	Men total	<i>absolute</i>	82 389	88 016	122 287	90 375	97 321	84 573	69 493	70 335	
			Never married		72 941	80 880	113 399	79 090	80 409	66 957	54 555	54 998	
			Widowed		4 053	3 221	3 187	1 794	1 270	1 811	1 670	1 739	
		Divorced		5 395	3 915	5 701	9 491	15 642	15 805	13 268	13 598		
		Women	Women total		82 572	88 713	123 521	89 978	97 148	82 773	68 200	68 731	
			Never married		74 620	83 014	116 141	80 307	81 476	66 769	54 347	54 612	
	Widowed			2 976	1 785	1 812	1 154	970	1 441	1 372	1 404		
	Relative	Men	Men total	<i>per 1,000 men 15+</i>	55.1	59.9	73.2	42.5	35.9	29.5	22.4	22.4	
			Never married	<i>per 1,000 unmarried men 15+</i>	58.0	64.1	78.4	45.1	37.9	28.8	21.8	21.7	
			Widowed	<i>per 1,000 widowed men</i>	29.2	22.8	21.9	12.2	8.4	11.1	9.8	10.2	
		Divorced	<i>per 1,000 divorced men</i>	182.8	123.8	127.0	67.5	56.9	41.7	31.3	31.4		
		Women	Women total	<i>per 1,000 women 15+</i>	52.5	56.3	69.6	40.0	33.8	26.6	20.4	20.3	
Never married			<i>per 1,000 unmarried women 15+</i>	63.3	72.4	95.5	56.4	46.9	35.2	26.1	25.8		
Widowed	<i>per 1,000 widowed women</i>		10.0	5.0	3.8	1.9	1.4	2.0	1.9	2.0			
Divorced	<i>per 1,000 divorced women</i>	108.5	73.5	76.1	45.4	41.8	29.4	22.2	22.2				

© Statistics Netherlands, Voorburg/Heerlen 2-4-2008

A graph can be shown:



In case a user directly starts with a table (such as entering through a hyperlink), the user can change the selection or select another theme or table through the path shown on top of the page.

23. The main differences with StatWeb 3 are:

- The user interface has been improved; modifying and selecting a table requires less user interaction, such as mouse clicks; selecting does not require pop-ups.
- StatWeb 5 has the same top-navigation as the SN Website and is now a part of the overall website.
- When entering StatLine through a hyperlink, the context is clear and the user can easily see how to retrieve more information on the table or theme.
- The meaning of the icons are clear and meaningful.
- The performance and scalability has been improved.

24. A number of challenges still remain:

- StatWeb 5 uses the Adobe SVG viewer plug-in. Adobe will no longer provide support for Adobe SVG Viewer as of January 1, 2009. The SVG viewer does not function correctly in StatWeb for Firefox and Safari. Similar to the Java applet plug-in, downloading this component proves troublesome for some users - even some PCs at Statistics Netherlands do not have it. Moreover, access to statistics through a picture can be made easier and more attractive.
- Some of the StatWeb 3 functionality is not available in StatWeb 5. In StatWeb 3 one could modify the layout of the table, so that instead of showing classification categories such as years as rows, they can be selected individually. Also additional functionality is suggested, making StatLine more attractive for inexperienced users. A StatLine light perhaps? CBS is interested in hearing how other statistical institutes promote their statistics.
- Webservices allow the use of StatLine data without the StatWeb layout. Statistics Netherlands has started with a first try-out version for internal test purposes. Questions such as performance impact still need to be answered.
- These requests require IT budget and capacity, which are hard to get these days.

V. LESSONS LEARNED IN PROJECT MANAGEMENT

25. Why is this programme more successful than its predecessor? Even with these projects not everything went well, according to plan or according to what we hoped the effect of our actions would be. The quality of the text improvements is not as good as we hoped it would be, for instance.

26. Critical success factors for managing projects are described in the CHAOS reports of the Standish Group International Inc. These also apply to the projects within the StatLine CoOP programme and include:

- Involvement of the management; when it comes to making a choice for one project or another, they have to choose to support your project.
- Involvement of the end-user in defining the requirements and acceptance criteria. The guiding principle is that if the end-user is not able to help, you do not have a project.
- An experienced project manager, a multi-talented champion
- Focus on a clear main goal, leaving things that are not necessary aside. This also means breaking the project down into small parts. Each must have a result that can be implemented in the organization. We have done so in dividing the improvements into three projects, each of which is divided into separate tasks.

27. There are a number of lessons learned from these projects:

- Communication in organization-wide projects is highly important; it starts by listening very carefully to the problems and extracting the underlying message. Communication in organization-wide projects should be well organized, structured and extensive: all management lines should be incorporated in the communication.
- The bottom-up approach works
- Have professionals in your project teams
- Be pro-active in managing risks, in sticking points, facilitate in time; stay one step ahead of things
- Have your facilitating products, such as plan for approach, guidelines, courses and tooling ready by the start of the project; try them in pilots so that they are well-tested beforehand
- Adapting almost a thousand of the most frequently used tables takes over 16 hours per table on average (double what was anticipated)
- Scope creep when making text improvements can be expected, unless you use a way to measure when a text is good enough and then hold it there.
- Would the results have been better if each table was reviewed after it was improved? And if the dissemination departments could have been “ordered” to spend more time on quality? At what costs?

28. Finally, we will continue with the three pillars of StatLine CoOP:

- Achieving Coordination level 1 is continued up to the summer 2008 for further improving tables and guidelines. The future project Achieving Coordination level 2 will be started on a small scale and in pilots to test out the ideas and tools. A new tool will be developed for reference texts for tables.
- Improving the Output process will be continued up to Quarter 3, 2008.
- Improving the Presentation of statistics will be continued with maintenance releases of StatWeb 5, with webservices and with replacing the mapping module.

29. The main issues now are how to obtain the resources and the budget to accomplish these goals.
