**⦀ Statistics Finland**

Nordisk Statistikermøde i København 11.-14. august 2010
Tema 4: Formidlingen
Riia Arvela (riia.arvela@stat.fi)
Dissemination and databases

# Effect of the revised publishing system on the development of the contents of Statistics Finland's free data supply

**Abstract:** Over the 2009-2010 period, Statistics Finland will start to publish its statistics comprehensively with a new FastWeb-XML-based publishing system suited for multi-channel distribution. When the new system is introduced, free data supply on the Internet will be extended and at the same time, the division of data into different publishing modes will change.

## 1. Progress of the publishing system renewal

### 1.1 Change from the old publishing system to the new one

Over the 2009-2010 period, Statistics Finland will start to publish its statistics comprehensively with a new FastWeb-XML-based publishing system suited for multi-channel distribution. When the new system is introduced, free data supply on the Internet will be extended and at the same time, the division of data into different publishing modes will change.

In Statistics Finland's old publishing mode a printed publication, PDF and HTML were prepared with totally separate processes. Tables were also made to the database service with different processes and their content was not thought to be necessarily connected to other publishing at all.

In the new XML publishing process different publishing modes (PDF, HTML, printed) are automatically formed from one XML file created with the Arbortext editor. The main focus has simultaneously shifted to electronic publishing. The new basic publication combines the statistical release made in the old system and the related appendix tables and figures and the printed publication. The basic publication is formed of HTML pages[1] and of a free PDF publication[2] with a fully corresponding content. If so desired, it is possible to produce from the basic publication a printed, chargeable version; a printed original (pdf) is created in the same process as the HTML and web PDF.

The introduction of the new publishing system is still in progress. By June 2010, a total of 160 sets of statistics out of Statistics Finland's 200 statistics have moved over to the new system. The work first started as implementation of the new technology, but as the transition advanced, it was found necessary to develop both the structure and content of publications and publishing at the same time. To carry out the transition a support project was set up for organising training and support to producers of statistics in implementing the renewal.

---

[1] Prices of dwellings, HTML: http://tilastokeskus.fi/til/ashi/2010/01/ashi_2010_01_2010-04-30_tie_001_en.html
[2] Prices of dwellings, PDF: http://tilastokeskus.fi/til/ashi/2010/01/ashi_2010_01_2010-04-30_en.pdf
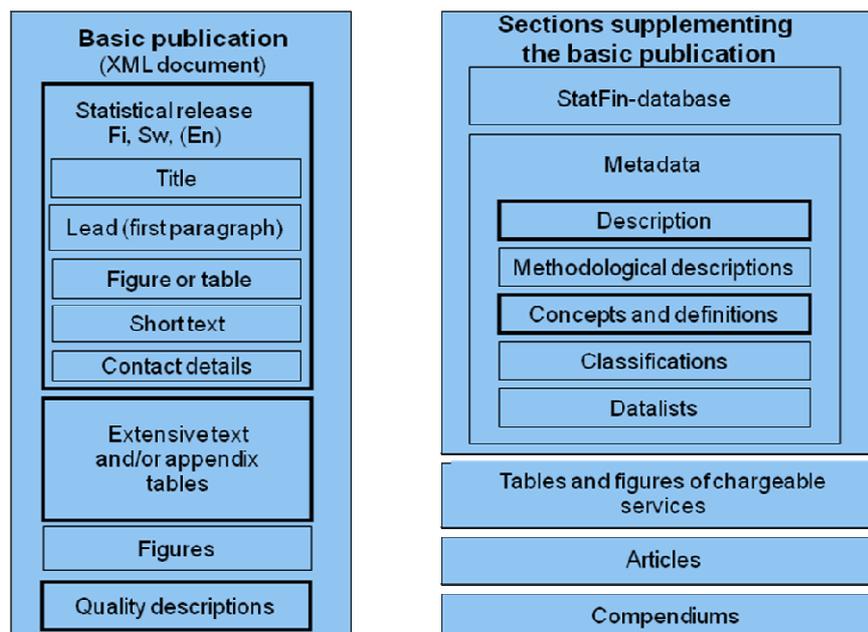
# Statistics Finland

## 1.2. Re-division of content

The change of technology has brought about a re-division of the content especially as concerns tables. The old printed publication had an important role in data dissemination and it contained large tables for several statistics. Before the introduction of the new publishing system, free of charge distribution of statistical data was increased by widening the content of the PC-Axis file-based PX-Web service (StatFin). With the renewal of the publishing system StatFin has also enlarged further. Because the publication is automatically transformed into HTML and PDF files, large tables made for the printed publication are no longer produced. Earlier in connection with the release of data, Excel tables were also published, but now they are converted into tables suited for an XML publication or bigger tables are converted into PC-Axis files and added to StatFin.

For many sets of statistics the change is sizeable and it requires re-planning the content of the publication and an appropriate division of data between the database tables and the basic publication.

In Statistics Finland's web service statistical releases are always available at least in Finnish and Swedish. Releases of the major statistics on economic trends are also published in English. English-language and Swedish-language releases may be more condensed. At the moment the aim is to make all Finnish releases in the same extent as the actual basic publication (Figure 1.). The number of statistical releases in different languages is meant to grow in future.

**Figure 1. Connections of the basic publication to the rest of the website**

# Statistics Finland

## 1.3. Challenges met when moving to the new system

### 1.3.1. Challenges connected to the production process

The transition to the new publishing system was started as a mere technical change, but quite soon it was seen necessary to change the content of the publication and the division of its sections as well. Many producers of publications wanted to transfer the old publication as such to the new system. That is not always possible for either technical or content reasons.
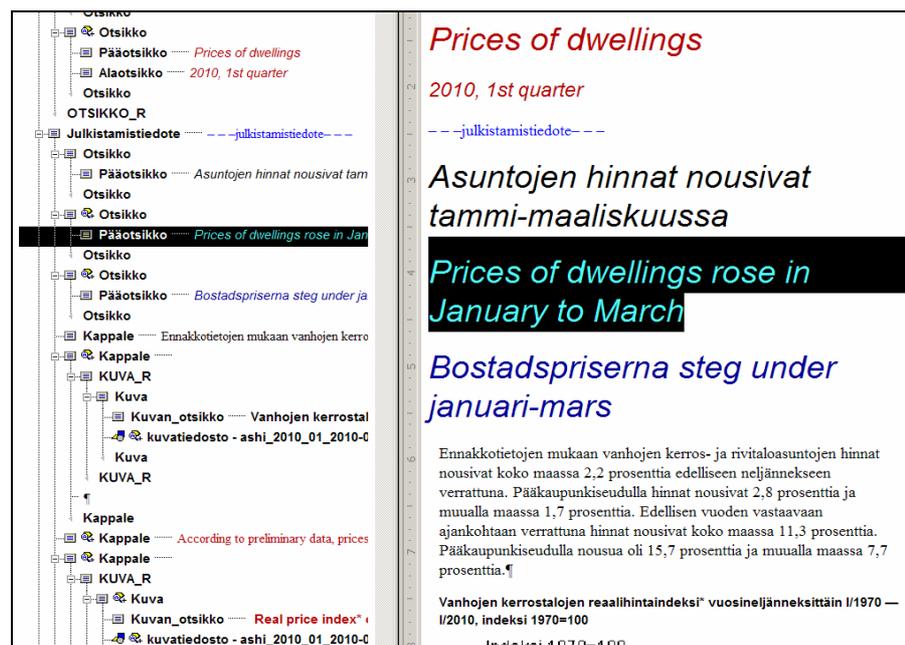
The text part of an XML document can be written either with Arbortext or a text written with Word can be copied to Arbortext. Tables are attached to Arbortext as XML/CALS files. For the time being, figures are appended to Arbortext as gif images[3].

As in the earlier production mode, figures are produced with Excel or Harvard Graphics. SAS was taken into use as a new production tool for figures, with which tables and figures can be made from the same data. To help figure making with SASA an Excel file was developed, from which parameter data can read by SAS programming to change figure data. This use is under further development.

SAS, SuperStar and PX-Edit are used for table production. The same production modes are also used for producing database tables.

Differences in statistics production systems have brought challenges to the adoption of the publishing system. The idea is to increase the use of SAS so that table and figure production could be better automated.
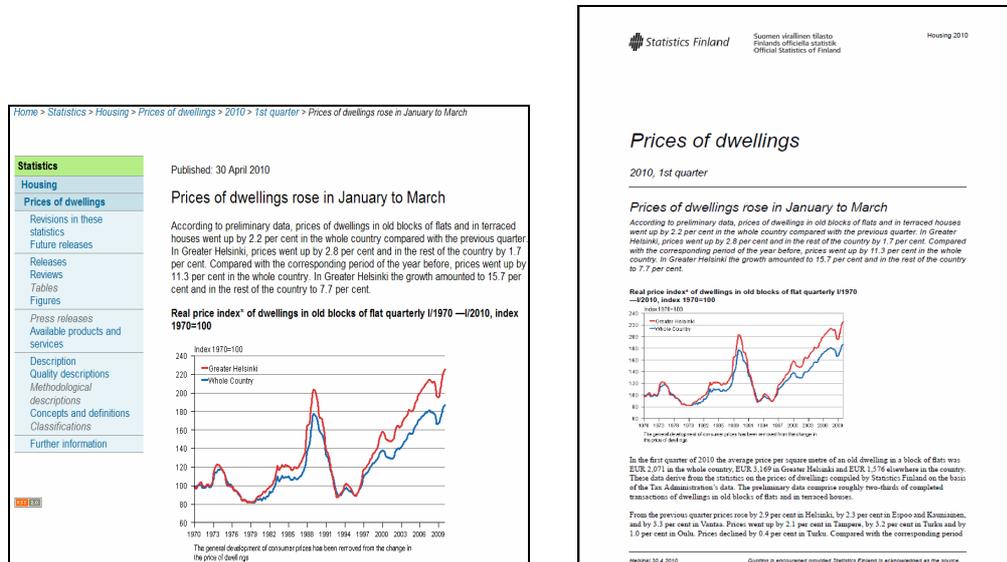
**Figure 2.  Arbortext editor interface**



---

[3] The production is prepared to make SVG images. Because of non-standardisation of browsers this it not in use.

**Figure 3. Basic publication HTML and PDF**



## 1.3.2. Standardisation, training and support

Production of XML documents, on which the publication is based, is decentralised to producers of statistics. It has been challenging to standardise the structure of the publication, although an open general editor has been in use and there are many producers. Standardisation of the structure is important so that the XML document can be utilised effectively by means of automation and a high-quality publication can be produced.

There are no technical means in the XML editor to make producers come up with a high-quality result, although the content of the publication is treated as structured elements in the editor. A high-quality contentual result can only be produced when the possibilities and limitations of technical tools are known. Harmonisation requires interaction and extensive training.

Producers of an annual or quarterly publication use the new editor seldom, so learning the new technology and using it effortlessly has at first been found difficult. To make the work run well many statistical units have reconsidered their division of work in producing the publication. Some units have solved it so that one person compiles publications for several statistics so that the tool can be used efficiently. In some units the person responsible for the statistics concerned makes the publication alone.

The aim of the training programme designed for the implementation of the new publishing system is to provide the large group of producers with the required technical competence and skills to construct the content of the publication. Training has concerned publication standards and work processes, planning of tables and publications and attaching database tables as part of the whole publication. Technical training has been given on PC-Axis and PX-Edit, SAS tabulation and graphics and use of the publication editor (Table 1). Instructions were prepared as to the order in which the courses should be taken according to the work roles, in order that the right people would be

on the courses. At first there were no such instructions and there was uncertainty about the necessity of the courses. Except for one course, all course instructors were from Statistics Finland.

Table 1. The training programme of the new publishing system

| Name of the course | Duration in hours | Number of courses 2009-2010 | Training provider |
|---|---|---|---|
| Publication standards and work process | 2.5 | 7 | Statistics Finland |
| StatFin as part of publishing | 2.5 | 3 | Statistics Finland |
| PC-Axis | 4 | 5 | Statistics Finland |
| PX-Edit | 5 | 9 | Statistics Finland |
| SAS tabulation | 5 | 9 | Statistics Finland |
| SAS graphics | 3 | 6 | Statistics Finland |
| Planning of tables and publications | 3 | 7 | Statistics Finland |
| Use of the publication editor (basics in Arbortext) | 7 | 11 | External instructor |
| Production and publishing of the statistical publication (Arbortext and Webajastin) | 7 | 9 | Statistics Finland |

In autumn 2010 all the courses will be arranged for the rest of those transferring to the new system. In addition, three content-oriented courses will be combined into one course.

In addition to training, producers also need support. There is one full-time person providing producers of publications with technical support in the use of the publication editor and four others besides other work. Although the making of publications is decentralised to producers of statistics, the progress of publications is monitored in a centralised manner through publication documents saved in the XML base and where needed, support is given before publishing. One person works full-time at this, four persons besides other work. In addition, there are several support persons for different tools in table production.

Instructions have been supplemented throughout the system change.

## 2. Practical experiences of the renewal from the viewpoint of statistics producers

### 2.1. Prices of dwellings

The content of the old publication on statistics on prices of dwellings was dominated by a 32-page long table and the publication also had many other several page-long tables. In connection with the change of the publishing system the tables connected to the publication were re-planned and large tables were placed in the StatFin database. Eight smaller tables remained in the publication showing the latest released data on the indices and change percentages. More time series had been used in the old publication. No changes were made to the figures.

Now the publication is produced in about two days' shorter time than during the old system. The database tables are made with SAS, so it requires one run.

Statistics Finland has produced six publications with the new system. As experience grows, the work already progresses well and improvements are being planned to the data content. The training related to the renewal was considered good and support has been well available when needed. No changes needed to be made to the division of work in connection with the renewal.

Plenty of feedback was first received from customers about the move of tables to the database but as guidance was given to customers it has decreased. Feedback is mostly from older users whose IT skills are not necessarily sufficient for using the database without guidance. Inquiries of the release date from journalists have decreased in comparison with the old publication, as it was learnt how to use the StatFin database.

### 2.2. Education statistics

The education statistics unit got well prepared for the change in advance. Before the changeover started, the unit's all statistics had, in addition to other training, requested a joint event where the changeover was discussed with regard to the production techniques and special features of the statistics. A few employees became more acquainted with the technical support of publications and made a model template for the use of the unit's statistics, for example. Before the new publishing system was introduced, education statistics had just taken their tables to the StatFin database.

Each person responsible for the statistics in the unit prepares the publication totally alone. The statistics are annual statistics. Earlier the statistics had Excel tables on the web. Their data contents were transferred to the StatFin database or into HTML tables. Due to the changeover of the publishing system, some of these HTML tables were moved to the new publication, some were still added to the StatFin database. Figures are not used in most of the unit's statistics. In the first publishing round of the new system the content remained narrow for some statistics, but it is meant to be extended in the following publications.

**⫙⫙⫙ Statistics Finland**

Nordisk Statistikermøde i København 11.-14. august 2010
Tema 4: Formidlingen
Riia Arvela (riia.arvela@stat.fi)
Dissemination and databases

The tables of the statistics are made with SAS mainly from the production database, but not all information is easily available from there. This has resulted in compilation of tables with other tools and the work has been slower than before. Although the producers have long experience of using SAS, more varied skills would have been needed in connection with the change. Now skills in using several different tools were required. The education statistics unit felt the tool had set restrictions on content production for publications.

The new publishing system has helped to enlarge publishing in Swedish and English as well.

## 3. Summary

Use of a structured XML document in statistical releases enables flexible and effective use of different distribution channels. The same XML document can be utilised for both the conventional publishing mode and for the adoption of new distribution channels.

The changeover in the publishing system has supported Statistics Finland's strategic objective to widen the free data supply on the web, when all previously chargeable printed and pdf-format basic statistical publications are now available free of change on the web. It is possible to make any final estimates of the intensification of processes only when the new publishing system is fully implemented and its use is well-established. The development of production and metadata systems also makes it possible to use the publishing system more effectively. The estimation finally hinges on clear intensification of processes.

In connection with the renewal the quality of publications and releases has been assessed systematically, and technical and training and instruction-related development needs have been recognised. Although this was a change of the publishing system, it has an effect on all publishing modes. For data users a closer connection than before between releases and database tables aids in finding information.

## 4. Sources

Some papers about the implementatation of the XML-based publication process based on CoSSI-model.

http://www.stat.fi/org/tut/dthemes/drafts/impl_en.html