Dynamically Updated Publications

By Jacob Chr. Solander, Statistics Denmark

Introduction

The strategy of Statistics Denmark clearly states that the Internet is the most important media for dissemination of statistics. The present situation is that traditional paper publications are typically available on the Internet as PDF documents, and more detailed data are available in the on-line databank called "*StatBank Denmark*" (www.statbank.dk). Furthermore, we have dynamically updated Internet publications, taking all their information from different sources (Back-office systems), using XML.

But what is meant by the term "Internet publication"? This issue has been discussed within Statistics Denmark the past year. Some will use this term about PDF documents, while others think of dynamically updated web applications, with a high level of interaction.

When talking about using XML in the publication process, it becomes important to distinguish between different concepts of publications. Not only the difference between paper and electronic publications, but also the level of interaction and dynamics within electronic publications.

This paper will focus on the experiences we have with publications that are fully or semi-automatically updated, and the use of XML in these processes. Due to the reasoning above, we shall distinguish between dynamic Internet publications and automated production of paper and PDF documents.

Internet publications

As mentioned above, a discussion is going on in Statistics Denmark about the form and shape of future Internet publications. However, we have had completely dynamic Internet publication for several years now. The best example is our "Key Indicators". Below you can see one of the indicators (in Danish). The interesting thing about these indicators is that all pages are 100% dynamically updated from other sources of data. Thus, there is no editor doing any work on these pages, when new data are released.

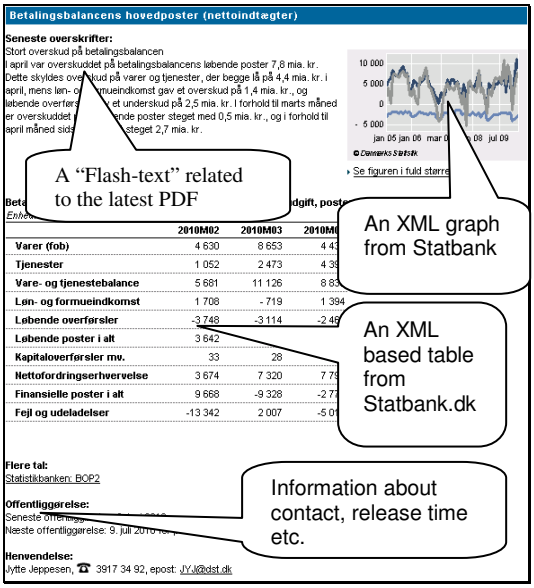


Fig. 1: A completely dynamic Internet publication taking all its information from other sources.

The page can be split into four elements, which are described below. When a user enters a key indicator page, all data are loaded into the four elements dynamically, by retrieving the latest data from the respective sources. Thus, none of the elements on the page are written specifically for the internet page. So all the information is "reused" from other information sources.

1 - The text element
Data for this element is taken from the latest PDF “News release”. A short abstract for each news release is stored in the publication database. We refer to this as the “Flash text”. The same text is used on the front page of www.dst.dk, when the news release is made public.

2 - The graph element
The graph is created dynamically from a saved query in StatBank Denmark, which defines the general contents to be presented as a graph. When calling for a dynamic graph, an image file is created on the fly (if it is not already there) and the result returned is an XML file which specifies the temporary path for the graph image, as shown in fig. 2.

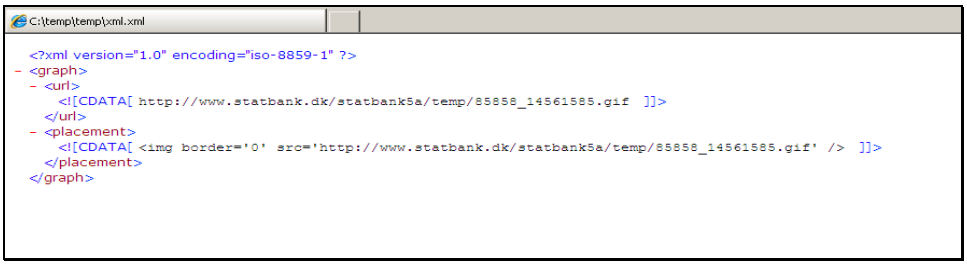


Fig. 2: XML wrapping the URL of the temporary graph image file.

3 - The table element
The HTML table presented as the third part of the page is generated using the XML feed / Web Service. It is possible to request the *XML feed* directly:

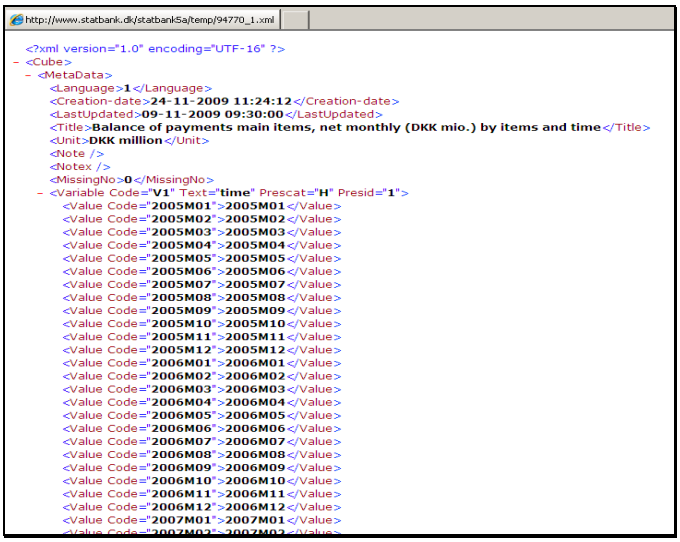
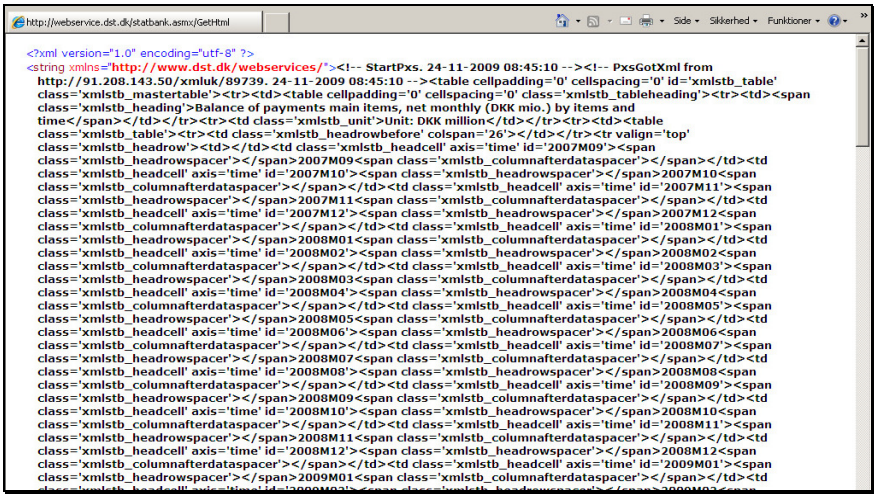


Fig. 3: XML feed for a saved query.

Or by using the Web Service, which returns HTML ready for presentation, wrapped in XML:



<?xml version="1.0" encoding="utf-8" ?>
<string xmlns="http://www.dst.dk/webservices/"><!-- StartPxs. 24-11-2009 08:45:10 --><!-- PxsGotXml from
http://91.208.143.50/xmluk/89739. 24-11-2009 08:45:10 --><table cellpadding='0' cellspacing='0' id='xmlstb_table'
class='xmlstb_mastertable'><tr><td><table cellpadding='0' cellspacing='0' class='xmlstb_tableheading'><tr><td><span
class='xmlstb_heading'>Balance of payments main items, net monthly (DKK mio.) by items and
time</td></tr><tr><td class='xmlstb_unit'>Unit: DKK million</td></tr><tr><td><table
class='xmlstb_table'><tr><td class='xmlstb_headrowbefore' colspan='26'></td></tr><tr valign='top'
class='xmlstb_headrow'><td></td><td class='xmlstb_headcell' axis='time' id='2007M09'><span
class='xmlstb_headrowspacer'>2007M09</td><td
class='xmlstb_headcell' axis='time' id='2007M10'>2007M10<span
class='xmlstb_columnafterdataspacer'></td><td class='xmlstb_headcell' axis='time' id='2007M11'><span
class='xmlstb_headrowspacer'>2007M11</td><td
class='xmlstb_headcell' axis='time' id='2007M12'>2007M12<span
class='xmlstb_columnafterdataspacer'></td><td class='xmlstb_headcell' axis='time' id='2008M01'><span
class='xmlstb_headrowspacer'>2008M01</td><td
class='xmlstb_headcell' axis='time' id='2008M02'>2008M02<span
class='xmlstb_columnafterdataspacer'></td><td class='xmlstb_headcell' axis='time' id='2008M03'><span
class='xmlstb_headrowspacer'>2008M03</td><td
class='xmlstb_headcell' axis='time' id='2008M04'>2008M04<span
class='xmlstb_columnafterdataspacer'></td><td class='xmlstb_headcell' axis='time' id='2008M05'><span
class='xmlstb_headrowspacer'>2008M05</td><td
class='xmlstb_headcell' axis='time' id='2008M06'>2008M06<span
class='xmlstb_columnafterdataspacer'></td><td class='xmlstb_headcell' axis='time' id='2008M07'><span
class='xmlstb_headrowspacer'>2008M07</td><td
class='xmlstb_headcell' axis='time' id='2008M08'>2008M08<span
class='xmlstb_columnafterdataspacer'></td><td class='xmlstb_headcell' axis='time' id='2008M09'><span
class='xmlstb_headrowspacer'>2008M09</td><td
class='xmlstb_headcell' axis='time' id='2008M10'>2008M10<span
class='xmlstb_columnafterdataspacer'></td><td class='xmlstb_headcell' axis='time' id='2008M11'><span
class='xmlstb_headrowspacer'>2008M11</td><td
class='xmlstb_headcell' axis='time' id='2008M12'>2008M12<span
class='xmlstb_columnafterdataspacer'></td><td class='xmlstb_headcell' axis='time' id='2009M01'><span
class='xmlstb_headrowspacer'>2009M01</td><td

Fig. 4: Result returned from a Web Service, using a saved query id.

4 - The information element

The remaining metadata about contact person, date of publication and link to the source data in StatBank Denmark, are retrieved from the publication database.
The Danish version of the "Key Indicators" is used in this example, because it has more elements than the English version. All key indicators are found on:
http://www.dst.dk/focuson (English version)
http://www.dst.dk/keyindicators (English version)
http://www.dst.dk/hovedtal (Danish version)

Caching XML files

Both XML files for tables and graphs are cached, to reduce the work load on the servers. When an XML file is requested, the name will determine if the file is the correct version. The first user who enters a page, after 09:30 in the morning, will cause a fresh file to be created. After that, a cached version is used. If data are revised or an error should occur, all related temporary (cached) files are deleted. From an implementation point creating proper caching procedures is a significant challenge which requires great care.

Services for dynamic updating

In this section some of the services offered by Statistics Denmark on the web are presented. We shall focus on the services related to the dynamic updating of tables.

Saved queries

When a subset of a cube in StatBank is selected, it is possible to save this table request for later use. Each saved query gets a unique identifier (e.g. *pxs-id:* 136631) that is used when accessing data afterwards. The query can be used in a number of different ways. Each query has a number of attributes, such as the automated time period selection for future updated tables. One can choose between "rolling time periods (the same number of periods each time)", "fixed starting period (all available time periods since the first selected)" or "Last updated (the time periods updated at the last data release)".
Example: http://www.statbank.dk/pxs/136631

Standardised deep links

The example above gives you a saved query in the same way it would be presented to you if you had just made the selection yourself in StatBank Denmark, but always with the latest updated time periods. However, the system offers several other ways of presenting the same query. If you need the result as an Excel Web Query, you access the query as displayed in the example below. This will give you just the data table itself, without thousands delimiter and comma presentation according to the language of the table. Thus, this table is appropriate for loading into Excel as a normal Excel table. Accessing the table as demonstrated in the example concerning "XML feed through URL" is a third way of deep linking to the same saved query.

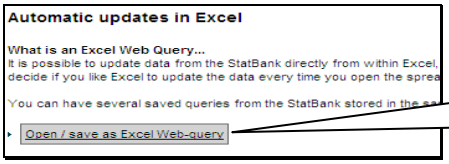
Example: http://www.statbank.dk/xls/136631

Maps and diagrams – XML wrapped
If the saved query can be presented on a map or as a diagram, this is done in a similar way. Several standard chart presentations are prepared with different sizes and types of diagrams. Try one of the examples below. The diagrams can also be accessed wrapped in XML, for dynamic use on other web sites. This technique is used for diagrams presented on the "Key Indicator" pages.

Examples:
http://www.statbank.dk/map/83054 (web page with map, legend, menus etc.)
http://www.statbank.dk/gif/136631 (A gif with a graph)
http://www.statbank.dk/fs_graf_uk0/85858

XML feed through URL
Any saved query can be returned as XML. We refer to this method as *XML feed*. This technique is used for dynamic table presentations of the "Key Indicators" and is the input used in our StatBank Web Service.
Example: http://www.statbank.dk/xml/136631

Excel Web Queries
Knowing the URL where data can be found is really all that is needed for a user to create an Excel Web Query. However, it is possible for the provider to take this a step further. Thus, from any saved query in StatBank Denmark, you can download an IQY-file, containing the necessary information for Excel to find table data over the Internet. Opening this file on a standard computer with Excel installed results in a standard Excel sheet with updated data from the saved query in StatBank Denmark. In the Excel sheet the user can specify whether data should be 1) updated every time the spreadsheet is opened; 2) with a specific time interval or 3) manually via "left click" on the table (which is the default setting).



Example:
http://www.statbank.dk/136631.iqy
(This link is generated when you click on the link "Open/save as Excel-Web-query")

Web Services from StatBank Denmark
Finally saved tables can be accessed as a Web Service. There are unlimited ways of returning data when using Web Services. The example below returns the result as a *variable* of the type *string*. The information in the string variable is a finalised HTML table that can be dynamically inserted into a receiver's web site.
Example: http://webservice.dst.dk/statbank.asmx?op=GetHtml (test with *pxsid = 85858* and *lang = 1*). Thus, the basis for all the different services is the saved queries, which can return data in many different forms and shapes, by changing just one parameter in the URL that activates the query.

Conclusions
The Internet is considered the most important media for dissemination of statistics, and XML is a natural part of the production process. The format in itself is not crucial as long as the process is efficient. This is achieved through a high level of automation and standardization. It is important to use formats that are well structured and adaptable. To ensure coherence between statistics that are disseminated through different channels, using a single-source-strategy is essential. However, a high level of flexibility and efficiency of the output database that delivers data is required. Before deciding on the tools and formats to be used in the future Internet publications, it is necessary to determine the level of ambition for these publications. It is also necessary to address the problem of archiving. This issue becomes highly relevant if publications are dynamic by nature, and gather their information from different sources, using XML.