

Influence of Internet on Dissemination of Official Statistics

Tomaz Smrekar¹, Statistical Office of the Republic of Slovenia

Summary

In the pre-internet period official statistics was first available in paper publications, computer printouts and later on magnetic tapes and CD-ROMs. In time statistical tables became more detailed and gradually more written material such as statistical analysis, explanation notes on concepts, sources and methods, and charts were added to facilitate user interpretation and maximisation of the use of statistics. Due to electronic data processing time series became available. The main users in the pre-internet period - researchers and government officials - were statistically literate. Less statistically literate users such as journalists (who are among the most important disseminators of official statistics) and the general public were not major user groups at that time.

The Internet has introduced a revolution in dissemination, communication and accessibility of official statistics in the second half of the 1990s. Web 1.0 is mostly about publishing static HTML pages on a server and it upgraded dissemination, communication and accessibility of official statistics substantially with the creation of National Statistical Institutes (NSI) websites. Official statistics became more visible, accessible and used more than ever by different user groups (media, general public, researchers, students, policy makers, businesses, non-governmental institutions, etc.). Involvement of user groups became more influential than ever before. It resulted in developments of new products and services (extraction of data from electronic databases, storytelling and writing for web, better metadata, reports aiming to build trust in the system of official statistics).

Web 2.0 (from 2000 on) is more dynamic and interactive. It has enabled NSIs to make decisive steps in the so-called knowledge continuum (data, information, knowledge, wisdom) mainly through self-help tools for data visualisation and animation and upgrading statistical literacy. Notification e-mails, RSS and bookmark and share tools have become frequently used by NSIs. In some countries with a large number of institutions cooperating in the production of official statistics portals for official statistics have been introduced. In some NSIs wiki technology to produce Wikipedia-style portals for presentation of statistics has been introduced. A decisive step in communication among official statisticians was achieved by sharing visualisation tools.

Emerging Web 3.0 (semantic web, data web) is all about improvements in technology. We have to remember that technological improvements have always resulted in major improvements in official statistics (computers, Web 1.0). Web 3.0 technology will enable remix, reuse and repurpose of official statistics on the web in quantity and ways never seen before. Different data will compete in the virtual world of the internet and official statistics has a lot of advantages in "competing" with other data: attached metadata, quality reports, long time series, regional breakdowns, comparison in international context.

1. Official statistics

Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information².

¹ Tomaz Smrekar is head of dissemination at the Statistical Office of the Republic of Slovenia, Tomaz.Smrekar@gov.si

² Fundamental Principles of Official Statistics

Statistics make it easier for people, politicians and the business community to stay informed and to make informed decisions. Official statistics must be presented in such a way that the main results can be understood without expert knowledge of statistics, coherent and easily accessible.

2. Users and dissemination of official statistics in the pre-internet period

In the pre-internet period the main users of official statistics were experts who were statistically literate (high degree of interest and understanding of statistics). They were mediators of statistics and analyses to governments and politicians.

In the pre-internet period the main dissemination channel for statistics to businesses and the general public was the media.

Official statistics was first available in paper publications, computer printouts and later on magnetic tapes and CD-ROMs. In time statistical tables became more detailed and more written material such as statistical analysis, explanation notes on concepts, sources and methods and charts were added to enable user interpretation of statistics and maximisation of the use of statistics.

3. Influence of internet on dissemination of official statistics

The web³ has introduced a revolution in dissemination, communication and accessibility of official statistics in the second half of the 1990s.

Typical products on the web are electronic versions of paper documents, electronic-only documents, tables, databases of aggregated data and micro data⁴, metadata and knowledge databases, spreadsheets, static and animated graphs and maps and podcasts⁵.

NSIs present data and metadata, data release dates, business information for reporting units and users, school curriculum materials for teachers and students, research and information papers and podcasts promoting statistical literacy and access to micro data for researchers through the web.

Statistics is visible and free of charge

Being on the web official statistics has become more visible, accessible and more used than ever. The cost effectiveness has improved by maximising the opportunities offered by the web and enabling more and more data to become free of charge for the end users.

It should be easy to get statistics from the web

A good website of official statistics can be defined according to whether or not the users get what they are looking for quickly and easily and whether or not statistics are suitable for further manipulation of data. The key elements are meaningful and up-to-date content that is prepared for web use,

³ Web 1.0 is mostly about publishing static HTML pages on a server and it upgraded dissemination, communication and accessibility of official statistics substantially with the creation of NSI websites. Web 2.0 (from 2000 on) is more dynamic and interactive than Web 1.0. It has enabled NSIs to make decisive steps in the so-called knowledge continuum (data, information, knowledge, wisdom) mainly through self-help tools for data visualisation and animation and upgrading statistical literacy. Notification e-mails, RSS and bookmark and share tools have become frequently used by NSIs. In some countries with a large number of institutions cooperating in the production of official statistics portals for official statistics have been introduced. In some NSIs wiki technology to produce Wikipedia-style portals for presentation of statistics has been introduced. A decisive step in communication among official statisticians was achieved by sharing visualisation tools.

⁴ For researchers under special conditions only.

⁵ Podcasting is an audio or video visualisation tool.

consistent navigation, consistent “look and feel” and efficient search option. It has to be easily accessed by people with disabilities and with older web browsers, too.

When web development in an NSI is distributed across the organization it is a real challenge to prevent inconsistent presentation of the content and unconnected navigation on the site. Many NSIs use content management systems with fixed templates for efficient work, assurance of implementation of institutions web style standards and as a result consistent “touch and feel” through the site.

It is important to know and understand the users of statistics

NSIs are interested in information who is visiting their websites, what visitors are looking for, what problems they have with obtaining information and what suggestions they have for improving the websites. Most NSIs use feedback to determine new statistics, type of products and the mode of dissemination.

NSIs communicate with different types of users. Users differ in their attitudes (statistical literacy), efforts, time and money to give away to obtain statistics. The general public appeared as an important user group that influenced a range of new products and services (storytelling, writing for web, metadata databases, and data visualisation tools).

Enmon (1997) classified users into three groups: “tourists”, “harvesters” and “miners”. “Tourists” have limited statistical literacy; on the other hand, experienced users, i.e. “harvesters”/“miners”, are able to exploit statistics in different forms. NSIs should provide suitable products for all three user groups.

Understanding statistics-statistical literacy⁶ enables maximization of the use of statistics

Research shows that about 80% of the public and 90% of media personnel are number blind. This means that NSIs have not an easy job to perform in communicating statistics: training statisticians to try to be as precise as possible when on the other hand the majority of users ask for messages. To enable users to maximize their use of statistics, products such as static and animated graphs and maps, databases of metadata and podcasting have been introduced in official statistics.

Knowledge continuum

Knowledge continuum is a way of understanding how people move through a range of experiences at varying depths of cognition. Data are the basis for information, information leads to knowledge and with deep understanding comes wisdom⁷. For real benefits of the society and individuals, official statistics goes beyond numbers. Without context a number provides little useful meaning but by putting numbers in a context NSI publishes data. With more context and meaning added to data it produces information and enables users to develop knowledge and even wisdom.

Electronic databases on the internet add value for reusing of statistics

An important development in electronic dissemination has been the introduction of electronic output databases. By storing all published data in a database, it is possible to ensure that experienced/demanding users search and download exactly and only the data they need and reuse them according to their needs.

Metadata matter more than ever

The data about the data (metadata) are particularly important on the web and consist of discovery metadata (indexes, site maps and lists of terms (A to Z, glossaries)), which help users find what they are looking for, and interpretative metadata (links to descriptive information, footnotes, and metadata repositories as companions to online databases), which help users to understand statistics.

⁶ Katherine Wellman defines statistical literacy as the ability to understand and critically evaluate statistical results that permeate our daily lives - coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions.

⁷ Pollock J. T. Semantic Web for Dummies, Wiley Publishing 2009, Indianapolis, Indiana, p. 118

Revisions should be clearly indicated

The ease and speed with which the content of websites can be revised has led to an increase in the number of revisions. To properly manage the revisions, many NSIs have established a transparent way of data revision called data revision policies, which are published on the web to inform users. Revised data are clearly indicated and the date of the revision and the nature and effects of the revision are added.

NSI are (pro)active in disseminating statistics

The web has enabled NSIs to change from observing the use of statistics to becoming proactive disseminators by pushing statistics through Real Simple Syndication (RSS) and e-mail notification.

Self help tools enable NSIs to better manage financial assets and users to be less dependent on NSIs

Users became less dependent on NSIs in searching for statistics using self help tools to generate their own statistical outputs. Self help tools are mostly used in extraction of data and metadata, release calendars, ordering of statistics releases, updates of self created tables, using visualization tools (graphs, maps) and downloading answers from knowledge databases ("frequently asked questions").

Writing for the web in journalistic style

NSIs are obliged to make statistics they disseminate useful for the users. We know that stories get people interested in statistics and help them to understand what the statistics mean. With a special way of writing for the web, NSIs try to maximise the use and minimise the misuse of statistics.

People who are busy and usually in a hurry rarely read web pages word by word. Instead, they scan the page, picking out individual words and sentences (79% of users always scan web pages and only 16% read them word by word).

On the web the focus has shifted from plain numbers to text explanations of numbers. This is the shift from data to information. Writing specifically for the web is becoming increasingly important in communicating with users.

NSIs adapted to cognitive psychology research results of user behaviour on the web by writing short, scannable and to the point to minimise user's cognitive load.

Clear and simple everyday language is used as much as possible in short sentences, short paragraphs, with one main idea per paragraph and subheadings to guide the reader's attention. Contents are cross linked on the web and linked to metadata.

Main conclusions are at the top of the story, followed by details arranged in decreasing order of importance. Headings break the text into manageable sections, simple tables support the text messages and graphs show trends or relationships.

Data visualisation tools make statistics more interesting and easier to understand

If a picture is worth a thousand words, an animated and interactive picture could be worth even more. Visual animation touches our lives daily by television weather forecasts, computer games and transport navigation. It has enormous potential for presentation of statistics, too.

Professor Hans Roseling of the popular website Gapminder⁸ said: "Our approach was simple; let's apply the animation concepts from computer games on statistics! Our interactive moving graphics have been surprisingly well received."

Interactive maps⁹ and population pyramids¹⁰ are typical visualization tools NSIs provide for their users.

Visualisation and social software go hand in hand

A combination of visualisation and social tools is a step forward and a typical Web 2.0 product. For example, Swivel¹¹ is a forum to publish data and also to visualize data. Everybody can give his information and opinion about his favourite topic or as Swivel says "Swivel is a place where curious people explore all kinds of data".

Many Eyes¹² is a service offered by IBM that combines information visualization with social software, enabling collaborative visualization by groups of users. The service presents a set of interactive visualizations that provide insight into a variety of topics. In addition, visitors are able to upload new data sets and create their own visualizations.

More efficient access to micro data for researchers

The web has enabled researchers to analyze statistical micro data literally from their office/home. They can access the database relevant for the specific project and process data "on-line".

Physically the data are stored at NSI servers but they are visible for the researcher on the screen and the aggregated results from the processing of the data are sent by e-mail to the user after statistical confidentiality checks.

Usually researchers can choose from software like STATA, SPSS, SAS, GAUSS, Super Cross and MS Office. The advantages of using remote access or remote execution of programmed software are convenient for a researcher because there is no commuting due to on-line access for data, they are able to work in a familiar software environment, there are no investments in software or hardware for the user, the system is safe for the NSI and for the researcher.

4. Future of official statistics on the web

How to find the way in the mass of data?

Larger and larger volumes of information will become available through the web and it will be more and more difficult for users to find their way through this mass of data and to define fitness for the purpose of different data for their particular needs.

Emerging Web 3.0 (semantic web, data web) is all about improvements in technology. We have to remember that technological improvements have always resulted in major improvements in official statistics (computers, Web 1.0).

Web 3.0 technology will enable remix, reuse and repurpose of official statistics on the web in quantity and ways never seen before. Different data will compete in the virtual world of the web and official statistics has a lot of advantages in "competing" with other data: attached metadata, quality reports,

⁸ <http://www.gapminder.org>

⁹ <http://surs.monolit.si>

¹⁰ <http://www.stat.si/eng/Piramida.asp>

¹¹ <http://www.swivel.com>

¹² <http://manyeyes.alphaworks.ibm.com/manyeyes>

long time series, regional breakdowns, comparison in international context will enable users to find official statistics more useful than other data.

Users do not care who produced official statistics

It is likely that due to user friendliness NSIs will provide portals¹³ of all national official statistics. Obtaining data from multiple data sources is too burdensome because decision makers have limited time available and even if they are statistically literate and have appropriate technical skills they find it too time consuming to identify and organize the appropriate data on a particular issue. New software will allow topically oriented rather than dataset oriented presentation of statistics in text, statistical tables, graphs and maps.

International data will gain importance

International data sets will grow in importance due to globalization. Providing statistics to international agencies is burdensome for NSIs. To provide more efficient world wide data exchange, an international standard protocol for exchanging data and converting data to multiple output formats called Standard Data and Metadata Exchange (SDMX)¹⁴ for time series data is being developed. The BIS, ECB, EUROSTAT, IMF, OECD, UN, and the WB have joined together to focus on business practices in the field of statistical information that would allow more efficient processes for exchange and sharing of data and metadata within the current scope of their collective activities. The goal is to explore common e-standards and ongoing standardization activities that could allow them to improve efficiency and avoid duplication of work.

Artificial creation of new knowledge

Wolfram Alpha¹⁵ is called a new “search engine” that can have influence on dissemination of official statistics. It is a search engine only by “touch and feel”, but in reality it doesn’t search but it computes the answers and actually produces new knowledge. The answer which appears as a page of text and graphics is not the answer to the question. Wolfram Alpha figures out what the question means, looks up the necessary data to answer the question, computes an answer and designs a page to present the answer. It is about how we might build human knowledge from computational rules.

5. Conclusions

The web has introduced a revolution in dissemination, communication and accessibility of official statistics. Official statistics has become more visible, accessible and used than ever in history. Today the majority of official statistics is free of charge for the end users.

NSIs try hard to disseminate meaningful and up-to-date statistics prepared for web use. Users strongly influence new statistics, type of products and the mode of dissemination of statistics. NSIs provide suitable products for different user groups that differ a lot.

Understanding of statistics enables maximization of the use of statistics. Due to low levels of statistically literate people, NSIs have not an easy job to perform in communicating statistics. To enable users to maximize their use of statistics, products such as writing for the web in journalistic style, static and animated graphs and maps, databases of metadata and podcasting have been introduced in communication of official statistics.

SIs are proactive in dissemination of statistics by RSS and e-mail notification. On the other hand, self help tools enable NSIs to better manage financial assets and users to be less dependent on NSIs

¹³ <http://www.fedstats.gov>

¹⁴ SDMX is an initiative to foster standards for the exchange of statistical information, <http://sdmx.org/>

¹⁵ www.wolframscience.com

Access to micro data for researchers has become easier and safer through the web for researchers and NSIs.

In the future larger and larger quantities of statistics will become available through the web and it will be more and more difficult for users to find their way through this mass of data and to define fitness for the purpose of different data for their particular needs. Official statistics with attached metadata, quality reports, long time series, regional breakdowns, comparison in international context will enable users to find official statistics more useful than other data.