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

UNECE Work Session on the Communication of Statistics
(29 June – 1 July 2011, Geneva, Switzerland)

(i) Gathering user feedback from web sites

Statistics Explained user statistics via Piwik: who is looking?

Submitted by Eurostat, EU[1]

I. INTRODUCTION

1. On 28 September 2009 Eurostat, the statistical office of the European Union, launched **Statistics Explained** (http://epp.eurostat.ec.europa.eu/statistics_explained), a new and fully electronic way of publishing statistics. Statistics Explained targets the general public, offering them statistical and background articles, complemented by a glossary, using a Wikipedia-like structure and the same open-source software, Mediawiki. One of its major objectives, reaching a previously unserved audience looking for statistical information, seems to be successfully met: visitor figures are high, Statistics Explained web pages come up prominently in search engine result pages and they are picked up frequently and broadcast further by blogs, press articles, columns, discussion forums, as well as tweets and Facebook pages and Wikipedia articles.

2. In November 2009 an open-source Mediawiki plugin for collecting user statistics called **Piwik** was installed. Piwik is an intuitive and user-friendly tool generating customisable reports with very detailed and real-time data on visitors and usage. This information is valuable as feedback and input for further improvement, in line with users' behaviour and needs. The most relevant features in this context are search key words used, referring external sites, page rankings, bounce rates and actions.

3. Apart from the general site-oriented feedback, Piwik metrics are sufficiently detailed and recent to give an idea about the latest trends in the internet and how they affect people searching for statistical information – or not. They can answer questions on:

- The relative importance of alternative access modes: search, direct entry (bookmarks) or referring sites; and, at a more detailed level, on search terms used most frequently and on web sites sending through significant numbers of visitors.
- The demand for statistical themes and topics, and how these correlate with events in the wider world (e.g. banking crisis, greenhouse gas emissions, Euro area government finances, sustainable development, information society).
- The importance, or possibly the lack of importance for statistical dissemination via social media sites (Facebook, Linkedin, Twitter …) and how this develops over time.

[1] Prepared by Marc Debusschere, marc.debusschere@ec.europa.eu.

- The success as a reference and trustworthy source for Wikipedia, academic and educational sites, blogs, media, and the many heated discussions raging on discussion forums.
- The positive or negative comments, of a more qualitative nature, on the relevance of the statistical information offered and on the way it is presented.

4. The analysis below is based upon Piwik user statistics for May 2011, compared if relevant with May 2010 totals and with evolutions since the start of recording in November 2009.

II. VISITS AND ACTIONS

5. Since it opened to the public in September 2009, Statistics Explained received **over 8.3 million page views**. In May 2011 Statistics Explained registered **104 114 visitor sessions ('visits')**, 71 319 of which by unique visitors. An estimated 1% of visits are from Eurostat staff updating. Visitors performed 439 621 'actions': page views, outlinks (links in page clicked) or downloads. Compared to the May 2010 figure of 53 454 visits, the number has almost doubled (+ 95%), to somewhat over 4000 visits per working day. Compared to the figure of the first recording month (November 2009), 14 729, visits have multiplied by seven in somewhat over 2,5 years.
Exactly 50% of all visits 'bounced' (left the site after visiting only one page), which means that the other half performed on average 7.4 actions. This is an indication that visitors to a Statistics Explained article quite frequently click on the links to glossary pages, images, further information, similar pages and relevant external sites. The average visits lasts 3.5 minutes, but almost 60% of visits take 30 seconds or less.

III. HOW IS STATISTICAL INFORMATION FOUND?

6. Search engines provide the most important entry point to Statistics Explained: 50.3% of visitors enter via a search, but adjusting this for additional traffic from search engine sites (e.g. Google, Google translate, Yandex, Yahoo, Ask) an estimated **55% access via keyword search**. The next most important channel is **direct entry, somewhat over 40%** (typing or copying in URL in browser address bar or, more likely, using bookmark). Finally, some **5% enter via clicking on a link** on a non-search site. These proportions do not change a great deal over time; in spite of the doubling of visits, May 2010 showed a largely similar distribution (estimated at 56, 38 and 6%, respectively); and at the modest beginning in November 2009, with 14 729 visits a month, proportions were approximately 58, 40 and 2%.

A. Search engines

7. The May 2011 top 10 of keywords to access Statistics Explained:

	Search keyword	Visits
1	air pollution statistics	484
2	pollution statistics	196
3	trade flows eu database	118
4	eurostat unemployment	97
5	statistics explained	86
6	eu 15	71
7	construction cost index	69
8	climate change statistics	61
9	full time equivalent	57
10	npish	54

People looking for particular information via keyword search, for example via Google, will only access your website if it ends up high enough, preferably in the top-10 on the first page of a search result. On the whole, Statistics Explained's rankings are high and still going up, even for non-Europe-specific search terms like '*air pollution statistics*' (2nd out of 11.2 million results), '*ISCED levels*' (6th of 75 000) or '*full time equivalent*' (6th of 27 million). As a result, the Piwik ranking of keywords does not only reflect the frequency with which certain data are searched for, but also the score in a particular search, compared to competing sites. This is clear from a comparison with the top 10 of articles, both in total views since the public opening at the end of September 2009 and May 2011 views (between brackets is the May 2011 ranking, very similar to the overall one):

	Page	Total	May 2011
1	Unemployment statistics	117 522	8198 (1)
2	Minimum wage statistics	60 961	2455 (3)
3	GDP per capita, consumption per capita and comparative price levels	56 533	2962 (2)
4	National accounts – GDP	48 620	1869 (5)
5	Migration and migrant population statistics	45 727	1251 (9)
6	Wages and labour costs	40 192	1685 (6)
7	Air pollution statistics	38 492	2018 (4)
8	Tourism trends	35 255	1422 (7)
9	Waste statistics	34 008	863 (19)
10	Energy production and imports	29 940	1274 (8)

It is striking, when comparing the two lists, that a significant number of visitors enter via a glossary page (e.g. EU-15, NPISH, full-time equivalent), and that the two lists don't match perfectly. The most-viewed pages also contain themes like migration, tourism, wages, GDP. Both lists, however, correlate quite well with topics that have been in the news: global warming and Copenhagen conference, euro crisis (GDP, but also minimum wages), migration and asylum policies; others, however, seem to be 'evergreens': unemployment, tourism, waste.

8. Finally, searches are not limited to English-language page searches. Firstly, **Google Images** also comes up surprisingly often, accounting for almost 10% of searches, making the Statistics Explained maps, graphs and tables an important additional entry point. Secondly, **Google translate** in its many language versions is used very frequently to display Statistics Explained pages in other languages than English; quality is acceptable and gives a fairly accurate idea of the original content. Statistics Explained, although as yet only available in English, is technically fully equipped to handle multiple languages and recently the decision has been taken to publish 80 statistical articles also in German and French (compared to the 353 available in English) as well as some 20 high-interest ones in all 23 EU languages.

B. Direct entry

9. Without more detailed information, it is hard to find out who is behind the 40% direct entries. A significant part is undoubtedly made up of Eurostat and external contributors accessing Statistics Explained on a daily or at least regular basis. But the majority are probably repeat users who want to check at regular intervals what is new; especially in the case of pages used to release new data this makes perfect sense (e.g. the most-viewed page, Unemployment statistics, always displays the latest monthly figures). It is also expected that some viewers access a specific page to use it in a portal-like fashion, for the 'Further Eurostat information', with deep links to specific publications, main tables, database entry points, metadata, legal texts etc.

C. Referring sites

10. High search rankings make a website better known, resulting in more external sites linking to it. This in turn improves search rankings. So a high number of referring sites is an important asset for attracting traffic, both directly and via the impact on search rankings. Statistics Explained has a large and growing number of sites linking to it: in May 2011 654 different sites sent through 9467 visitors. Referring sites are discussed in more detail below.

11. It is important to realise that most referring links have an **expiry date**; this is especially true for quickly-outdated entries in discussion forums, blogs, press articles, social media pages and tweets, but less so for well-visited portal, home and special-interest pages. The first type of link, however, can also be responsible for huge visitor **peaks** (some of whom will create a referring link in their turn, or maybe become repeat visitors). To date, Statistics Explained has known two noteworthy peaks:

- In May 2010 Beppe Grillo, an Italian journalist, added in his widely read blog a link to 'Information society statistics at regional level', resulting in some 2000 direct and several hundreds of indirect additional visits;
- In February 2011 the Spanish newspaper El Pais linked to the newly-updated minimum wages article, resulting in some 6500 extra visits in 3 days (with 8049, 6374 and 4597 visits, respectively, against a 'normal' level of 4000-4200 per working day); by the 4th day, however, the visitor number (4028) had already returned to base level. Activity was even more concentrated than these day figures indicate, shooting up around noon on 14 February from a normal 282 at 11-12 o'clock to 1583 the next hour and 1367 from 1 to 2 o'clock, causing some access problems.

D. Some recommendations on maximising traffic

- The basis for increasing usage is 'findability' by Google and other search engines: a high ranking, preferably among the first 10 results, on page 1. Some elements are known to contribute to this: using key terms in URL and level-one header (page title), 'connectedness' (quality links out and especially in), ratings by viewers, …
- Being on page 1, however, does not guarantee click-through: viewers must be able to quickly judge the page's relevance for their information search. Consequently, displaying the most important information on top of the article is crucial; it is also partly displayed on the search result page and provides some help to decide whether or not to access the page makes sense.
- Internet surfers are notoriously impatient, so the most relevant and concrete information must be on top, rather than lengthy philosophical reflections. A web-adapted introduction allows viewers to decide in 5 seconds whether they have found what they were looking for.
- Be aware that visitor peaks may occur when a referring link is displayed at high-profile sites; if significant numbers on the BBC website should click through, the site is quite likely to crash.

IV. HOW IS STATISTICAL INFORMATION USED?

12. Piwik allows visiting the places on the Internet where the statistical information from Statistics Explained is linked to, analysed and discussed. This provides vivid and extremely useful quantitative and qualitative feedback. Although Statistics Explained is English-language only, for the time being, it is striking that it is accessed from a great many different countries and all continents, and translated in numerous other languages, as demonstrated by Google translate results. Apart from the social media, discussed in more detail below, two main categories of uses can be distinguished: information sites and comment and discussion sites.

A. Information sites: Wikipedia, academia and government

13. The highest-scoring non-search referring site still is Wikipedia, with some 600 visits in May 2011 (0,6% of the total and 6.3% of referring sites). Figures peaked in March-June 2010 (see table below), although visitor numbers are falling even in absolute terms.

Month	Visits from Wikipedia	Total visits	%
Nov 2009	74	14729	0.5%
Dec 2009	7	20819	0.0%
Jan 2010	62	26717	0.2%
Feb 2010	66	31049	0.2%
Mar 2010	994	40316	2.5%
Apr 2010	**1295**	40763	**3.2%**
May 2010	1282	52454	2.4%
Jun 2010	1012	41272	2.5%
Jul 2010	681	38847	1.8%
Aug 2010	679	41960	1.6%
Sep 2010	875	54688	1.6%
Oct 2010	936	68794	1.4%
Nov 2010	941	90434	1.0%
Dec 2010	738	74941	1.0%
Jan 2011	814	84548	1.0%
Feb 2011	582	103441	0.6%
Mar 2011	593	101990	0.6%
Apr 2011	529	92910	0.6%
May 2011	590	104114	0.6%

The reason is probably that pages linking through are mainly in the domains of national accounts and public finance, and this topic is somewhat less in the picture than it was during the 2010 'Greek crisis'. Statistics Explained is presently linked to from the English (some 50 pages and redirects), French (3), German (1), Czech (1), Estonian (1), Hungarian (1) and Vietnamese (1) language versions.

14. The websites of the European national statistical institutes constitute a second very important category of referring sites, collectively providing some 750 visits (0.7% of the total, 8% of referring sites). The position of the link is very important, as is shown by a home page link on the Romanian NSI website, alone accounting for 472 of those 750 visits.

15. The remainder of information sites consists of a large number of educational, library and documentation centres, either academic or business-oriented, ministries and administrations at European, national and regional levels, each sending a small but steady number of visitors.

B. **Comment and discussion: press sites, blogs, discussion forums**

16. The list of referring sites contains many examples of Statistics Explained material being cited and linked to in order to make a point in a press article or blog, or in a contribution to a discussion forum. Depending on the visibility of this link and the importance of the site this may result in additional visits ranging from a limited amount to a huge number indeed. The highest to date are the 6500 visitors resulting from the article on the site of El Pais cited above.

17. Finally, references to Statistics Explained crop up in the most unexpected environments, from a Russian or Canadian gaming site over a German 'Bodybuilding Szene' to 'Cycling Dutch Style'.

V. HOW IMPORTANT ARE THE SOCIAL MEDIA?

18. Social media sites are not terribly important at the moment, with the only possible exception being Facebook (see table below). Here are some observations on the visits from Facebook:

- Figures seem to increase gradually, and somewhat more rapidly than overall visitor figures: percentages are slowly going up, but they remain very modest; doubling from 0.1% to 0.2% does not make for a landslide change.
- A press peak causes a peak in Facebook entries (the February 2011 peak is the time of the El Pais minimum wages link - entries from Twitter also shot up in that same month, from a level of 2 to 8 a month to a peak of 52, but that represents only 0.05% of total visits!).
- Facebook obviously does not serve as a search engine (yet), but as a referring site where links are posted; in the context of referring sites (market share of 5%) it has considerably grown in importance and is now always in the top 10, occupying a third place, below Wikipedia and National statistical office links.

Month	Visits from Facebook	Total visits	%
Nov 2009	8	14729	0.1%
Dec 2009	6	20819	0.0%
Jan 2010	12	26717	0.0%
Feb 2010	13	31049	0.0%
Mar 2010	22	40316	0.1%
Apr 2010	17	40763	0.0%
May 2010	44	52454	0.1%
Jun 2010	25	41272	0.1%
Jul 2010	35	38847	0.1%
Aug 2010	73	41960	0.2%
Sep 2010	65	54688	0.1%
Oct 2010	66	68794	0.1%
Nov 2010	186	90434	0.2%
Dec 2010	71	74941	0.1%
Jan 2011	177	84548	0.2%
Feb 2011	**685**	103441	**0.7%**
Mar 2011	358	101990	0.4%
Apr 2011	213	92910	0.2%
May 2011	208	104114	0.2%

19. The two only other social media sites with noticeable impact are Twitter and Linkedin. But even at their peak levels, in February and March 2011 respectively, they do not account for more than 52 and 42 visits, representing 0.05% and 0.04% of visits. Peaks seem to be a rather weak reflection of media coverage and no upward trend is discernible.

20. From these figures, it seems unlikely that the social media will become a major channel in disseminating statistical information. People may be massively in the social media, but maybe not for the purpose of finding or transmitting statistical information. Not unlike a football or pop festival crowd.