

**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: Communicating statistics in the information age

**SOCIAL VISUALIZATION WEBSITES AND OFFICIAL STATISTICS:  
THE OECD EXPERIENCE**

Submitted by OECD<sup>1</sup>

**I. BACKGROUND**

1. The OECD Statistics Dissemination project was initiated in 2007 with the aim of reorganizing the statistical outputs into three services:
  - **OECD Statistics** (a portal giving access to all complete OECD databases, to be available on subscription using the free-at-the-point-of use model but subject to some Access Principles).
  - **OECD Core Data** (up to 1000 ready-made tables, with metadata, drawn from all OECD databases, aimed at students, informed and specialist audiences; to be freely available to all),
  - **OECD Facts and Figures** (a series of simple tables and graphs, with commentary, aimed at non-specialists and specialists; to be freely available to all).
2. The requirements for the third service “OECD Facts and Figures” are now being defined. To help its reflection in this area, OECD has launched several experiments to evaluate alternative ways of publishing data. One of these projects was to upload data tables on social visualization sites, [www.many-eyes.com](http://www.many-eyes.com) and [www.swivel.com](http://www.swivel.com).
3. Many-Eyes was created by information visualization experts from the IBM Collaborative User Experience research group, and it provides sophisticated visualization types. Swivel is a San Francisco start-up proposing simpler, more accessible visualizations which blend naturally in a web environment. Despite a very different culture and approach, the purpose of these sites is similar: enable users to upload datasets, create visualizations based on these datasets, and share their contributions with the world at large.
4. This paper studies various aspects of the co-operation of OECD with Many-Eyes and Swivel:
  - Why are we interested in such systems to publish data on the web?
  - What was the immediate outcome?
  - What have we learned?

---

<sup>1</sup> Prepared by Jérôme Cukier, OECD ([jerome.cukier@oecd.org](mailto:jerome.cukier@oecd.org))

## II. WHY PUT OECD DATA ON SOCIAL VISUALIZATION SITES?

### To reach a wider audience

5. The OECD has a core audience of experts who regularly use our data – economists, statisticians, and researchers – for whom the Organisation is a well-known, credible and reliable source of statistics. There is a second group, constituted of students, policy-makers and journalists, with whom OECD communicates directly. By embracing this new form of data dissemination, the OECD wanted to go a step further and interest people beyond those two groups – curious citizens, who already “consume” statistics through TV or the web.

### To turn statistics into knowledge

6. For non-experts, it can be difficult to draw conclusions from a given OECD dataset. But a graph created by a subject-matter expert to convey a certain message, and supplemented by a simple explanation as well as by links to further resources can help the user to understand the data. In fact, it is impossible to interest people beyond our traditional audiences if we cannot provide them with content in a form that is both engaging and easy to understand. Swivel.com and Many-eyes.com are able to help us to achieve that.

### To foster debates and interactions about data

7. The various OECD websites and dissemination platforms are very unilateral. Users of OECD statistics cannot react or respond directly on the site. This doesn't prevent them from doing so, by email or on other websites. OECD statistical content has also been difficult to cite on the web – bloggers who want to comment on a specific OECD graph often resort to sub-optimal solutions, like providing screenshots of Excel® graphs, or citing the entire document without being able to single out the graph. Here again, Swivel and Many-eyes can help as they can facilitate discussions about graphs, and they make them very easy to integrate in websites.

## III. THE EXPERIMENT AND ITS OUTCOME

### The material

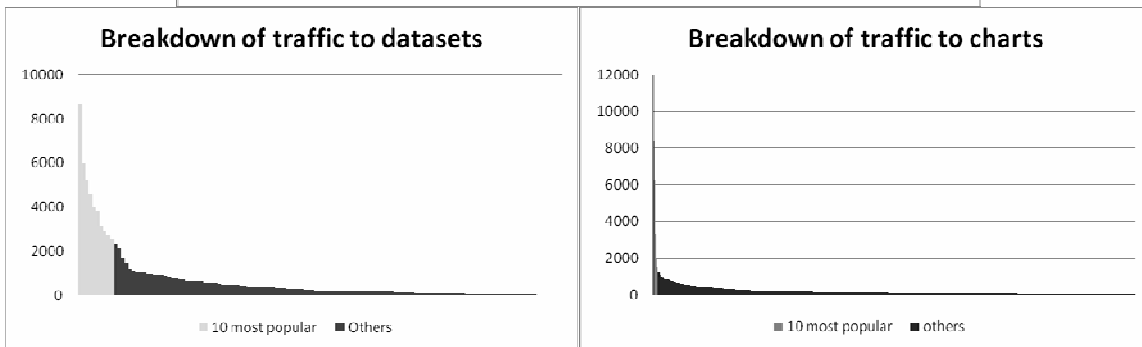
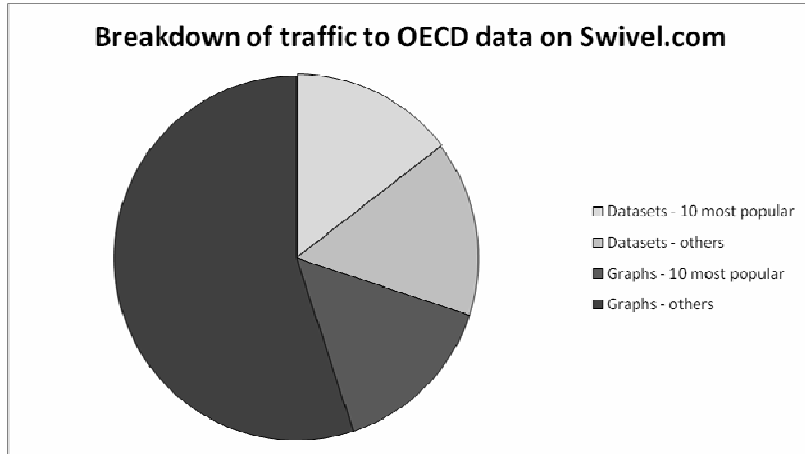
8. We have uploaded the contents of our highest-level publication, the OECD Factbook 2007, in both Many-Eyes and Swivel. The Factbook is a collection of sections that are representative of the scope of the OECD work. Each section is a page of text with one to three tables and graphs. While a page long, the text is understandable by non-specialists. The book contains:

- 85 texts,
- 135 tables and
- 155 graphs.

9. In Many-eyes.com and in Swivel.com alike, a graph needs to be supported by a matching dataset. So we had to upload all tables to datasets, but also the data underlying each graph, even though such data were often directly derived from those of the corresponding table. The difference between the two systems is that while, in Many-eyes, the user must choose exactly what type of graph they want to create, in Swivel, it is rather the system that generates graphs depending on how the users consult a dataset. When this is done, the user can edit the resulting graphs. For that reason, while there are only 165 OECD visualizations on Many-eyes, there are over 200 000 OECD graphs on Swivel, of which only 155 have been touched up by OECD.

### Traffic

10. The total number of hits on our datasets and graphs on Swivel is around 300 000. This compares with some 500 000 visits to the official Factbook website and some 650 000 queries made by external users to the country statistical profiles, which are the interactive database version of the Factbook. To put these numbers into perspective, the traffic on [www.oecd.org](http://www.oecd.org) is about 60 million hits per year.



**Top 10 OECD datasets and charts**

Datasets	visits	Charts	Visits
Gross national income per capita	8671	% obesity by country	1214 2
Population growth rate	5995	Gross domestic product of OECD countries, 2005	8895
Total fertility rates: Number of children born to women aged 15 to 49	5230	Obesity in developed countries	6696
Household net saving rates	4602	Statistic Value by Country	6489
Average Hours Actually Worked	4018	United States vs. Total fertility rates: [...]	3549
Road Fatalities per Million Population	3791	Unemployment Rates by Country	2900
Net Human Migration Rate	3140	Luxembourg by Switzerland	2038
Volume index of GDP per capita	2915	Overweight Percentages	2007
Unemployment rates of foreign-and native-born populations	2732	GDP per capita	1747
Gross domestic product	2587	World population by Country	1657

11. The main observation from the OECD Factbook traffic numbers is the large difference between the most visited datasets and graphs and the rest of the datasets and graphs. What is intriguing is that a few meaningless, computer-generated graphs managed to make it into the top 10 charts. “Luxembourg by Switzerland”, a senseless scatterplot, has accumulated over 2 000 views and 2 comments, making it our seventh most visited graph.

12. In Swivel, the default way in which datasets or graphs are presented is ranked by popularity (page views). This means that **an object** which manages to get enough views to make it **past a certain threshold** will subsequently **appear on top of searches**, which will in turn get it more opportunities to be viewed.

13. So how are objects viewed in the first place? There can be several sources:

- A graph can be featured by Swivel and appear on the home page, which gives it a boost of traffic.
- Users can also explore the contributions of one given author, which displays 12 graphs or datasets per page – objects on the first page gaining much more exposure than the others.
- Data objects can also be found by search engines, internal or external. Such searches are the way by which the less popular tables and graphs manage to attract traffic.
- Finally, since Swivel makes it easy for external sites to use Swivel content, these sites are another source of traffic directed at relevant content. Swivel also monitors the sites which use their data.

14. If we compare this to our other online representation of the Factbook, where every section is two clicks away from the home page, we reach the conclusion that on Swivel and Many-eyes 10% of our content get all the attention and 90% remain invisible. Items which happen to be at the bottom of that list would be difficult to access and are likely to be very seldom viewed.

15. The incoming traffic from Swivel or Many-Eyes to our online library, [www.sourceoecd.org](http://www.sourceoecd.org), was minimal: 1 400 visits from Swivel.com, ten times fewer from Many-Eyes.

### Interactions

16. We've had only 65 comments in total on OECD data posted on Swivel, 27 on datasets and 38 on graphs. That is roughly one comment per 5 000 visits. This is quite low compared to the 5% of visitors who contribute to non-OECD content on Wikipedia or even to the 0.6% who do so on YouTube.

17. Swivel does provide many ways for users to exchange with the authors of a dataset or with the community at large: beyond comments, users can rate the graph, feature it or propose new keywords. There are tools by which users can even re-appropriate those graphs – they can edit them and appear as the author of the resulting graph, they can easily import them on their blog or website, etc. On the author's side, it is possible to centrally monitor the activity around all of one's assets – each time a user interacts with a dataset or graph, this appears as an entry in a log. Many-Eyes also have similar features and the results have been comparable: few comments, by fewer people.

18. This low level of participation can be seen as a direct consequence of traffic. **All interaction is done at the object level**: users comment or rate a graph or a dataset, and in order to do so, they **must first discover it and access this object**. Measures have been taken by Many-Eyes and Swivel to address this and to spur a discussion among users with similar interest at a higher level. Swivel created groups, and Many-Eyes, topic hubs. Both are an open way to categorize content. If a user joins a group, they can affect relevant content to that group. Discussions can take place at a level above individual graphs and tables.

### Reactions of the statistical community

19. OECD, through its World Forums on Statistics, Knowledge and Policy, has committed itself to **defining key indicators to measure the progress of society**<sup>2</sup>. These **indicators will need to be communicated to a wider audience**, and the OECD wishes to find the best format to fulfil this ambition.

20. The experiment with loading OECD Factbook data on Swivel and Many-eyes received very mixed reactions, from vibrant enthusiasm to frank hostility, with arguments which were both technical and political.

---

<sup>2</sup> the full text of the Declaration of Istanbul can be seen at <http://www.oecd.org/oecdworldforum>

21. By the scale of the volume of statistics handled by official agencies, the experiment had a very limited scope. But it led to a tangible realization; it was there for anyone to see and to react upon. So while quantitatively, it can be difficult to measure, qualitatively it had the huge merit of **opening a discussion that is still going on**. It was arguably the first time that official statistics were published in a web 2.0 environment. Are statisticians, happy with that? Can we deliver more in this environment? Should we change our structure and the way we work to pursue this objective?

22. After the OECD, several organizations uploaded official datasets onto Swivel, including WHO, UNECE, UNESCO, several US bodies and EUROSTAT, which has uploaded more datasets than all of the other combined.

23. As of January 2008, there are over 14 000 references to OECD and Swivel outside of the OECD.org and Swivel.com domains, and close to 2 000 references to OECD and Many-eyes, again excluding the OECD.org and ibm.com domains<sup>3</sup>. This demonstrates the attention with which the statistical community is following this collaboration.

#### IV. FURTHER QUESTIONS

##### Media or content?

24. By disseminating OECD statistics through social visualization sites, we change the **form** in which we publish statistics. But the **contents** remained identical to what has been published elsewhere, instead of creating a custom collection tailored for the web. By OECD standards, one full page of text to support a graph is considered short and simple. Likewise, graphs with 30 to 100 datapoints, or plain 2-dimensional tables with 500 cells are among our simplest published statistical products.

25. But on the web, these outputs compete with graphs with 5 data points and tables with 10 cells – all described in a paragraph, or less. Objects from our data collection cannot be understood in a one-second time lapse. They are too bulky to reuse as is by a non-expert user, and too authoritative to openly criticize. In other words, existing OECD statistical outputs have little interaction potential.

##### How to organize large collections of statistical objects

26. Objects – that is, datasets and graphs/visualizations – appear to be primarily accessed from the homepage of Swivel or Many-Eyes, either from a list of all material from a user, or through an internal search engine query. Both methods yield **many results**. We have noticed that objects which were featured in the first page of results had a much better exposure than those which were not, the latter being practically invisible. The number of objects which will be visible doesn't depend on the total volume of data. If twenty tables receive all the exposure out of the 135 tables of the Factbook, that leaves only 16% of our content visible (approximately 8 000 datapoints out of 50 000). But that proportion would **sink to near zero** if we were to upload the whole data collection of the OECD, which is estimated at 2 billion datapoints.

27. Many-Eyes and Swivel seem geared towards users who would contribute with a small number of objects. They both provide systems (groups, topic hubs) to group the work of many users in a single workspace. What they don't allow is for one user with a lot of material to split and hierarchize their work. 135 tables and 155 graphs can seem a lot when they **are spread across a flat list that covers many pages**. But in the book, all of this material is distributed in 12 chapters, which are comprised of two to five sections, each of which contains two to five sections, which group one to three tables and one to four graphs.

28. **In theory, the tag system should supersede a rigid hierarchy**, and works well in web 2.0 sites, like Flickr or YouTube, where one user publishes a very heterogeneous content. In that case, keywords help users to

---

<sup>3</sup> <http://www.google.com/search?q=swivel+oecd+-site%3Aoecd.org+-site%3Aswivel.com>  
<http://www.google.com/search?q=many-eyes+oecd+-site%3Aoecd.org+-site%3Aibm.com>

make the necessary distinctions and find what is needed. Also, in those contexts, the user identity is less important than the description of the content, especially when the contributor is not the author of the content and merely shares it.

29. In official statistics however, all tables of one collection should be very homogeneous and receive the same keywords. Moreover, the identity of the user is at least as important as the keywords, e.g. the GDP data published by OECD is not necessarily the same as the GDP data published by the IMF or by national sources.

30. Finally, official statisticians may not be the best at guessing what keywords non-specialist users are likely to look for. A table could be described as dealing with “consumer price indices”, when non-experts would look for inflation or cost of living. In this context, **keyword navigation doesn’t deliver as much value as it could**. It can be a valuable tool provided users know what they are looking for, but it should not be the only way to drill down into statistical content.

### Traffic analysis

31. The analysis of the traffic raises two main questions:

**(i) Why is traffic to our datasets on Swivel.com not significantly higher than traffic to our traditional sites**, despite a much stronger design, a better integration in the web landscape and a decent level of PR exposure? Four explanations come to mind:

- The size of our collection prevented most datasets from receiving the necessary exposure to collect traffic through the internal search engine, as mentioned above.
- There is a misfit between the content and the intended audience. This makes our data not only less engaging to use on the site, but more difficult to re-use elsewhere – blogging a single, simple graph, for instance, is easy, but having to integrate a complex graph that needs one page of text to support it is not practical.
- Towards the beginning of the experiment our Swivel datasets were fairly visible on Google – some were consistently on the first page of results. That was because each was well described with detailed metadata such as keywords. But their algorithm increasingly takes into account the freshness of a page and as our data got old, the rank of the corresponding datasets dropped drastically in Google.
- A final hypothesis is that Swivel.com visitors seem more interested in Swivel.com, the service and the concept, than Swivel.com, the content, and don’t necessarily visit the service looking for data in a specific area, but rather are attracted by topical subjects and featured graphs.

**(ii) Why did the top 10 datasets and the top 10 graphs behave so differently?**

The hypothesis that people are more interested in the service than the content is reinforced by the finding that the top datasets and top graphs just don’t match. Datasets are visited by people who are actively looking for particular content. But popular charts are pushed to users – they see them featured on the portal, or posted on a blog, and they follow the links.

### How to name a statistical object for maximum impact

32. At the OECD, we have strict guidelines when it comes to naming tables and graphs. These objects are destined to be **found in tables of contents**, so we provide **very descriptive names** that help readers of tables of contents to quickly identify the item they are looking at. For objects found in tables of contents, that is, in smaller lists or in lists that can be broken down in smaller, manageable chunks, our approach is best. But **for objects which appear in search results, or news feeds**, objects which compete with others on the same level for the attention of the reader, we have noticed that **using an engaging sentence proves to be more effective**.

## How to thrive in a Web 2.0 environment

33. Intergovernmental organizations have little flexibility when it comes to assigning resources to unexpected tasks. Unfortunately, merely “dumping” content on the Internet will not produce any significant outcome unless the content is animated by human moderators and writers. In the words of veteran blogger John Biggs, **there is no surer way of getting traffic than having a constantly updated front page**. But frequently updating content, that is, the action of permanently editing fresh content to adapt it to the web environment, as opposed to uploading a large quantity of data once a year, is exactly what a structure like ours is not designed to do. If instead of uploading the Factbook at once, we had uploaded one graph every couple of days throughout the year, traffic would probably have been of a different scale of magnitude – as would the percentage of users who actually go deeper and complement their visit to Swivel or Many-Eyes by a tour of the OECD resources.

34. The same applies to comments. **We have had a very passive attitude towards comments**. Some comments are spam. Many of the remaining comments are largely irrelevant. And the comments which are relevant are usually met by a quick reaction from within the statistical community. As a result, the OECD hasn't responded to many comments, thus reinforcing the impression of inertia of our datasets.

35. The initial commentary that accompanied our datasets was copied directly from the Factbook. A fairly lengthy text for the web, it was descriptive, authoritative and objective, but not thought-provoking. It wasn't useful as a conversation starter, which would draw reactions and spark a debate. We didn't even provide such a commentary for graphs or tables in the Many-Eyes environment, which also hindered significantly the possibilities for discussion.

What is our role in this environment? Can we just be providers of raw content? If we step outside of this role and add an editorial layer to the numbers, are we losing our objectivity?

## V. WHAT NEXT?

36. For the reasons outlined above we don't think that uploading Factbook 2008 datasets in the exact same conditions as last year will be very effective. Not only is this data collection not well adapted to this exercise, but posting it all at once proved to be less than optimal.

37. However, our faith in the service offered by Swivel.com and Many-eyes.com remains intact. We feel they are fantastic services for publishing simple datasets on line. Many blogs or web-centric media, including prominent ones, publish data on a regular basis but resort to primitive methods to that effect – such as a low-resolution, bitmap image of an Excel® graph. We also feel that in the fields of virtual learning environments, especially with undergraduates where our own statistical dissemination systems prove to be too complex to be useful, the match with Swivel.com and Many-eyes.com would be perfect.

38. So this year, we are planning to use these services to power a “Factbook blog<sup>4</sup>” with very simple datasets and graphs, taken from the OECD Factbook and similar publications, but posted on a very regular basis throughout the year. We intend to use the lessons learned in this experiment to make this service as effective, as engaging and as interesting to our audience as possible.

\*\*\*\*\*

---

<sup>4</sup> <http://oecdfactbook.wordpress.com>.