Open Data in real life: some reasons to keep it closed

Submitted on 18 Mar 2015 by David Osimo

I have just registered my kid for school in Barcelona, it was an exhausting experience. Basically, if you have a two year old and live in Barcelona, then the chances are you've devoted the last three months mostly to this.

By the end of the process, I learnt what data are most used to help you choose the school for you:

- Number of teachers per pupil
- Number of external activities per trimester
- Non-usage of textbook in primary school (indicator of openness / modernity)
- Socio-economic standing of the families
- Facilities (garden, sport, etc) but even basic things such as windows in rooms (some schools don't have it)
- Own kitchen
- Additional services (e.g. logopedy)
- Average results of kids in secondary school test

On top of this basic data, many also use the feedback of other parents – provided they think like you.

But because of the procedure used for subscription, the fundamental datum is oversubscription rate. You should choose a school that is not oversubscribed, and therefore the oversubscription rate is fundamental.

Obviously, these data are often not available and you have to chase them. Although I expect that an app that gathers such data will soon be made available.

But the fundamental point is that these data are not being published in an easily readable format **on purpose**. The argument is that **if data, and in particular the average results of pupils, were made available, parents will flock to the best performing schools and enhance the inequalities of the system.**

In other words, **transparency would create a "rich getting richer" effect**. The difficulty in getting the data ensures that only those who are really interested are getting this information.

Is this argument sound? Is it applicable to other domains? What are the main counterarguments? For sure this can be applied to mortality rates of hospitals.

It reminds me of another apparently sound argument for limiting the openness, i.e. the embargo period for scientific publication coming from specific datasets in order to allow the author of the dataset to write his conclusions. This is also recognised as a valid limitation, but is it really so?

I wrote in the past on the need for "sharing literacy" rather than sharing culture, because one should know when and where to be open.

Can we identify a set of valid arguments to limit openness across domains?

(Note: this post was originally published on Benchmarking e-Government in Web 2.0's blog)