

TOWARDS CORPORATE AND USER-GENERATED OPEN DATA: BENEFITS AND ISSUES AROUND THE EXTENSION OF OPEN DATA

Daniel Kaplan, CEO, Fing

Open Data Week Keynote Speech – Nantes, May 25, 2012

So we want to expand the amount and the diversity of open data, be they data "released" from their corporate vaults or natively open data, contributed by user groups in the likes of Open Street Maps and Wikidata.

We assume that is a good thing to do. But why? Especially on the corporate side, why should they contribute to open data? Sure, companies like to be good citizens, but let's not kid ourselves, this will never be the key reason for their actions.

We have never taken data for serious

The first reason, to me, is that we've never taken data for serious. And now they're kicking back.

Data are for the most part ad hoc constructions designed to fill variables in programs. They are narrowly defined according to the program's goals. Because programs are often designed as if they were alone in the world, their data are inconsistent duplicates of similar data all over the organization; often, they are also maximized: why bother selecting only the really necessary data out of consideration for the time of those who will input them, since no other program exists? More often than not, data are also badly maintained and tend to outlive their relevance and usefulness.

However, today, innovation becomes open; most value chains and service models combine dozens of organizations; the Web is one of mashups, applications and APIs. And, on the users' side, "Personal data stores" and "Vendor relationship management" (VRM) are changing the dynamics of interaction: Corporations are not the only ones to use information systems, customers have their own, too.

This new world of co-operating, loosely coupled organizations / projects / services / applications pivots around data and "services" (as in "web services"). It needs data that are as reliable, consistent, precisely described, linked, accessible (under clear conditions) and cheap as possible.

Trivializing openness

Therefore, making sure that many of their data are somehow "open" will soon be an ordinary business decision.

Information systems will *really* separate data and processes, something design methodologies have advocated for decades, with frustrating results. Data will exist in common servers and "stores", with visible and usable access mechanisms and rules. Decisions on openness will not be on/off decisions, rather matters of degree: open to other groups within the organization? Well-known partners? Broader participants in our ecosystem? Paying guests? Everyone?

Corporations will also constantly be accessing external data, especially the personal data stores of their customers and users. And to gain access, they will need to reciprocate.

Accountability, transparency, CSR, will be part of that trend, but will (as usual) play a rather secondary part.

It's just how we'll work. Because the alternative is just too dumb and inefficient.

The real issues with open corporate data

So the real issue is not whether companies will open their data: they will, with a few rules attached. Nor whether they will sell their data: they mostly won't, because "information wants to be free" (Steward Brand wrote that in... 1987!) and its business value lies in its uses, not its raw existence.

So what are the real issues? I'll concentrate on two of them: Minimization, and Empowerment.

Minimization

Open Data is often related to the world of "Big Data", a recent trend in computing that promises to usher a new world of knowledge, of previously unseen correlations, of predictability.

However, Big Data's main field of application is in fact marketing, where, for privacy reason, another key word today is also *data minimization*. "Privacy by design" is about collecting as little personal data as possible to get anything done, using even less, and keeping it for just as long as it is needed.

Sure, most open data are supposedly *not* personal. But there are plenty of examples where the aggregation of disparate "anonymous" data provides pretty good information of individual situations and behaviors. In the world of Big Data, most data can get personal.

Minimization doesn't only make ethical sense, it also makes business sense: less clutter, less chances for mistakes, less complexity, more trust. When you minimize what you know, you make it so much easier to become transparent if you're asked to.

And there's another reason for "minimizing", i.e., thinking hard about what data you *really* need to get a specific job done. Data are constructs. They are created by specific people, artefacts and interfaces, for specific reasons, in specific contexts, under specific constraints. Unthinkingly reusing them in other situations is often likely to give us a misleading sense of security about how tangible and measurable reality is. It may induce us to automate too much of our decisions or relationships.

The tension between open data and minimization is a productive one. It helps us think. It should not stop us from sharing data. We should just keep in mind that it's not because data become more open and usable, that we should turn everything into data.

Empowerment

Who does openness empower? We'd like to believe it's everybody. However, it may not be the case unless we really want to make it happen.

Initially, opening something up often serves to empower the empowered: Those who have the means and the know-how, those who have known it was going to happen and planned accordingly. As an example, several studies on the release of landowning and housing data have shown that they were used to evict the poor from lands they occupied informally, or (in more urban areas) those who did not pay the maximum rents that could be expected in their neighborhood.

Then it tends to pave the way for new dominant player who have proved better at using the new resource – think Google...

If these were the only results of open data, it would defeat its whole purpose. The point of opening data is to generate as many productive interpretations and uses as possible (which means we also need a lot of unproductive ones); to get people to relate and contribute to the company's data; to create fertile ground for interactions as well as discussions on common issues.

If data, although open, remains inaccessible and mysterious to most people, we will not benefit from the innovation, the contribution, the additional proximity, and the quality discussion we expect from releasing them.

Therefore, we need to be active in creating platforms, tools, places, in which the existence and meaning of open data becomes visible, and that help people use them. We need to create a culture of data among all kinds of players: value chain partners, public agencies, "traditional" SMEs and media, NGOs, consumer groups... We need data mediators and dataactivists, within and without our organizations. Open innovation needs innovators and an ecosystem that allows them to meet with one another and with the larger players. Open data needs the same.