Proposal for a Session at the Lisbon Workshop: Encouraging open data usage by commercial developers

Version 2.0 (27/11/2014)

Open Market Dilemmas

Hashtag: #OpenMarketDilemmas

While certainly data markets have existed long before the advent of the internet (matching information supply and demand has been driving many media revolutions from newspapers to the telegraph), the wholesale exchange of large structured dataset is relatively new.
The Deloitte Study „Market assessment of public sector information“ for the UK Department for Business, Information and Skills identifies data marketplaces and data enrichment as important business models utilizing open data. Many early entrants have been forced to shift their focus from “statistics on speed” (timetrics) to more cautious approaches or they left the market altogether. Data enrichers - often specialist for a narrow business area (like geo marketing) seem to fare better but here the public data sources are often not spelled out in detail. All this may be because these business models and open data policies geared towards them face a couple of dilemmas often overlooked.

This interactive session addresses practical problems stemming the interplay of licensing policies and commercial re-use as well as the functional design of public sector data portals themselves resembling data marketplaces.

Target audience:

Entrepreneurs and start uppers
Policy makers involved in drafting Open Data Policies on all levels of government
Researchers interested in PSI Reuse business models
Intellectual Property lawyers

Background

Internet Market places abound

- airbnb connects owners of spare rooms to enterprising travellers.
- expedia connects airplane ticket buyers to fares and routes.
- uShip connects specialty shipping companies with people who need to ship things UPS doesn't accept (like motorcycles or horses).
- SpareFoot connects available public storage units with people trying to find units online.
- angel.co connects investors and startups (among them 3.446 Marketplaces)

What do data marketplaces do?

- They function as search engines for (public or exclusive) datasets or data services, API access,
- they provide quality indication for data (often crowdsourced),
- they facilitate the comparison of datasets,
- they allow the download of data,

How do they do it?

They build an enormous collection of structured data through

- automated methods,
- editorial work and
- crowd sourced commits and edits.

How do they earn money?

- They charge (buyers, sellers or both) a commission of any data that is sold on the site or after referral,
- they charge sellers for listing (akin to the yellow pages),
- they distribute premium data for a commission or sell value added custumer services (the so called Freemium model),
- they monetize t traffic with targeted advertising (rarely),
- they earn money through other cloud services, which become more attractive to developers through the ready availability of third party data,
- they trade data for data or
- they do nothing of the kind and , burn investor money and hope for the future

Proposed Agenda

1. Introduction of Participants
2. Collection of topics to be discussed (Brainstorming)

Chicken and egg vs. market forclosure
Like all marketplaces, those dealing with data initially face a classic chicken-egg problem: Specifically, buyers don't visit the site because you're obscure and lack inventory, but sellers aren't interested in listing because there's no buyers. Furthermore, with limited mainstream market participation in the business of data, it is difficult to assess what a given dataset is worth. When you set up public data portals you seemingly make it easier for private data market places to enlarge their inventory. Yet on the other hand the very marketplaces might suddenly have to compete with tax financed portals, not only diminishing their returns but also cannibalizing the nascent market.

Example questions:

- Is the call for public sector information portals premature?
- Can public sector information portals through an unfair economic advantage deny commercial competitors the necessary access to suppliers an buyers of data?
- Where can public portals help and which services should better be left to private data exchanges?
- Would it be better for the state to pay commission fees?

Technical restriction prohibition and share alike vs. monetization of added value.
If your business model relies on selling the added value that stems from aggregating and combining several open data sets it may fail if the licenses the open data are supplied require you to make the data freely available?

Example Questions

- Which licenses in which areas carry such risk?
- Are there examples of Open Data Policies addressing such risks?
- What should businesses sell: The enriched data o

Privacy vs. information density
It is generally agreed upon, that open data shall respect the privacy of the people the data is derived from. Yet it is a fact that the very information that is useful to business is also the information responsible for the mosaic effect, i.e. facilitating the re-identification of anonymized data.

Example Questions

- How does anonymization affect the usefulness/value of open data?
- Which anonymization techniques diminish the values more than necessary, which retain more “information” density?
- Is the perceived “mosaic effect” a potential barrier to more open data?
- Do we need regulation (e.g. by certification) which companies or types of service are allowed to deal with data potentially contributing to the “mosaic effect”?

Attribution vs. trade secrets

Licenses that require attribution might force business to give away trade secrets. If your company sells the spending power data derived from public sources (e.g. the number of upper class cars licensed in the area), revealing your data sources amounts to giving the very formula your business relies on.

Example Questions:

- How can one protect ones business from the risk of „attribution“ leakage?
- Which licenses in which areas carry such risk?
- Are there examples of Open Data Policies addressing such risks?
-

Moving the data vs. moving the computation

Making public service information available is more or less always equated with the release of the data to the public. However, as datasets grow in volume, not only the necessary data traffic rises but the data analysis might be executed most efficiently on distributed storage with distributed processing of the data Big Data on clusters of commodity hardware. In such circumstances it might be better to move the computation. This might give the data owner also more control over the use of the data where issues of privacy arise.

Example questions

- Could a strategy of providing a PSI Cloud that accepts computational payloads be a viable complement to releasing the data?
- What kind of practical applications profit from such an approach?

3. Voting on Agenda
4. Discussion
5. Conclusion/Recommendations

Facilitator: Dietmar Gattwinkel

Projektleiter Open Government Data | Project Manager Open Government Data

STAATSBETRIEB SÄCHSISCHE INFORMATIK DIENSTE | SAXON IT SERVICES
Riesaer Straße 7 Haus D | 01129 Dresden
www.opendata.sachsen.de |dietmar.gattwinkel@sid.sachsen.de | opendata@sid.sachsen.de

Background

Dietmar Gattwinkel is Project Manager of the Open Government Data Project in the Free State of Saxony. Into this project he brings 12 year of experience in setting out Saxony's Web Strategy and a strong involvement in the overall e-government process. He is also representing Saxony in the project group implementing Germany's open data portal “govdata.de”. Prior to his work for the government he worked for a geo marketing company that pioneered the reuse of PSI in Germany. He holds a master’s degree in Communications, Law and Philosophy from Johannes-Gutenberg-Universität Mainz.