Open Market Dilemmas

Attendees:

Simon Whitehouse, Peter Winstanley, Amanda Smith, Dolores Hernandez Marot, Noël Van Herreweghe, Makx Dekkers, Clemens Wass, Georg Hittmair and others

Introduction

Dietmar Gattwinkel, is the project leader of Open Data in Saxony. A law was just passed fostering open data. Some years ago he was also working for a company marketing data. From 2011: Doing business with PSI.

The session will deal with intermedia in a literal sense, i.e. those companies (slide) do not create data themselves.

Slide 3 -

Matching demand and supply e.g. AirBNB, angel.co

Marketplaces are one of the most tried business ideas on the web. – Why should it not work for open data → Data Market places.

Data enrichment providers

Share-PSI 2.0

What do they do?	How do they do it?	How do they earn money?
▮ They collect data necessary for a certain task	They large datasets enhanced through e. g.	▮ Sale of datasets,
▮ They enhance, refine or otherwise improve this data,	▮ extrapolation,	▮ Service charges,
▮ They sell datasets or services specifically de-signed for the needs of a certain business.	▮ error correction,	▮ Sale of software that uses the data.
	▮ matching with proprietary data.	

3 | 4. December 2014 | Facilitator: Dietmar Gattwinkel

Slide 4 – data enrichment providers

What do they do, How do they do it, How do they earn the money?

GeoMarketing – data examples

Share-PSI 2.0

Data on precinct or street level
- ▮ Retail Turnover
  - ▪ Food, Construction/DIY, Books, Garden Centres, Watches and jewellery ...
- ▮ **Number of residential buildings**
- ▮ **Household inhabitants**
- ▮ **Approved construction projects for buildings**
- ▮ Living environment typology
- ▮ Purchasing Power
- ▮ …

Data on business companies
- ▮ Description
- ▮ **Legal form**
- ▮ **Date of incorporation**
- ▮ Number of employees
- ▮ Financial ratios
- ▮ Rate of change
- ▮ Auditors
- ▮ Listing on Stock Exchange
- ▮ Shareholders
- ▮ …

3 | 4. December 2014 | Facilitator: Dietmar Gattwinkel

Slide 5 – GeoMarketing – data examples

Many player out there in a competitive market

Overview over proposed agenda:

Chicken and egg vs. market foreclosure

Is setting up an open data portal a competition to the industry?

Or is it necessary because it is so complex: demand will not come without supply. Supply will not come without demand

PSI Alliance Georg Hittmar: According to the PSI directive you cannot restrict access via technical restrictions – Share alike or not?

Privacy vs. Information density:

The more we anonymise data the less valuable it is for industry? How do we reconcile data protection and business demands?

Question: Any more subjects to suggest?

If you want to foster entrepreneur ship you have to distinguish the sector (legal information, geomarketing, chemical, etc)

Georg Hittmair: When we go to the end user a lot of promotion will be necessary. Political activity is to enable the data, not to invest or foster market places.

- Voting on topics to discuss –

Topic 1 – Chicken egg vs. market forclosure	11
Topic 2 – TRP share alike vs. monetization of added value	7
Topic 3 – Attribution vs. trade secret	5
Topic 4 – Moving the data vs. moving the computation of data	8
Topic 5 – Privacy vs. information density	7
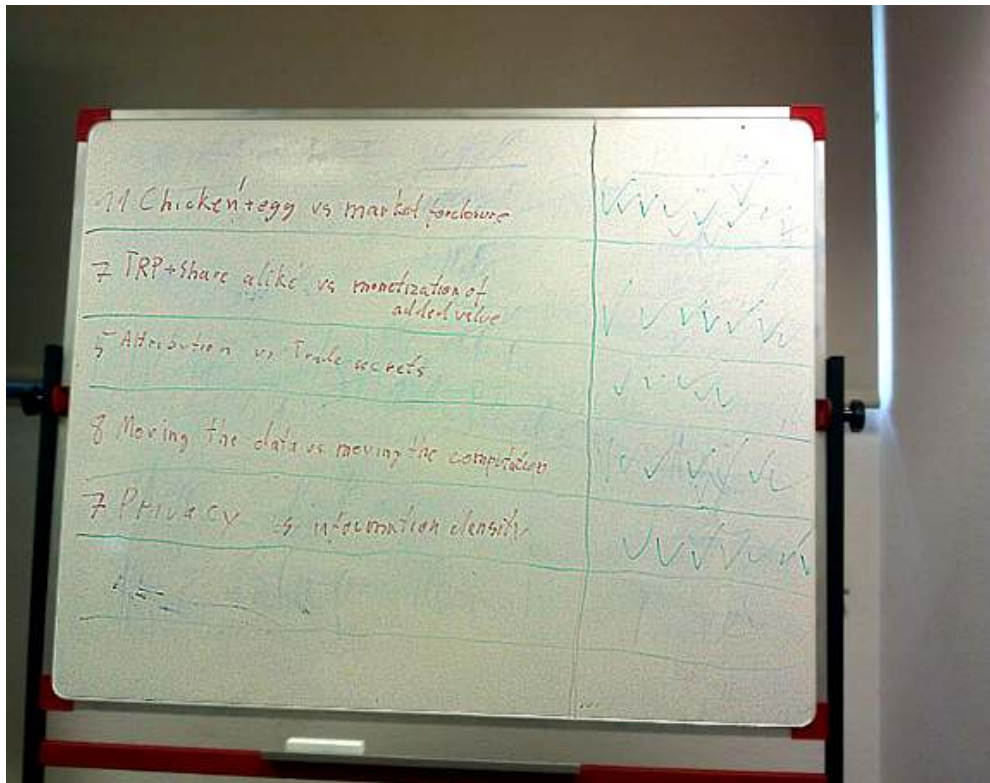
Discussion

Noël van Herreweghe: Are we talking about the ex-amount of open data portals? E.g. the region of Flandern has an open data portal. I do not want to see 100s of open data portals.

Dietmar: Open data portal by the administration only show what data is out there - everything else is a marketplace.

Noël: For every region in Belgium we have an own portal – it is difficult to agree on a common DCAT scheme. Data brokers is different to data portals. We aggregate the city portals –

Makx: I am working for the semic programme. I think there is a large market for data brokerage. I don't think that government organisation have time, money and knowledge.

Portal is good – Maintenance is one key question. The data brokerage is helped by having open data portals.

Noël: This is a business2business question - We do not want our companies to have to search for the responsible public agencies.

Georg: We are talking about the wrong end of the value chain – [Public agencies] should not care for business models.

Clemens: When I started an app for the Austrians Law. I took the telephone and I ask the ministry (without a portal) can I get the data please?

Noël: If there is one point ok – if there is a complicated system as in Belgium –

Makx: When a government organisation is creating a portal to foster the market – it is wrong – government portals should be used by government – side effect is that companies might use this data

[…]

Georg: Public data portals often present their data as gifts. Community comes together every 3 months.

Noël: It is not your data [public organisation] but it is our data. Portals do facilitate the access.

Hermandez Dolores: In the Spanish open data portal we do the same – 4 different levels of administration in Spain.

Noël: Only by way of our portal, I can control that our agencies do not create any apps but to leave this to the market.

Clemens: Is there no point where government apps come into play?

Noël: I tell my agencies to let this to the private sector.

Dietmar: When preparing for this session I got to know http://enigma.io/. They collect data and have a very nice user interface. A company creating open data portal would always use use sales force and promotion, s.th. that it difficult for the public administration.

Noël: - Example of Belgium: Analysis of legislation changes – every month this subscription product informs you on legislation changes (e.g.

Makx: Enigma.io as far as I understand is a data broker. Enigma.io sells aggregation services.

Dietmar: It is sort of a freemium service. To support such private offers, my suggestion is that public portals concentrate on the API – I have to deal with politician – they always want a nice interface.

Noël: Exactly

Peter Winstanley: I have not heard a lot of discussion of machines gathering data – a 21 century . I haven't heard about content delivery networks – We do not talk about what to do to optimise for a global market. Two kind of questions: Machine based discovery – standard access to DCAT repositories – and Content Delivera Networks CDNs. What I proposed as Barcamp: What is the infrastructure we need to foster open data. If I got Terrabytes and Petabytes of data. Do we forget about big data?

Dietmar: What you mean is whether we are concentrating to much on todays technology, when there are already other technologies around the corner?

Clemens: This sound fancy but it is 3 steps ahead. As an entrepreneur, I am happy toeven get what I asked for a year ago.

Peter Winstanley: As the government, do you want to provide what companies asked you 1 year ago or do you want to provide, what will be needed in the future.I want to make the barrier as low as possible.

Noël: How does this relate to the API of a European data portals?

Peter: We need to put in place infrastructure and standards – phoning up to a desk is simply no solution.

Spain: The situation you describe is great – but me as a entrepreneur – I currently have to fight for the data.

Peter: I do not want to do something that was asked last year - What should the next step be?

Dietmar: Let's dive into this very interesting topic in the bar camp and let's move to the second topic "Moving data vs. moving the computation". The whole concept in the last two days is about "How do we get the data out to people willing to use it?". If you think on plethora of data our traffic agencies have -> Petabytes easily come together if you think about real time data. This currently is happening – often cloud based – all the customer has to do is to send in the computation. Is this an infrastructure for the future?

Makx: Your example of traffic data is a bit unfortunate- I use google and not public data for traffic data. I think you are right you need something to provide the data and I am not…

Noël: You should be careful of mixing big data, linked data and open data and Internet of things. I am not saying that this will have. Why should government bodies provide a computational service to process data?

Georg: It could make sense that government bodies do this kind of thing.

Peter: It already works with full text search: You send a query to the organisation and the organisation processes your query. This is often better than what does google does. We used to the notion of sending a data search to a chunk of data. It is all data it does not matter whether it is text or numbers – we are already doing it with Small Area Statistics. We got quite a history of doing that – what we need are standards – like a vocabulary for searches.

Dietmar: Thank you for this example – even being responsible for the search engine - I never thought of it as an example for computation at the data site.

Peter: There must be a point when organisations come up with such solutions.

Dietmar: There might be another reason: If you are not moving the data sets but compute yourself you have more control. This brings us to the next topic: "Privacy vs. Information density"

If you want to match data – you have to have some indicators that helps you in matching, e.g. the gender of the respondent. The more you anonymize the data, the lesser value it has for data enrichment. On the other hand there are examples of reindividualising anonymous data, for example movie ratings (I think it happened as part of an app competition).

Amanda: I disagree – that anonymous data becomes useless – It all depends on the granularity.

What I do is publishing crime records on a very low level – but anonymised.

Dietmar: Are there any checks that we can provide?

Amanda: It depends on the safeguard you have in place. It is the responsibility of the government, not the private sector user.

Noël: We have a rule of three in Belgium: If you can bring it down to 3 people that it is privacy. Also we should treat that as a risK. If we fear risks we should stop building roads.

Peter: Another thing you can do with linked data.

Amanda: We cannot publish court details –but other people can go to the court and get even more detailed information.

Peter: There are ways that one can put a kind of a safety and allow for open data.

Noël: I think there is a large difference between the British approach and the Flemish approach to data privacy.

Amanda: Just as there is a huge difference between the British an the American approach.

Noël: Entities will hide data sets behind the privacy concerns. In Belgium we put that in the license, that you are not allowed to re-identify the data.

Noël: Do you know about any portal that has a share-alike licence.

Dietmar: I definitely saw individual data sets that required share alike.

Noël: Which entrepreneur would put a business model on share-alike licence data sets? Share-alike license is kind of a stupid license for open data!

Makx: The directive says it should be available commercial and non-commercial.

Open Market Dilemmas

Dietmar Gattwinkel, Saxony Open Government Data [paper]
Scribe: Sebastian Sklarß,]init[[notes] (WB)

1. **What** X is the thing that should be done to publish or reuse PSI?
2. **Why** does X facilitate the publication or reuse of PSI?
3. **How** can one achieve X and how can you measure or test it?

Topic 1 – Chicken egg vs. market foreclosure	**What**? Distinguish between the task of a data broker and that of an open data portal. **Why**? Because they are not designed for data brokerage, rather they should help government to streamline its own processes, provide a single, reliable point of access for government data and help managing the open data progress (e.g. prevent departments or agencies from prematurely developing their own apps. Also because concentrating solely on optimizing end enriching todays portals with brokerage functionalities might distract from advancing the technologically possible services (e.g. Machine based discovery – standard access to DCAT repositories – Content delivery Networks) **How?** Withstand pressure to come up with brokerage solutions, instead provide Services and APIs commercial Data Brokers can use.
Topic 4 – Moving the data vs. moving the computation of data	**What?** Start thinking about accepting computational queries and delivering results instead of data (as we already do with our search engines, only much more elaborate). **Why?** Because it is much more efficient when it comes to large datasets and might also be a solution to privacy concerns as it allows more control. **How?** Don't be content with today's solutions (meta data catalogues), but keep an eye on data industry developments already underway.
Topic 5 – Privacy vs. information density	**What?** Treat privacy issue as a risk to be managed, not a yes/dichotomy. **Why?** Because privacy concerns can easily be used as a smokescreen for other motivations not to publish PSI, but is really a matter of the granularity of data. **How?** Keep in mind cultural differences. Not everything that

	is considered possible in one country is feasible in another. Include the prohibition of re-identification in the use condition, or make it even a criminal offence.
Topic 2 – TRP share alike vs. monetization of added value	Do not demand share alike for open data as it will prevent commercial usage.