

“Towards self-sustaining data reuse in Europe”

18th of June Madrid, 2015

The EuDEco initial clustering workshop was a highly interactive half-day event held in co-location with the Big Data Value Association Summit in Madrid on the 18th of June 2015.

The objectives of the initial clustering workshop were to kick-start the EuDEco User Expert Group (UEG) and to share ideas and experiences on data reuse in Europe. Most importantly, the EuDEco project used this opportunity to present its approach and preliminary results, especially on the research framework and the analysis of framework conditions.

The first session of the event comprised:

- A welcoming speech and an introduction of the agenda;
- A speed dating session for all participants to get a small introduction of each other;
- A presentation of the EuDEco project as a whole;
- Presentations of the legal, socio-economic and technological challenges towards a self-sustaining data economy in Europe.

The second session consisted of parallel sessions where participants were free to circulate between 3 tables (20min / table) focused on the legal, socio-economic and technological solutions towards self-sustaining data reuse in Europe. The second session ended with an open discussion of the solutions provided during the parallel sessions.

Event output and outcomes

The workshop was significant for getting a large number of feedbacks on several topics and helped clarify a number of legal, technological and socio-economic questions regarding the data economy in Europe. The workshop was a successful event with interesting and

Event report

informative talks, which stimulated very active discussions. A total of 14 participants attended the first clustering workshop, consisting mainly of the EuDEco UEG members, but also attendees of the BDVA summit who were interested to join.

Following are some key takeaways from the initial clustering workshop table sessions.

Challenges and barriers discussed at the technological table:

- The choice of outsourcing data (for computing or storage) has increased and is nowadays used by many companies and research institutes.
- With the data storage becoming cheaper and cheaper, industry faces a trend – since a few years ago – to collect as much data as possible without any pre-filtering. This tendency is still very intense, but companies are slowly moving to a point of only collecting those datasets that they actually need, especially because of three facts: irrelevant data, data duplication and outdated data.
- The first challenge mentioned by participants was the notorious difference of technology status and the technological gap between SMEs and the big players.
- The second challenge was the prediction of needed traffic, because a small structure or a trial can easily grow to a big scale and this change may not be foreseen.
- In data transmission the challenge is the big variety of formats, and the fact that there is no standard exchange format.
- Another key barrier identified was different accuracy standards. Frequently different kinds of data are required to be managed based on different purposes.

Challenges and barriers discussed at the legal table:

- Privacy is not seen as the most relevant area. The burning issue is licensing. In particular the differences in approaches to licensing (e.g. in Eastern Europe and the

Event report

UK are considered problematic).

- Law is often associated with high legal costs and “black suits”. Participants felt that it is necessary to bring lawyers closer to the technology.
- When dealing with large amounts of data it is impossible to be clear on intellectual property rights and to contact and to pay royalties to everyone involved. There are businesses that basically make revenue out of deficiencies of the legal restrictions.
- The participants recognized a need to share the data although there are barriers such as differences in the industries, no clear legal definitions etc. A checklist of requirements would be helpful.
- The participants pointed out the difficulties they have with terms and conditions of data sources (e.g. Twitter, Facebook). Terms are not clear enough, too long and restrictive (e.g. some data sources do not allow you to delete the data).
- Participants also mentioned that technology focused companies’ concentrate on development and simply ignore legal restrictions. They rather take risk than seek for legal support.
- They also identified problems with data aggregation. When combining public Twitter posts and anonymous TripAdvisor posts, an anonymised user is actually re-identified, so this might be a tricky privacy issue.
- Participants also pointed out that it is not always clear when the law allows for free access of data.

Challenges and barriers discussed at the Socio-economic table:

- The participants agreed that data may be reused no matter if it is raw data or data which has already been processed or analysed. Accordingly, data reuse may happen at all stages of the data value chain.
- It was highlighted that although most open data is currently provided by government bodies and research institutions, there are also reasons for companies to provide open data; product catalogues, for instance.

Event report

- There was strong support among the participants with respect to statements such as “data is one of the most important assets companies have” or “data is becoming a major source of competitive advantage”.
- Participants agreed that the need for human involvement in the processing and analysis of data increases its value considerably.
- The participants agreed that culture affects both data sharing within organisations as well as data sharing across organisational boundaries. Some people seem to have a stronger “feeling of exclusive ownership of data” than others.
- In small and medium-sized companies, lack of expertise leads to a very cautious approach to data sharing and reuse. They do not have the staff and resources to deal with data as large companies do. This means that many actors are reluctant to data sharing and in consequence there is a lack of data that may be reused.
- In general, companies, government bodies and research institutions but also citizens were considered to be not fully aware of the risks and benefits of data reuse.
- It was further pointed out that “companies increasingly give data to research institutions” to allow them to collect experience and do research with real-world data. However, it must be acknowledged that trust will remain an essential prerequisite for data sharing and reuse, at least for now.

Save the date!

Save the date for our next event on the 19th of October in Lisbon just before ICT2015, where a first heuristic data economy model will be presented and discussed.

Event report



Please visit our website more information: (eudeco-project.eu)