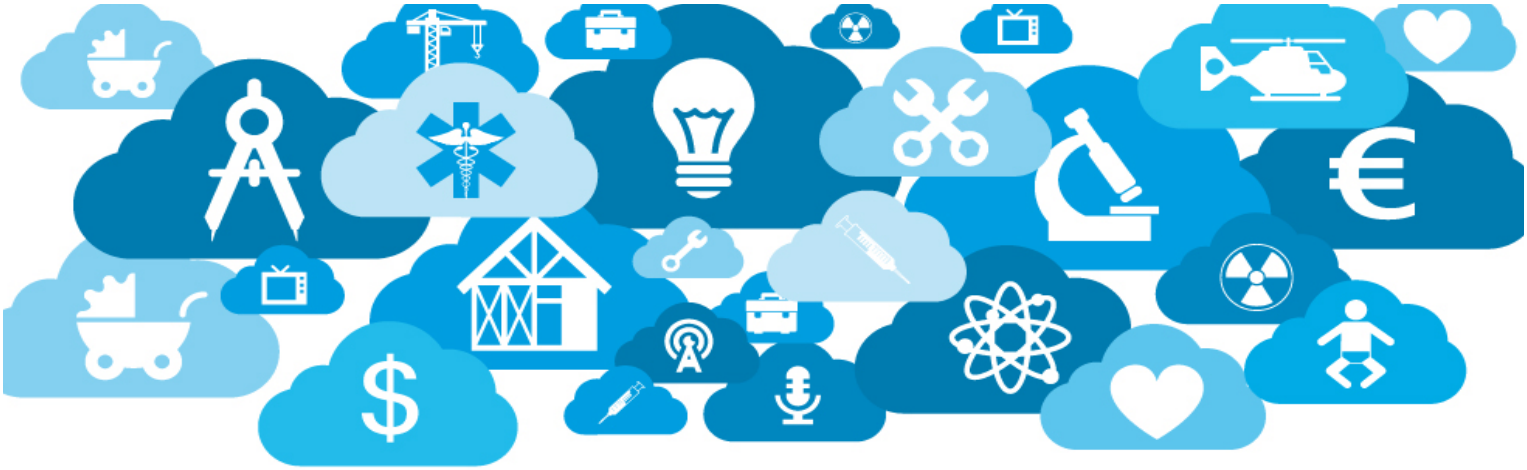


EuDEco Forum for Data Reuse

19th of October 2015, Lisbon

Event report

The forum for data reuse was a highly interactive half-day event held in Lisbon on the 19th of October 2015.

The forum gathered 26 participants of the EuDEco User Expert Group (UEG) and the Advisory Board (AB) to discuss the first version of the heuristic data economy model developed by the project.

The programme was divided into two sessions. The first session comprised:

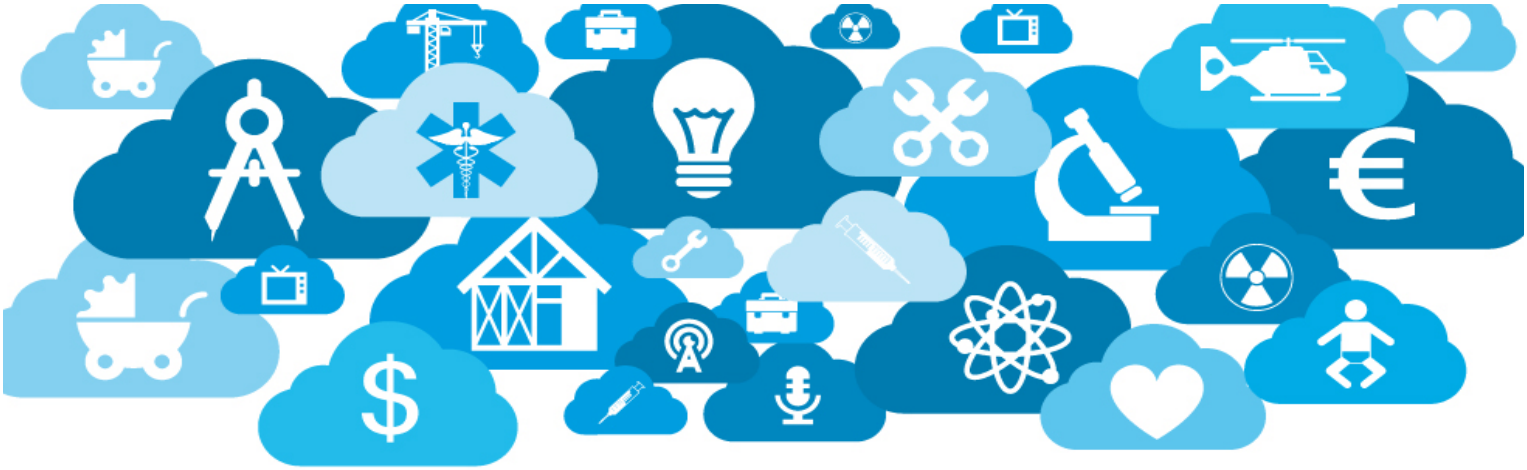
- A welcoming speech and an introduction of the agenda;
- A speed dating session for all participants to get to know each other;
- An introduction video of EuDEco;
- A presentation of the heuristic data economy model;
- Presentations of the legal, socio-economic and technological dimensions of the data economy model.

The second session consisted of three parallel discussions about legal, socio-economic and technological aspects of the data economy model. The session ended with small wrap-ups of the table rapporteurs. The discussions focused on the completeness and accuracy of the EuDEco model and the main challenges faced by European stakeholders.

The event was useful for getting a significant amount of feedback on topics related to the data economy in Europe and allowed clarifying numerous legal, technological and socio-economic questions.

For more information on the forum, please visit: http://data-reuse.eu/ueg-meetings/

Funded by the Horizon 2020 Framework Programme of the European Union



Summary of discussions on the legal dimension

Firstly, the discussion focused on data protection with special regard being given to the forthcoming GDPR, (General Data Protection Regulation). The participants at the table shared the opinion that there is an undeniable general lack of awareness regarding the content of data protection requirements. However once the regulation is adopted, the two-year transition period will provide enough time for companies to adapt to the new legislation.

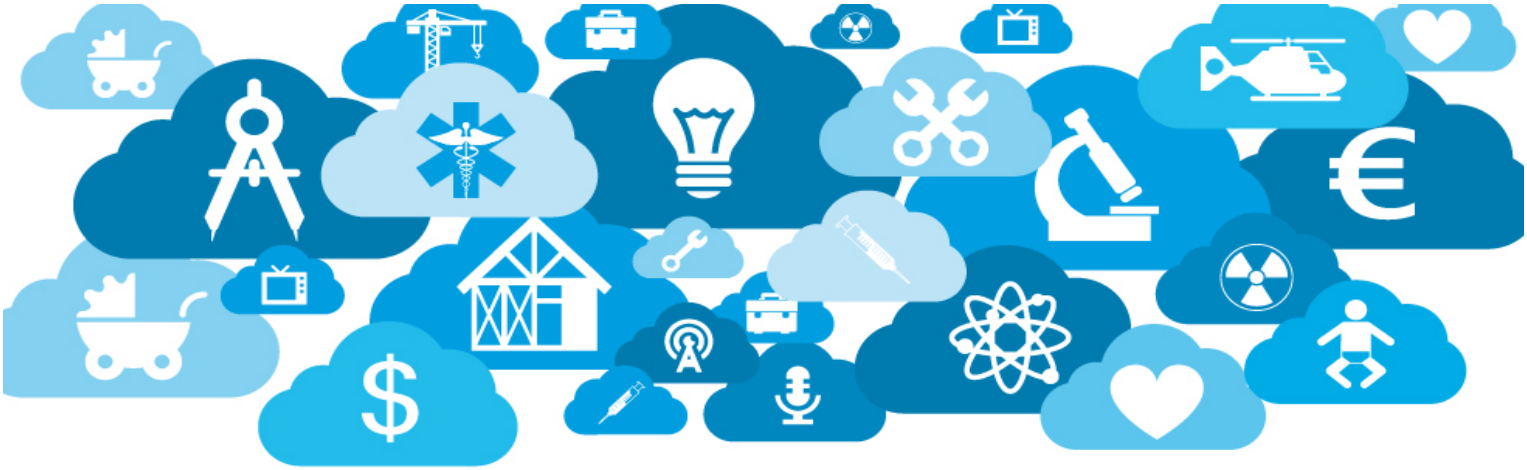
The second topic of interest was the impact of Intellectual Property law on the reuse of data, with the participants having a hard time reaching consensus on a legal definition of “data reuse”. In order to assess possible impacts of the intellectual property (IP) requirements, it is significant that the concept of data reuse is refined.

The third topic discussed was the issue of data localization, this is, the legal requirement enacted or proposed in some countries, such as Russia, which imposes an obligation to locate the data of the citizens within the borders of their countries. The discussion asserted that this could become a barrier to the digital economy, if it was too costly for companies to maintain servers in each and every country. As a consequence, they would only operate in the countries that have no or only limited restrictions on data localization and in the countries that could be a market by themselves.

The next topic was cyber security and the foreseeable implications of the proposed Networks and Information Security (NIS) Directive as well as its relation to the proposed GDPR. The focus of the conversation was on the consequences of a data breach for data controllers and processors. Namely, both proposals contain amendments to the current system of data breach notifications and impose fines that companies would eventually have to face.

The last point that the table addressed was the impact of the contract law on data reuse, where the group quickly reached consensus by supporting the idea of the case-by-case analysis. Everyone agreed that there is a misbalance in the freedom of negotiation between the parties (e.g., when a big corporation negotiates with a start-up or a small or medium-sized enterprise (SME). The law that applies to these contracts should be assessed on a case-by-case basis, depending on the applicable legislation the parties choose to adhere to, but always bearing in mind the jurisdictional rules.

Funded by the Horizon 2020
Framework Programme of the
European Union



Summary of discussions on the socio-economic dimension

The discussions at the socio-economic table focussed on some of the propositions comprising the heuristic model of the data economy.

The deployment of a self-sustaining data economy requires creating a data-reuse culture

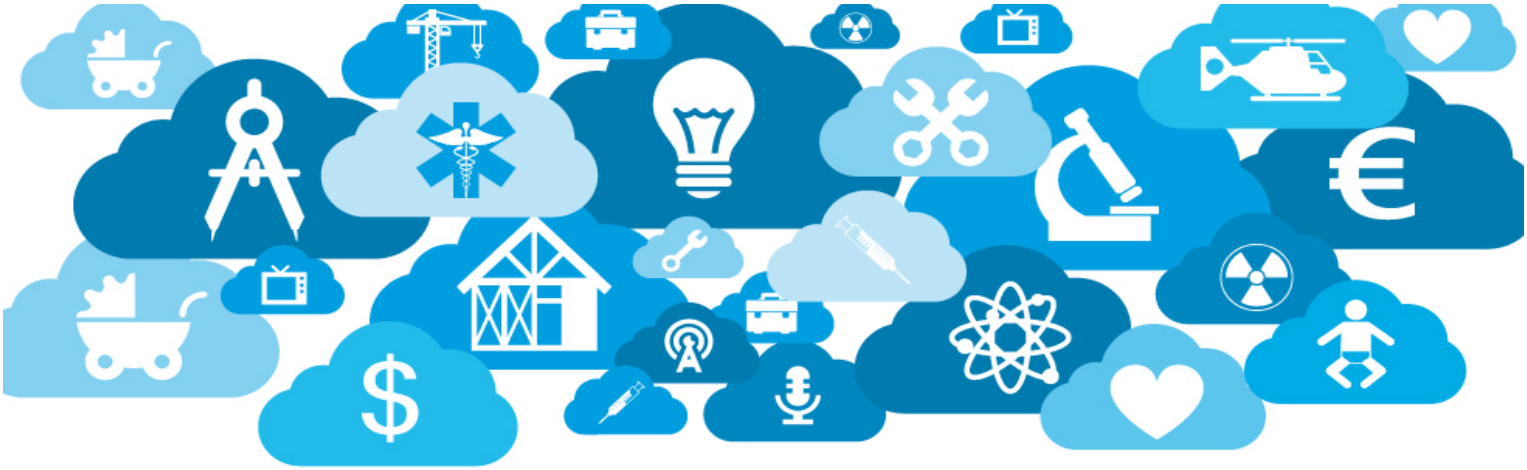
First, this question needs to be considered at different levels: the necessary *culture* may be very different on the data owner/producer's side and on the data user's side. Therefore, this topic requires further analysis.

The UEG members generally agreed that cultural, or non-technical, barriers, are among the most important obstacles to widespread data reuse. However, the concept of culture may be too wide and ill defined to satisfactorily describe the situation. Instead, the major issue might be rather a lack of awareness, in the sense that most actors still need to understand the potential of data reuse, in terms of uses and in terms of use and business cases. This lack of awareness is especially salient in SMEs. In addition to the lack of awareness, the importance of *social barriers*, such as lack of trust, must be acknowledged.

The suggestion that cultural change should be induced by the public sector taking the lead in opening public data was met with scepticism. Even though, on the one side, there might be the need for a top-down approach led by the public sector, in particular in terms of awareness-rising, the key lies in innovative practices, from any type of actor, demonstrating the value of data reuse in real conditions and in context-specific environments. Such a flagship experience is the open data programme around the London Olympic Games. On the other side, not all data should be shared or reused, which leads to the focus on sector-specific use cases rather than general approaches.

To support this bottom-up vision, the key question should be to identify and highlight the various interests that different types of actors can have in data reuse. Those may include financial benefits, but also non-financial advantages such as data quality improvement.

Funded by the Horizon 2020
Framework Programme of the
European Union



Monopolies regarding data ownership and analysis capabilities can hamper a well-functioning data market

This proposition obtained wide consensus, and is true also for the public sector, where public authorities might keep their data as a tool of control. Solutions should couple state intervention, in particular based on anti-trust and competition regulations, but also *softer* incentives. One key point is the development of alternative solutions, such as OpenStreetMap in the field of geographic information systems.

Data interpretation requires analytical skills of traditional experts in the field, in addition to data experts

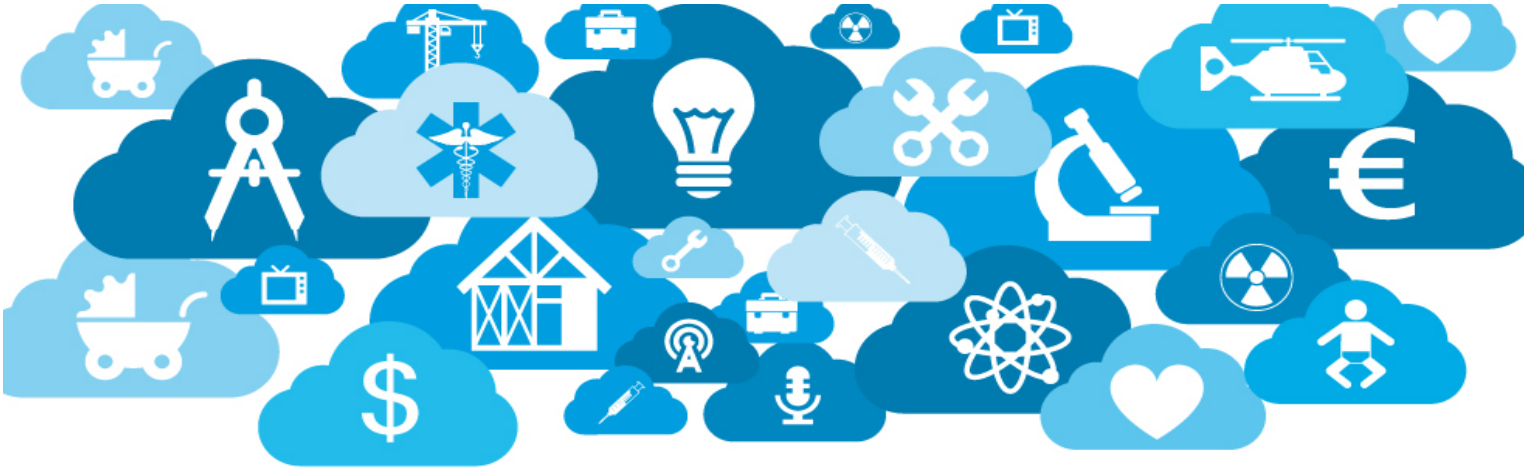
Again, the question of skills needs to be considered at different levels. On the users' side, a well-functioning data economy should not require any specific skill, since data analytics could be provided as a service by specialized companies.

From a more general point of view, the availability of the necessary data skills and how they are created are the key questions. It seems that a number of universities and business schools have already adapted their curricula and provide data-related courses.

The bottleneck in some (smaller) countries may not be the number of such professionals though, but the difficulty for interested organisations to identify them. However, the exact situation throughout Europe remains to be further studied and linked to the technological perspective. The two questions of vocational education and lifelong training have to be considered.

Another related topic is that the data value chain does imply a multidisciplinary process. In some sectors, this requires a dramatic change of mind-set, as traditional experts may not be used to working with people from different backgrounds.

Funded by the Horizon 2020
Framework Programme of the
European Union



The lack of broad availability of reusable data in many areas reduces the range of feasible business models

This question can be split into two issues:

- the availability of reusable data, and
- its discoverability.

On the first one, no clear consensus could be reached, and further studies at sectorial levels are needed, also in relation to the technological perspective.

However, it is commonly acknowledged that the big data *3Vs* (volume, velocity, and variety) do make discovering the right data a difficult task, which increases the costs of data search. Generally, information on the data available is lacking. So far, the development and availability of tools for searching/mining data has not kept pace with the availability of data.

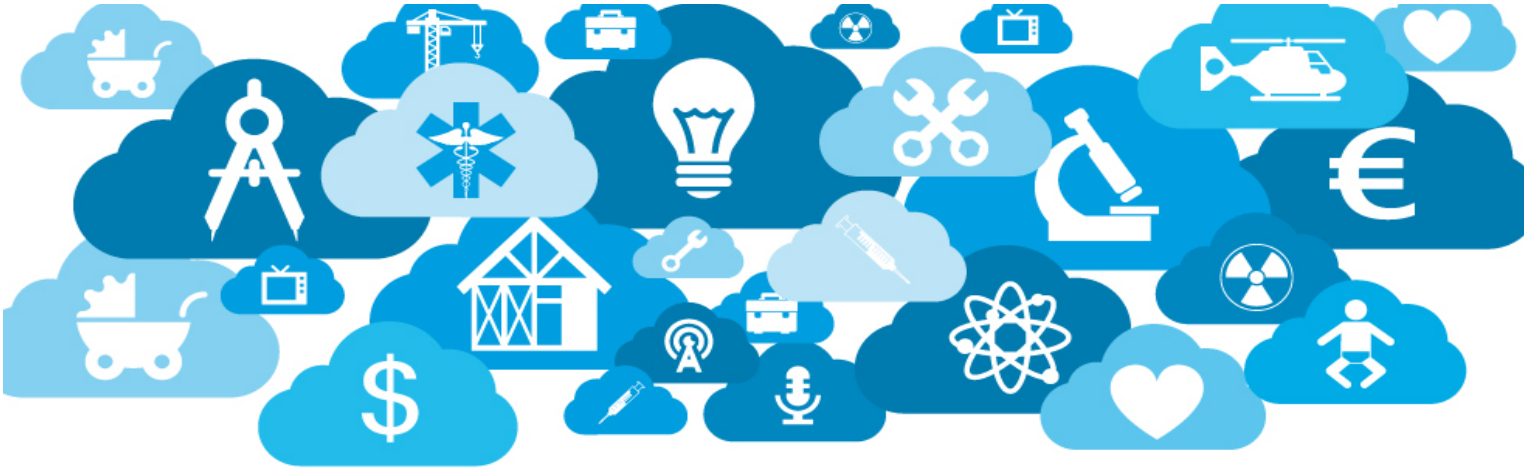
Difficulties with respect to determining the value of data make both potential data providers and data users cautious

A crucial preliminary question for determining the value of data is the one of data ownership. In many cases, who is the rightful owner of data is not clear yet, which logically prevents stakeholders to claim the value that can be extracted from it.

It is agreed that finding ways of evaluating the value of data is key, both for private and public actors, as it can justify the costs for producing/finding/aggregating the data or, for public institutions, for providing linked open data.

Moreover, the value of data varies significantly with the different uses it can have, and potential uses are by nature hard to determine once and for all. Therefore, it may not be relevant to give an intrinsic value to data, but the question is rather to evaluate the added value of its potential uses. This added value may or may not be restricted to financial value, and is given by the specific business models of the data uses. Therefore, the key question is not to evaluate the value of data *ex ante*, but rather to focus on how value can be extracted from a given dataset.

Funded by the Horizon 2020 Framework Programme of the European Union



An interesting example is the one of banks collecting data on the quality of the maintenance carried out by their clients, and adapting the interest rate of the loans they grant them.

Summary of discussions on the technological dimension

The first topic discussed was the discrepancy of broadband connection across Europe. As much as we have high-speed broadband connection in Denmark, for example, there is still a big difference between countries and within countries, particularly between cities and rural areas. While in several countries companies need to check for availability, in Finland, regulations guarantee the availability and broadband transmission speed.

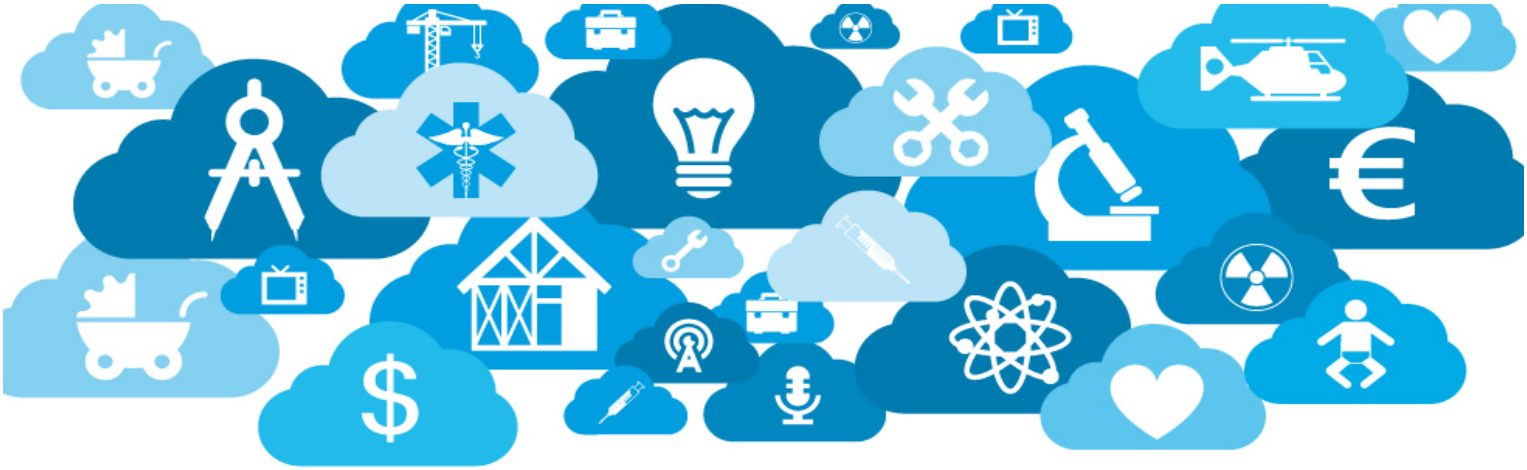
The second topic discussed was the application of data in business. In terms of the usage and sharing of data, the members have seen a high chance that data may soon follow the same model as other assets such as oil, gas or other goods of value. Over the years, more and more people may see value in data, especially if data and its processing meet high quality standards.

For the evolution of the European data economy, the sharing and aggregation of data is crucial. Companies may make use of shared (public and private) data to innovate and to make business decisions. However, it was also noticed that citizens and companies are tending towards getting more and more careful in sharing their private data.

The discussion table round agreed that there is a trade-off between usability and privacy concerns and that both citizens and companies are often willing to sacrifice their privacy to a certain extent in exchange for specific benefits.

The next topic discussed was related to personal data. The issue of privacy raised a lively discussion. Concerns over privacy and the US National Security Agency (NSA) surveillance have grown and people are being more careful about data sharing, and about the type of information they share, however personal data sharing is still not of great concern for them. Policy makers have reinforced a rule that youth should be aware of personal data privacy stressing how important data is and what they share, however from a personal perspective of

Funded by the Horizon 2020
Framework Programme of the
European Union



the actors at the table, youth do not consider that a major concern, and usually share sensitive data in a careless way.

As regards personal data and privacy best practices, the majority of the table expressed the impression that big players as Facebook, Twitter, YouTube, guide the market and that there was no specific rules, guidelines or boundaries applicable.

The majority of the table agreed that even if it had been possible to hide the IP of a machine, it would still have been easily possible to actively trace people and the table discussed a paper mentioning that only four datasets are needed to guarantee 80% of identity of a website visitor. However one question remained unanswered: How can the web-users secure their privacy? The answer was suggested by one of the guests: an identity management interface and/or privacy centres could be the answer. In this way the user could select the datasets or personal information that he would like to share, and the kind of information that it doesn't give allowance.

An additional topic discussed was based on the feedback from a EU project that addressed challenges faced in the context of geographical data. After a quick summary over the EU NUTS (Nomenclature of territorial units for statistics) classification, the debate continued with the critique that European geographical database have only three parameters in common:

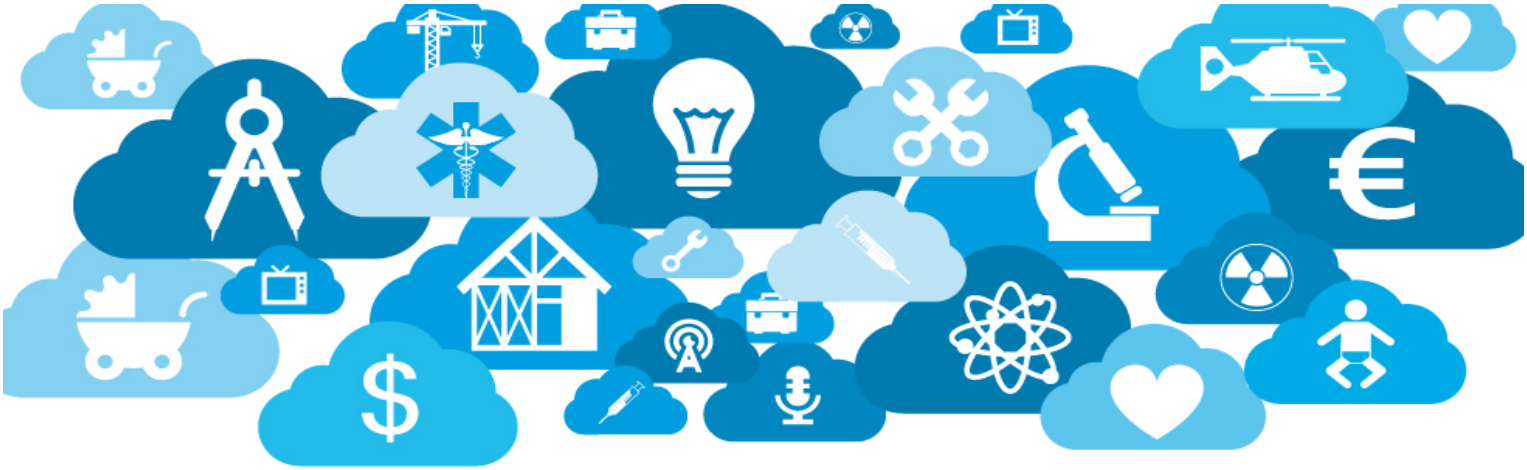
- The name of national parks,
- The date of publication of the law and
- The bulletin.

The majority of the group agreed that these barriers are not only technical but also have socio-economic and political aspects.

The fifth topic discussed was raised by the discussion group. The group stated that it does not know any rule or guideline regarding the quality of data. A key question was identified: How can the quality of data be measured and verified?

Data is coming from different sources, which raises more questions than answers: Where did data come from? For which purpose was it collected? Can the source of a dataset be validated?

Funded by the Horizon 2020
Framework Programme of the
European Union



According to members of the group, a CEO from a major bank in Europe reported that 90% of the activities of their technical team consist of verifying identity, if datasets are legal, if they can be tracked to the source and if they are real.

The discussion that followed this question focused on the possibility of making data marked, on certifications that would be needed to handle this data and on social media.

When data is *certified*, there may be more trust in it. When something is verified by, for instance, a stamp or a seal, you gain trust. It has been discussed that data – even when provided free of charge – is often considered to be useless for some sectors without an assurance of its quality. Could a data quality certification be similar to a product certification and how could it be realized?

The last point addressed was the engagement of companies in the big data industry to big data, and how the governments could provide incentives in this area. Tax benefits and standardization were named as possibilities.

Stay tuned for our next event in spring 2016!

Please visit our website for more information: data-reuse.eu

eudeco
Modelling the European data economy

Funded by the Horizon 2020 Framework Programme of the European Union