

DG CONNECT draft report: 26 June 2017 workshop on access to privately-held data for public bodies

1. Introduction

1.1. Context

In January 2017 the Commission adopted a Communication on "Building a European Data Economy" in which several aspects of access to and transfer of data are discussed. The main objective of the data economy initiative is to facilitate data sharing in a way that will enhance the overall European data economy, foster European innovation and ensure that Europe is able to compete in the global digital economy.

The Communication on Building a European Data Economy primarily aims at making it easier for European businesses and consumers to take part in the data economy, and facilitating access to data for purposes of innovation and business development. However, public sector bodies are also stakeholders in the data economy.

Access to data for public interest and research purposes is mentioned in the Communication as one of the possible options for addressing the issue of access to data, and the concept of what has been dubbed 'reverse PSI' is encompassed here. The concept of 'reverse PSI' refers to the Public Sector Information (PSI) Directive, which lays down a right to re-use all public documents (data) held by public sector bodies of the Member States. Reversing the concept of the PSI Directive would entail access for public sector bodies to re-use privately-held data.

Access to privately-held data for public interest purposes is also mentioned in the Digital Single Market mid-term review, which was published in May 2017 to take stock of the progress so far, but also announce some new actions in the Digital Single Market strategy.

In the Communication on the mid-term review, the Commission announced it would launch an initiative on accessibility and re-use of public and publicly funded data, as well as further exploring the issue of privately held data of public interest, by spring 2018. The main vehicle for this initiative will be the review of the Public Sector Information Directive, which is due by July 2018.

Although the issue of access to privately-held data for public interest purposes does not have an immediate connection with the PSI Directive (the PSI Directive is meant to open up data sets for the indiscriminate reuse by third parties for any purpose, and the 'reverse PSI' concept would not have such an effect), the purpose of fostering innovation and services through access to data is a common denominator.

1.2. Main issues discussed

This workshop was organised to discuss the possibilities of access for public sector bodies to privately-held data for public interest purposes and what can feasibly be achieved in this area of data access.

The Commission wanted to hear from actors from public sector actors about their need to access privately-held data for public interest purposes. It also wanted to hear the positions of the private actors typically holding the data in question, including what types of data they would be willing to share and for what purposes.

The following main questions were put to the participants:

- To public sector bodies: what kind of data do you have the greatest need for, and why?
- To private data holders: what data can data holders share, for what purposes and under which circumstances?
- Which measures will need to be taken in order to facilitate this kind of access from private to public?

Participants were also asked to reflect on whether they would favour any of the following measures, or whether other measures could be considered:

- A legal provision to grant access for certain specified public sector bodies to raw data sets held by companies.
- The introduction of guidelines or a cooperation framework for access for public bodies to privately-held of public interest.
- The introduction of standard procedures/standard contracts for public bodies' requests for access to privately-held data in situations of urgency.
- The facilitation of access through data "mediators" or through technical solutions to allow the public sector body to analyse the data in question without having to physically transfer any data to their own servers.

2. The four workshop sessions

2.1. Introductory horizontal session: how the public sector can do great things with data

The workshop started with three horizontally themed presentations, in order for the participants to get an idea of the broader issues.

Francois Levin and Judith Herzog from the **French Digital Council**, which consists of 30 colleagues appointed by the president to help shape the digital policy of France, talked about the French developments in access to public interest data. The French Digital Act was passed in 2016, and includes among its objectives to extend open data policy in France. The definition of public sector information has been extended to public funding, including concessions, public service delegations. Furthermore, data access rights for statistical

reasons has been improved, and also access to data held by certain private entities when the data are understood as infrastructure, such as data on electricity and gas distribution network. The French colleagues believed that introducing a category of public interest data on EU level in order to enhance access to data for public good would be a good idea – one could envisage obligations e.g. to provide data to improve public policies, for scientific research purposes or for public information. However, public authorities first need to identify the data they need and make sure there are infrastructures in place to share data between different actors, and to process the data securely.

Nuria Oliver from **Vodafone/Data-Pop Alliance** gave a presentation on what has been done in the context of mobile data. Vodafone have projects ongoing in many different areas using their data for the public good, including for natural disasters, humanitarian crises, public health, urban studies, transport, energy and agriculture. She also mentioned several other examples of initiatives aiming to use privately-held data for public interest purposes. The Data-Pop Alliance carries out research projects for public good, as well as being involved in data literacy and engagement activities. Using their OPAL platform, actors who want to leverage data for public good can access privately-held data using pre-approved algorithms that comply with privacy and security legislation.

Oliver underlined the importance of public sector bodies not only being able to access data, but also their ability to make sense of the data they access. She also stressed that voluntary data sharing should be encouraged, but that there is no need to force access through legislation. She held that access to privately-held data should not necessarily be free of cost. Oliver suggested that for access to privately-held data to become possible, the following conditions ("five guiding principles") need to be met:

- 1) Legal compliance
- 2) Socially acceptable purpose
- 3) Commercial viability
- 4) Technical feasibility
- 5) Sustainability of arrangement over time

Isabelle Vandoorne from **DG MOVE** gave an overview of Commission actions in the data-intensive transport sector. The use of privately-held data is mandated in different instruments, including the ITS Directive, the delegated regulation on EU-wide multimodal travel information services, the delegated regulation on road safety data and the legislation on railway transport services for passengers and freight. When it comes to multimodal travel information services, dynamic data, such as real-time fare data, seems to be more difficult to access than static data (e.g. timetables, estimated travel times, refuelling, etc.). The goal is to start exchanging at least static data EU-wide via National Access Points by 2021. Whether the exchange of dynamic data will be mandated is left up to the member states to decide.

Upon questions from the participants, the issue of how to handle commercially sensitive data, and balancing commercial interests with public interest was discussed. The possibility to reveal trade secrets is often used by companies as an argument not to share data. It was suggested that a principle of mutual interest in sharing should be applied, but it was also held that the legislator should be cautious in defining general policy that may harm the commercial interest (including the interest in producing the data in the first place).

2.2. Thematic session I: data access for official statistics

Presentations were given from representatives of the **Danish, Spanish, French, Dutch and Slovenian national statistics agencies**. There is general agreement between the statistical offices that they need access to more privately-held data than what they have today. Official statistics are produced by the national statistical agencies. Good official statistics are absolutely necessary for the performance of public institutions and the EU, and as such they are considered as a public good. There are more and more requests for official statistics on new areas. Without access also to private data, official statistics will fall short and lose relevance.

The statistical agencies primarily need access to **raw micro-data** (i.e. data that have not been transformed or reprocessed) for quality and standards conformity. Examples of data of interest to the statistical agencies are mobile phone data, financial transactions data, scanner data, eCommerce data, social platforms data, transport data and smart energy data.

Certain Member States are working to enhance the (free) access of statistical offices to data, including in the UK, Netherlands and Denmark, as well as the recently adopted French Digital Act (mentioned also in introductory session by the French Digital Council).

EU-level legislation to further enable data access for statistical agencies is welcomed by most of the representatives from national statistical agencies, possibly in the context of the EU Data Economy Initiative or the DSM Strategy. The above-mentioned French Digital Bill, as well as the UN recommendation on access to proprietary data, could serve as inspiration. According to the statistical agencies, main issues that would need to be addressed in such legislation include provisions on the **structure of the data** to be accessed (i.e. whether data can be pre-processed or not), what kind of **cost structure** to apply (i.e. who pays for what) and how to deal with **proprietary, personal or otherwise confidential data**. There is also the question of what level of anonymisation or aggregation of data one should expect from the national statistical agencies, and whether further standards for information and data security need to be set.

"Data collaboratives", ensuring continuous dialogue between data providers and data users, are getting increasingly popular in statistics. Establishing **Public-Private Partnerships** on data sharing for statistical purposes could be a solution if softer measures are considered.

2.3. Thematic session II: Could data access save lives? Cases for better public health in the EU through privately-held data

This session provided many interesting use cases for the need to access different kinds of privately-held data. **The European Centre for Disease Prevention and Control** have accessed mobile data to define the location and possible spread of outbreaks of Legionnaire's disease, a serious lung infection caused by water-based bacteria with a considerable mortality rate. Similar cases to look at using the same data include individual case location (e.g. of airborne diseases from contaminated water, infected animals, vector borne diseases) and population density (e.g. to measure outbreak attack rate and keep mass gathering events at risk under surveillance).

The **Danish Ministry of Environment and Food** is increasingly using data as part of the services they provide, and food safety is one of the application areas for data analyses. They have successfully managed to introduce a system where owners of livestock report to the authorities prior to any export of animals, e.g. about what animals will be exported, when, and where from/to. This information helps the authorities plan for their follow-up. One type of data they have been less successful in accessing is the so-called "own-check data" from food retailers (self-verified data to ensure compliance with food safety regulations). These data are subject to on-site inspection during official food controls by the authorities. It would be useful for the authorities to access these data in order to prepare inspections and also save time during inspections. However, retailers have resisted such access, arguing data security issues and that the cost of making these data electronically available is too high.

As part of the project Crowdhealth, the **Health Research Institute La Fe** is trying to learn more about what determines patients' health by looking at data not generated in the health care context, such as data on dietary, physical activity and sleeping habits as well as sensor data and socio-demographic information. The goal of this project is to create more holistic health records, e.g. to help battle obesity. It has proven difficult to get consent from patients to access

As this session showed, the types of data needed for different purposes related to public health vary greatly. However, the challenges in accessing the data seem to some extent to be similar. The confidentiality and security of data is often mentioned as an argument not to share the data. The accessibility of the data is also an issue. Although real-time tracking is often more useful, providing data in this manner could prove very costly for the data holder.

Although data protection regulation and the need for consent from data subjects are not within the scope of issues to be addressed in the Commission's initiative on access to data, it should be mentioned that these issues were brought up by many participants.

2.4. Thematic Session III: Access to privately-held data in the context of smart cities

Both the **City of Amsterdam** and the **City of Hamburg** have broad experience with data partnerships with different private actors, e.g. through establishing Private-Public Partnerships. Amsterdam believes cooperation and finding common interests with partners from the private sector, always with a specific objective in sight, is the best way to make the most out of the data accessed. They gave the example of a project to reduce poverty by using statistical data to find out how to best allocate the means at their disposal, as well as a 'Moonshot' project to create the real-time city, e.g. using (close to) real-time traffic data from Google, TomTom and more traditional partners such as financial institutions, retail and IoT and more recently also telcos. Important prerequisites for their agreements with private actors on access to their data were for the city to adjust their way of working and thinking to that of the companies, as well as agreeing to very strong non-disclosure agreements. One dilemma they have often met is whether or not to pay for privately-held data. The City of Hamburg presented projects to establish an information system on subterranean cable routes and on mapping heat supply and demand in the city. Both projects were based on agreements with the private companies (including utilities, housing companies, energy companies, etc.). They had experienced companies that were unwilling to share data with them, e.g. for security reasons.

In response to the question whether they were free to commercialise the work they did with private companies, the cities answered that often they could not. On the contrary, companies were often left to sell the solution they had cooperated with the city on to other cities. A challenge lies in the need for cities to develop comprehensive or standardised approaches to data-driven policies and projects: cities would benefit from sharing data and solutions between themselves, but this interest needs to be balanced with the interest of companies to be able to commercialise the solutions they cooperate with cities to develop.

Conclusions

The workshop gave many useful examples in different areas of cases in which privately-held data can or should be accessed for public interest purposes. There seems to be advantage in taking a cross-sector approach to this issue, as there are many similarities in both the types of data needed and the main issues faced.

Several participants also raised the issue of data protection, and how to reconcile the need for data protection with wider access to data. Another issue frequently mentioned was that of proprietary and commercially sensitive/confidential information, and how to deal with this while opening up for broader rights to access privately-held data. A third, and rather fundamental issue is that of cost: who should bear the costs of facilitating access to and re-use of these data for public good? There will inevitably be a need for additional infrastructure, data protection and security measures, as well as better data readability and interoperability. More information on the exact cost involved, and a discussion on who should bear the different costs, is needed.

Different opinions emerged during the workshop on whether a principle of access or data sharing should be set in legislation or be upheld by voluntary arrangements or softer policy measures. The DSM Mid-Term Review Communication is cautious to present any preferred option in this regard. It is clear that the Commission wants to be stronger on PSI, but on reverse PSI there is less evidence. The Commission is therefore very open, and the issue needs further studying before any policy options can be presented. No preconceived ideas from EC on what should come out of the process.