For more information about this interview, please contact Mr. Maxim Chantillon (KU Leuven Public Governance Institute – maxim.chantillon@kuleuven.be)

The Unit

Role of Data Innovation and Policy Unit within DG Connect? Relations with other units and DGs within the European Commission – e.g. INSPIRE Unit, European Interoperability Framework Team?

This unit is in charge of making policies on data and also to follow part of H2020 actions related to Big Data. Not the only ones in DG Connect and EU Commission to do this but they were at the initiative of the Communication on the Data Economy. They concentrate on facilitating the re-use of PSI: idea is that data are resources that can be reused multiple times. The issue is that technical reasons for not being able to re-use it yet.

On technical side: ISA^2. There is not really implicated in standardisation processes, but they support standardisation activities and re-use results of ISA^2 activities. Ex: metadata, they are responsible for European Data Portal and implement DKAT that was the result of ISA^2 activities. In the coming months, they plan to start extension of DKAT with statistical and Geo-data => bridge between complicated modal of INSPIRE and simplified one of DKAT. Therefore, they work more as a user than as a contributor to standardisation.

They have also concluded PPPs (Private-Public Partnerships) under the H2020 vision: PPP on Big Data started a couple years ago and became operational with first projects => find ways of improving existing standards on Big Data (Cross domain Big Data) and development of standardised architecture for Big Data.

Linked Data solutions implemented in the Open data portal. Mature at the EU level but not so much on the field. Here there is a discrepancy between the real situation and their perception of it.

PSI

Why did the EU Commission feel, in the early 2000s, that there was a need to adopt a legally binding instrument, the Directive of 2003, to regulate the re-use of PSI?

Open data has always been linked to internal market. In member States, often seen as a tool for increasing transparency of government. From the Commission perspective, it is more a market driven initiative. In the USA, general rule that federal level data is open and can be re-used and whole business of PSI re-use was booming. There was a need for this in Europe as we were lacking behind for this type of economy. In Scandinavia, long standing principle of opening government information but not at all the case in other countries where no such PSI policies.

Moreover, the general vision in the early 2000s was that the most important and valuable sets of public data were only re-used based on exclusive contracts. Therefore, the market was closed so need to open up the market (make it competitive and abolish exclusive agreements) and harmonize the (licencing) conditions under which this re-use takes place in the internal market, especially in light of the development of the digital economy.

Do you think that, if it had not been legally binding, the Member States would have overlooked it? Is this also why did they decided to amend the Directive of 2003 and to compel the public sector to offer their PSI for re-use (only a possibility in 2003)?

They monitored the level of implementation of the PSI Directive, and after a couple of years, there were still many institutions that did not open up data. Moreover, when they did open

the data, they were setting very high prices so the objective of the Directive was not reached. For example, the UK was doing good but other countries were just doing the minimum and no proactive action. Therefore, they had to go a step further and to compel the administrations.

According to the PSI Directive, re-use of PSI should be done in full compliance with the privacy rules, and soon the GDPR. When we talk about this to administrations, especially the local level, they are quite confused and don't really know what to do. In your mind, how should this balance be reached? Anonymisation?
Personal data protection was one of the main obstacles to a wide re-use. There are different ways to deal with this, but it is certain that there is a need to accommodate with the GDPR and somehow make the administrations more aware that personal data is not always an obstacle to re-use of large databases, at least partial re-use.
The Data Innovation and Policy Unit cannot do anything specific about this, but they have started working with the EDPS (European Data Protection Supervisor), in order to see what kind of practices are emerging already. For example, there are specific committees, within the national data protection authorities at the National level, which help administrations to deal with this. For example in Spain, they have guidelines.
This shows a real need for cooperation between those who hold the data and the data protection authorities. Guidance published on this topic by WP29 (Working Party 29) but too much data protection oriented so not satisfactory for the Commission. In the UK, in their licence, personal data is simply excluded so reassurance given to the administrations.
In any case, the GDPR has precedence over the PSI Directive so there is a need to work based on best practices. They encourage public bodies to team up with data protection authorities and to define guidelines. The solution should be found at the practical level. However, there will be, in the context of the review of the Directive (July 2018), a study on the interplay between privacy and PSI.

In the context of re-use and compliance with privacy rules, do you believe that the public administration should check that the re-user is going to respect the privacy of the citizens' whose data is transferred? Or should this solely be the re-user's liability, who should be accountable of the compliance with privacy rules?
They hope that technology will be able to help them with anonymisation techniques. Some datasets are very useful from an economic perspective but personal layer on it. However, if these identifiers could be taken away by technology to anonymise the dataset, it would be great, especially because no real value on that personal data layer. Rather it is seen as contaminating the data.
There is no 100% certainty that a database can be made privacy proof, but it is possible to ensure a strong protection thanks to technological means. Even though no 100% security, there is a need to prevent most of the data abuses.

What is your experience with the implementation of the Directive (2013 review) in Belgium? What about the other Member States?
Belgium is not the best example as they were late in the transposition of the 2013 review of the PSI Directive. Half the countries missed the deadline of transposition so hard to assess the implementation so far. However, they observed a general trend regarding charging fees for re-use that are accepted and more and more studies about charging fees for data. These studies show that, in fact, in some cases, it creates costs for the administration that are almost as large as the fee they obtain. The Directive is one of the causes of this as it created the rule of charging a fee corresponding to the marginal costs. But even when charges of the administrations could be higher as they are not compelled by Directive to use the marginal

costs, the charges tend to go down.

The Directive is also quite light on the Licencing conditions and encourages use of standard licences, such as the creative commons licences. Here, they perceive Belgium as a good example, as they have heard about the proposal to use the creative commons licence. They seem to be unaware of the fact that this raises a lot of opposition from the administrations.

They added that they see that more and more data are opened. According to them, the whole area of work ahead is to make sure that the supply matches the data => that the data released is the data re-used. They believe that, at this stage, there are not enough matchmaking processes, where public sector discusses with private sector about what data they need and at what level of quality. Such debates take place in Italy and in France but this is the next big challenge. There is a need to make sure that investments made by the governments and administrations in Open Data pays off and that re-use is indeed done.

It's all a question of a funding model, as if the institutions were ensured that they will keep functioning, then they will of course open their data. In this regard, they haven't received any signals of regret from administrations that opened their data. Signal that they get is that still extremely hard to measure the real re-use of data, as very few processes of identification of who re-uses are in place. So mainly gets it through feedbacks and surveys but no real measurement tool.

Geo data and PSI

Geospatial data is one of the top priorities of the PSI Directive. Why is this the case?

In general re-users perceive the importance of such geo-data and a lot of apps built on it. The issue is more the way of using it, we have to step away from traditional use. It is cited in the top priorities of the PSI Directive and efforts to standardise the offer thanks to INSPIRE. But it is complex so enormous amount of time to set it but once set, it should allow seamless exchange and interoperability of data.

Very often users would rather have data that is not 100% correct but have it more rapidly. The point here is to get a good compromise between official, reliable, qualitative data and the need to feed the market with more regular updates, even if not 100% correct. Analogy can be done with official statistic data that is often released late, as there is a need to be 100% correct from the public sector point of view, but the data loses value for re-users as it comes too late. Another point of difficulty is that geo-data are complex instruments to re-use and it makes it hard for small companies who cannot invest in the re-use of such data. For example, there is an enormous amount of data generated thanks to Copernicus, but many re-users are not able to re-use it, as it is too complex for SMEs etc. So maybe idea of "first degree" relationship between public sector and big companies who have the means to re-use these complex and enormous datasets, which could then offer "second degree" services to smaller companies and SMEs.

For autonomous driving and robotisation, geo-spatial data is an important enabler, but not the only relevant one. Also weather data and commercial registries matter. Some countries came up with initiatives, in a more coordinated matter, to open geo-data. For example Denmark with postal codes: involved government, ministries and local level. But the interviewees don't feel that any country pushed for geo-data, it does not especially stand out. Finally they mentioned thoughts revolving around the idea that geo-data is a public utility and should be made as open as possible. However, no real understanding about where this could lead. Maybe idea to merge geodata portal with the open data portal. They are built on different legal obligations (geo portal is an obligation under INSPIRE, but no legal obligation to create a general open data portal). But idea to allow specific and/or generic streams of information between the portals. Not total merger as they have their own specific utility, but

idea to create interoperability, maybe through DKAT to simplify the complexity of INSPIRE for general data.