68.

-----Original Message-----
From: [redacted]
Sent: Monday, January 23, 2017 1:39 PM
To: JUST ARTICLE29WP SEC; presidenceg29@cnil.fr
Subject: comments on data portability

Hullo

I am a professional working with large scale computer systems for many decades.

I have scanned your documents 'Guidelines on the right to "data portability"', (http://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf)
and 'WP242 ANNEX – Frequently asked questions', (http://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_annex_en_40854.pdf).

There are two areas that these documents do not seem to address well:
1- the semantics of the data transferred
2- security issues

Taking these in turn.

1/ Any data provided is subject to interpretation that is based on implicit and explicit semantics in the source system and processes, including, but not limited to:
- object models
- precision, accuracy, timeliness, completeness, consistency
- units

All of these issues would normally be significant in integrating disparate systems as the original design assumptions are likely to be lost and violated in the actual data in the system. To give some simple examples:
- Object model: Name. A persons name is commonly used for different purposes, including addressing the individual and identifying the individual, and they may vary over time and in different cultures. Some enterprises mix these purposes. So, there is no data that is both unambiguous and universally available. For data such as Playlists - a cited example - what would be relevant? Surely not a description of the played media clips, as that is ambiguous. But names (in the sense defined by Roger Needham as the shareable reference) are often not meaningful in other organisations.
- precision, accuracy, timeliness, completeness, consistency must be understood so that data elements can be meaningfully combined. A common example here would be smart meter data, where representations vary over time and model of device. If they are normalised, then a/ they have been processed and so fall out of the scope of the advice, and b/ they can only be delivered at the lowest common denominator. There are further complications where meters actually measure different things (eg. total energy used, vs rate of use of energy over a period of time).
- units: numbers alone are meaningless, the dimensions and units must also be defined.

There needs to be clarity for data processors and data subjects as to what the semantic information that is required is, and the limitations that such requirements entail.

2/ The data portability requirement presents both an interesting new attack surface and a potential burden on individuals and data processors to provide mutual satisfaction of identity. It also presents

a new information leakage channel and it is not clear how responsibility for an information leak could be easily established. Additionally, work on smart metering systems shows some specific confidentiality threats from smart meters as fine grained data can show when someone was present at a location, and in some cases, who that person was. Also, it should not be assumed that the user of some data input device is necessarily the person that used it to input the data. Some of the device/ownership related issues seem to me to be particularly burdensome on data providers and to provide unexpected and unnecessary surveillance capabilities for individuals.

regards

[redacted]