A focus on 'information' instead of on 'data'

Emile Douilhet, University of Bournemouth

European Data Protection relies on a distinction between "personal data" and other data, "personal data" being the focus of the Data Protection framework. This paper will attempt to show that this distinction is becoming more and more obsolete as technological means advance, and that a focus on "information", instead of on "data", is required.

Data Protection in the EU is dependent on the concept of "personal data", which is defined in the Data Protection Directive as "information relating to an identified or identifiable person". The "identification" test is applied upon collection of the data by a data controller. However, this data is then transformed, aggregated and mined, and produces information. The content - and identifiability – of this information might be widely different from the original data's. In particular, pieces of information representing little original chance of identification (such as anonymised data or data only tangentially connected to individuals) can be aggregated with other data to produce very powerful information, an example being the practice of data profiling.

As such the "information" and the "data" are not entirely connected, and this connection is getting weaker as technological means of data processing and analysis develop.

"Information" is created and developed at every stage of data processing, and the European all-or-nothing test of "identifiability" could not be applied every time a piece of data is processed or mined or combined with another. Because of information's fluid nature, it is necessary to use a standard which follows and develops alongside the information.

In this paper, we propose an answer to this problem. Inspiration from this answer comes in part from Helen Nissenbaum's model of "contextual integrity". That model focuses on "appropriate information flows", appropriateness being dependant on the information-sharing context's socio-legal norms. There are forces constricting the flow and creation of information, and these forces are what protects (or fails to protect) privacy.

A complete picture of these forces, not restricted to only socio-legal forces, allows us to understand how and what information is created, and acting on these forces to bind information could answer the difficulties of the evolving landscape of data processing.

For this purpose, we will adapt Lawrence Lessig's four forces to the field of privacy: legal, social, market and architecture (in this case, architecture being the technological infrastructure). Many proposals have been presented to correct the existing imbalances in privacy. New laws, new social movements, new economic systems (such as data monetization), and new technologies and technological options (the TOR network, search engines like duckduckgo.com, etc.) all have a part to play. The goal of this taxonomy is to evaluate these different proposals on a common frame of comparison, and see their impact on information creation and dissemination, in order to find the most effective tools to protect individual privacy.