Government Information Quarterly 33 (2016) 338–345

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Data protection legislation: A very hungry caterpillar
The case of mapping data in the European Union

Bastiaan van Loenen [a,*], Stefan Kulk [b], Hendrik Ploeger [a,c]

[a] Faculty of Architecture and The Built Environment, Knowledge Centre Open Data, Delft University of Technology, The Netherlands
[b] Centre for Intellectual Property Law, University Utrecht, The Netherlands
[c] Faculty of Law, VU University Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 14 March 2014
Received in revised form 7 April 2016
Accepted 9 April 2016
Available online 30 April 2016

Keywords:
Data protection
Privacy
Open data
Mapping data
European Union
PII2.0

ABSTRACT

The European Union's policy on open data aims at generating value through re-use of public sector information, such as mapping data. Open data policies should be applied in full compliance with the principles relating to the protection of personal data of the EU Data Protection Directive. Increased computer power, advancing data mining techniques and the increasing amount of publicly available big data extend the reach of the EU Data Protection Directive to much more data than currently assumed and acted upon. Especially mapping data are a key factor to identify individual data subjects and consequently subject to the EU Data Protection Directive and the recently approved EU General Data Protection Regulation. This could in effect obstruct the implementation of open data policies in the EU. The very hungry data protection legislation results in a need to rethink either the concept of personal data or the conditions for use of mapping data that are considered personal data.

1. Introduction

It has been estimated that every day 2.5 Exabytes (2.5×10^{18} bytes) of data, an equivalent to 200 million DVDs of 5 Gb, are created (IBM, 2013) and added to the already enormous amount of 'big data' mostly available through the internet. Data may vary from the holiday snapshots of Mr. and Mrs. Jones from London and the daily tweets of their sixteen-year-old daughter Elsie to the commercial datasets of Google and Experian, or national datasets collected by the public sector, such as census data, topographical maps and elevation data.

Developments in information technology have significantly improved our ability to process data. Also the data itself (level of detail, currency, and interoperability) has improved. In addition, open data initiatives resulted in a greater availability of (public) data that can be freely re-used by anyone for any purpose. It has been claimed that the economic value of billions of Euros will be created by the reuse of open government mapping data alone (Dekkers, Polman, te Velde, & de Vries, 2006; Pira International Ltd., University of East Anglia, and KnowledgeView Ltd., 2000; Vickery, 2011). Therefore, mapping data, such as topographical maps and the underlying earth observation data, are top-listed by the European Commission and the G8 for release as open government data due to the high demand from re-users (Cabinet Office, 2013; European Commission, 2014).

However, the open government data policies may conflict with the individual's right to information privacy as protected by the EU Data Protection Directive (European Parliament and Council, 1995) that sets rules to the processing of personal data in the European Union. At first glance, mapping data may not necessarily refer to individuals. However, the data may become personal data by combining it with other data or when de-anonymized. Mapping data have a special role to play in this linking of anonymous data to a person. Linking anonymous data to a location on a map may turn such data, and the mapping data, into personal data. This is important to note because the use of personal data should be in full compliance with the principles relating to the protection of privacy. The EU Data Protection Directive dictates that the data cannot be freely re-used by anyone for any purpose, but should be processed for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes.

In this article we argue that the increased computer power, advancing data mining techniques and the increasing amount of available open data are transferring previously non-personal mapping data into personal data. We argue that the EU Data Protection Directive has turned into a 'very hungry caterpillar', which could in effect obstruct the implementation of open government data policies for mapping data in the EU.

The structure of this article is as follows. We first briefly discuss open data (Section 2) as well as data protection in the European Union (Section 3). Then, we define mapping data (Section 4) and discuss the

* Corresponding author.
E-mail addresses: b.vanloenen@tudelft.nl (B. van Loenen), s.kulk@uu.nl (S. Kulk), h.d.ploeger@tudelft.nl (H. Ploeger).

http://dx.doi.org/10.1016/j.giq.2016.04.002
0740-624X/

key question "Is mapping data personal data?" (Section 5), and in Section 6 we discuss the implications of mapping data being personal data. After an intermediate conclusion (Section 7), we continue with five possible directions for open data release while safeguarding data protection. In Section 8 we discuss the implications of the recently approved EU General Data Protection Regulation for our research findings. Section 9 presents our conclusion.

2. Open data in the European Union

Open data are data that are available without any restrictions to its use, are machine-readable, and adhere to open standards (Kulk & Van Loenen, 2012). The European Commission strongly advocates open data in its Digital Agenda for Europe program (European Commission, 2010; European Commission, 2011). The Commission's hopes are that the greater availability of interoperable public data catalyses the secondary use of such data, which leads to growth of information industries and better government transparency.

The total potential value of re-use of open public sector information in Europe is estimated to vary from €27 billion (Dekkers et al., 2006) to €68 billion (Pira International Ltd., University of East Anglia, and KnowledgeView Ltd., 2000). The economic value of commercial exploitation of public mapping data has been assessed to account for over 50% of the total estimated value (Dekkers et al., 2006; Pira International Ltd., University of East Anglia, and KnowledgeView Ltd., 2000).

The EU Directive 2003/98/EC on the re-use of public sector information aims at stimulating re-use by third parties (European Parliament and Council, 2003). The directive is the key instrument to arrive at the Commission's objective of enabling the availability of public sector data to third parties at low prices and with non-restrictive conditions (Janssen, 2011). The 2013 amendment (Directive 2013/37/EU) extended the scope of the directive and took the "open data, unless" standpoint (European Parliament and Council, 2013). Public organizations are stimulated to provide their data for re-use under open data policies: this means no charges and no restrictions in the use. However, this policy should be applied in full compliance with the principles relating to the protection of personal data (Recital 11 Directive 2013/37/EU).

3. Data protection in the European Union: the EU Data Protection Directive

The (re)use of open data is not without legal limitations. Article 8 of the Charter of Fundamental Rights of the European Union guarantees a citizen the right "to the protection of personal data concerning him or her". The automated processing of personal data is also covered by the Council of Europe's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. This fundamental right is further elaborated by the Data Protection Directive (European Parliament and Council, 1995).

3.1. The concept of data controller and personal data

The data 'controller' plays a key role in the EU Data Protection Directive. The data controller is the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data (Article 2(d) of the EU Data Protection Directive).

The directive defines personal data as "information relating to an identified or identifiable natural person". An identifiable person is "one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity" (Article 2(a) of the EU Data Protection Directive).

Typical examples of data that relate to a person are names, e-mail, Internet protocol, or portal addresses, postal addresses and telephone numbers (see Article 29 Working Party, 2000; Article 29 Working Party, 2007; cf. Watts, Brunger, & Shires, 2011; Robinson, Graux, Botterman, & Valeri, 2009). Personal data are, however, more than just names and addresses. The Article 29 Working Party, which is the group of European Data Protection Agencies with advisory status, also emphasizes that the purpose or result of how that data is used should be taken into account in order to determine whether data is personal data: "data relates to an individual if it refers to the identity, characteristics or behaviour of an individual or if such information is used to determine or influence the way in which that person is treated or evaluated" (Article 29 Working Party, 2005). Moreover, the Working Group argues that "data can be considered to 'relate' to an individual because their use is likely to have an impact on a certain person's right and interests, taking into account all the circumstances surrounding the precise case. It should be noted that it is not necessary that the potential result be a major impact. It is sufficient if the individual may be treated differently from other persons as a result of the processing of such data" (Article 29 Working Party, 2007, p. 11).

On some occasions data concerning objects may be personal data. For instance, the value of a house is at first glance, 'just' information about an object, i.e. information to which the data protection rules do not apply. However, "the house is the asset of an owner, which will hence be used to determine the extent of this person's obligation to pay taxes, for instance. In this context it will be indisputable that such information should be considered as personal data" (Article 29 Working Party, 2007, p. 9; European Commission, 2012a, p. 16).

The assessment whether data should be considered personal data also depends on how easy it is to link data to a person. Or as the Data Protection Directive reads: "account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person" (EU Data Protection Directive, Recital 26). When identification of the individual requires a disproportionate effort it should not be considered personal data (EU Data Protection Directive, Recital 40). This may be the case when the identification of individuals would cost many days of computing time (Dutch Government, 1999b). However, in this instance, the on-going developments in computer technology pose a serious problem: data that are today considered not to be personal data may very well become personal data tomorrow. One example may be the publication on the Internet of a picture including anonymous individuals. Ten years ago, it was almost impossible to uncover the identity of individuals in a picture. Today, facial recognition software (not only commercially used but also made available to the general public by e.g. Picasa and iPhoto) allows identifying these persons with a simple mouse click (GAO, 2015). Since it is very difficult to effectively remove data from the Internet once it has been put online (see Article 29 Working Party, 2013a; Gallo, 2012), one may argue that any data that in the future might be linked to individuals, should be considered and treated today as personal data (Kulk & Van Loenen, 2012; see also Article 29 Working Party, 2007).

Not only technological advances in software and hardware, also the increasing number of available open datasets increases the risk of identification. "A person might still be "identifiable" [if] information combined with other pieces of information (whether the latter is retained from the data controller or not) will allow the individual to be distinguished from others" (Article 29 Working Party, 2007, p. 13). This effect is called the 'mosaic-effect' (OMB, 2013). It occurs when the information in an individual dataset, in isolation, cannot be used to identify an individual, but when combined with other available information, it could pose such risk (OMB, 2013). This effect is likely to make much more data subject to data protection legislation than currently assumed and acted upon. As we will show, the key element in this possibility is geographical data.

4. Geographical data

Geographical data are data that, in one way or another, refer to a location on the Earth (Longley, Goodchild, Maguire, & Rhind, 2001,