



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

PII: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.

© 2016 Elsevier Inc. All rights reserved.

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).

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,

pp. 64–65). Examples are a map of roads in a country or a list of addresses in a town. Linking data to a location on the Earth makes the object or subject easy to identify, and as a result easy to reach (Van Loenen, 2006): geographical data are often a key element in identifying or re-identifying individual data subjects (Scassa, 2010). A simple example will illustrate what value geographical data adds to other data. Imagine a situation of Mr Smith whose annual income is €200,000. This information alone is insufficient to trace him, to approach him physically and to exploit the information. However, adding the location of his house to this information allows the public tax office to send a tax form to Mr. Smith's address, and the sales representative of Cadillacs to send a folder of the latest models. Mr. Smith has now become more than his name; he is an asset that is easy to reach. When we include his attributes in a database with all inhabitants of the area he lives in, we can map the income distribution, the distribution of sexes, or the distribution of all people owning a Cadillac. This example can be applied to many more human activities and decisions (see Scassa, 2013a; Scassa, 2013b).

In this article, we focus on basic geographical data of the kind typically provided by governments: mapping data and the underlying earth observation data. Mapping data, such as topographical maps and aerial imagery, are identified as the foundation of national information infrastructures (see Nebert, 2004). Furthermore, the European Commission has ranked these mapping datasets as the highest priority for being made available for re-use due to the high demand from re-users across the EU (see European Commission, 2014; see also Cabinet Office, 2013).

5. Are mapping data personal data?

In Section 3 we explained that in determining whether a dataset identifies individuals all the means likely reasonably to be used either by the controller or by any other person should be taken into account (see Recital 26 of the EU Data protection Directive). In this section, we will assess to what extent mapping data identifies individuals.

Typically, the Data Protection Directive applies to mapping data with a high level of detail such as maps showing individual houses, (e-mail, Internet protocol, or portal) addresses, house numbers and cadastral parcel numbers (see Article 29 Working Party, 2007; Article 29 Working Party, 2011; Graux, 2011; Dutch Government, 1998a,b, 1999a,b, 2000; Registratiekamer, Dutch Data Protection Agency, 1996). These data can easily be linked to individuals through, for example, publicly available phone books (in print or accessible and searchable online), or public information available in the land registry. Along the same lines, aerial photographs (CBPL, Commissie voor de bescherming van de persoonlijke levenssfeer (Flemish Data Protection Agency), 2006a; Karg, 2008) and 360° images of buildings, such as shown on Google Streetview are within the scope of the Data Protection Directive (Van der Sloot & Zuiderveen Borgesius, 2012; see also CBPL, Commissie voor de bescherming van de persoonlijke levenssfeer (Flemish Data Protection Agency), 2010; VTEBG, Vlaamse Toezichtcommissie voor het elektronische bestuurlijke gegevensverkeer, 2011).

This leads to the conclusion that almost all data, if linked to a sufficiently detailed map, can be considered to be personal data, even if the information as such does not necessarily identify a person (Graux, 2011; see for examples also Scassa, 2010, 2013a,b).

But there are deviating opinions. For aerial photographs in the Netherlands, the Amsterdam Court of Appeals (Hof Amsterdam, 2014) ruled that the data concerned the data of an object and therefore should not be considered personal data: "An address, an aerial image (location) and fuzzy images of surroundings of the address, without depicting individuals, are solely data about an object, not personal data about an individual" (translation by the authors).

This ruling is in line with the restrictive interpretation by the Dutch Government of the data protection legislation and its applicability to building and address data. These data include the floor space, year of construction and function of the buildings. According to the Dutch Minister for Housing, Spatial Planning, and the Environment, address and

building data should not be considered personal data because they are not relating to identifiable natural persons (Dutch Government, 2006). Only if the data controller understands that the data requester will be able to link the data to individuals by combining the data with other data, the building and address data should be regarded as personal data.

Similarly, in the United Kingdom, the Land Registry accomplished a privacy impact assessment for releasing its price paid information (PPI), information for all full value residential property sales. It was concluded that "PPI is property related and not personal: the focus of the PPI remains the property and not the owners of the property" (Land Registry, 2012, 2013; see also NSGIC, 2013).

However, we feel the reasoning in all three deviating opinions mentioned above is not in accordance with Recital 26 of the Data Protection Directive: "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". In all three cases, the publicly available records in the land registry make it very easy to link the provided data to a natural person, and as a consequence make these personal data.

We have provided a mixture of (legal) opinions on when mapping data should be regarded personal data. We strongly believe that mapping data of a high level of detail (large-scale maps) are highly likely to be considered personal data.

Our examples show that today's technologies almost always allow data to be linked to a place on a map and as a consequence most often to a person. This makes much more data personal data than foreseen in 1995 when EU Data Protection Directive was enacted.

6. Mapping data is personal data: the implications

Processing of personal data is not by definition unlawful, but the EU Data Protection Directive sets requirements that should be followed when personal data is processed. We present here three requirements of the EU Data Protection Directive that in our opinion are problematic for mapping data. We refer to the review of the Data Protection Directive by Robinson et al. (2009) for other Data Protection Directive issues and challenges.

6.1. Specified, explicit and legitimate purpose principle

The EU Data Protection Directive requires that if personal data is processed, it should be done fairly and for specified, explicit and legitimate purposes (Article 6 of the EU Data Protection Directive). The purposes for which the data is processed must be explicit and legitimate and must be determined at the time of collection of the data (Recital 28 of the EU Data Protection Directive).

The requirement of 'explicit' and 'specified' purposes may give rise to problems when mapping data are personal data. This will especially be problematic in the case of open mapping data. The purpose of open data policies is to allow access and re(use) of data without any limitations to the use. Unconditional (re)use of data is then certainly not specific enough to fulfil the requirement of a specified purpose (see Kulk & Van Loenen, 2012).

In addition, in instances where mapping data are not available as open data, the current purposes for mapping data processing are often not specific enough to fulfil the requirement of the processing having a specific purpose. For example, the Flemish Geographical Data Infrastructure (GDI) (Belgium) aims to optimize the collection, maintenance, exchange, use and re-use of mapping data and services. Stakeholders can use mapping data and services for the execution of tasks of public interest, including those regarding the environment. The Belgian Data Protection Commission argued that the GDI of Flanders concerns personal data (among other data), but does not meet the legally required data protection standard: "there is a lack of specific and well determined goals, the allowed use is too abstract, the need and proportionality of the use of personal data for the realisation of SDI goals are difficult to assess, and also the assessment whether use is incompatible or not

with initial purposes of data processing is difficult to determine” (CBPL, *Commissie voor de bescherming van de persoonlijke levenssfeer* (Flemish Data Protection Agency), 2008)(translation by the authors).

6.2. Data quality principles

Regarding data quality, the EU Data Protection Directive states that the processing of personal data should be adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed (Article 6(1)(c) EU Data Protection Directive). The controller must also ensure that the processing of personal data is accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified (Article 6(1)(d) EU Data Protection Directive). Again, this seems problematic for mapping data. For example, one may argue that individuals may require that an outdated aerial image, not showing the actual building (e.g. a newly built garage), will be rectified. The same will apply for an image showing a building, which was demolished after the image was taken. While this requirement for accurate and up to date information may benefit the information infrastructure of an organization or country, maintaining it for mapping data would put significant pressure on the budgets of data providers.

6.3. Processing for the purposes for which the data were collected is no longer necessary

Finally, the personal data must be kept in a form which permits identification of data subjects for no longer span of time than is necessary for the purposes for which the data were collected or for which they are further processed (Article (6)(1)(e) EU Data Protection Directive).

In Belgium, detailed satellite images are considered personal data. In this specific case, the Flemish Government used satellite imagery to identify and affirm building violations by comparing images on a regular basis (CBPL, *Commissie voor de bescherming van de persoonlijke levenssfeer* (Flemish Data Protection Agency), 2006a). The Data Protection Agency advised that if a comparison of the images does not result in an identification of a building violation and the images are not considered useful for future identifications, they should be deleted immediately. Also this has clearly a significant impact on the national information infrastructure.

7. Open data while safeguarding data protection

As we have seen in Section 3, the line between personal data and non-personal data is shifting because of developments in technology and datasets (see also Schwartz & Solove, 2011). Due to big data and open data, more mapping data will become personal data, which in turn has a major impact on the processing of these data in both the public and commercial sectors. Therefore, it is unlikely that the way the public and commercial sectors currently process mapping data will continue to be allowed in the near future. This is an undesirable situation from both information infrastructure and open data philosophy perspectives, both aimed to maximize the use of data.

7.1. Possible directions

Several solutions are identified to bring data protection and information infrastructure and open data interests to an acceptable common ground. We will discuss in this section several options that may bridge the gap between data protection requirements and mapping data utilization interests: (1) privacy-enhancing technology, (2) the Personally Identifying Information 2.0 concept, (3) moving the responsibility from the data controller to the data user, (4) licensing personal data, and (5) limiting access to directly identifying information sources.

7.2. Privacy-enhancing technology

The EU Data Protection Directive does not apply to data that is rendered anonymous in such a way that the person is no longer identifiable (Article 29 Working Party, 2013a). Therefore, anonymizing the personal mapping data seems to provide the solution. However, there is no clear answer to the question when data can be considered to be anonymous. According to the Article 29 Working Party, the assessment of whether data allows identification of an individual, and whether the information can be considered as anonymous or not depends on the circumstances of the case. Therefore, a *case-by-case* analysis should be carried out with particular reference to the extent that the means are likely reasonably to be used for identification (Article 29 Working Party, 2007). There are different ways to aggregate data in order for it to become anonymous. Here, we discuss two common practices:

1. only publish an average value for an area, and
2. reduce the level of detail of the mapping data (map generalization).

7.3. Only publish an average value for an area (with a minimum number of measurements)

The Statistics Netherlands (CBS) applies this method for demographic data. Data is published on a so-called ‘zipcode level’ map (approximately 15 buildings or households). Based on this data, a national newspaper (*NRC Handelsblad*) created an interactive website showing demographic data at zipcode level, such as the percentage of non-western inhabitants of a neighborhood, the percentage of people older than 75 years, the average net income, and one-parent families. They used the weighted average per pixel. The closer a pixel is to a measure, the more weight for the value. If there are less than 5 values in a radius of a hundred meters, then no pixel will be shown (Poort, 2012).

A similar strategy is used by the police in England, Wales and Northern Ireland for their UK crime maps. The UK Crime map provides street-level maps mapping a variety of crime data to an anonymous point, typically the geographical center of the street. For data protection reasons, they only do this if there are at least eight postal addresses in a street. If there are fewer than eight postal addresses on a street, the crimes may be repositioned to and added to the values of a nearby street (UK Comptroller and Auditor General, 2012, p. 25; see also Graux, 2011, pp. 13–14; see also The Task Force Smart Grids Expert Group 2, 2011).

Although these solutions look promising, they do not always mask the underlying personal data. In the case of the CBS data, research (Koot, 2012) revealed that adding other datasets, e.g. data related to birth or gender, to the ‘anonymous’ zipcode level resulted in uniquely identifying 99% of the individuals. In addition, in the example of anonymous UK crime maps, the de-anonymization process may be made far easier because of the police forces using social media (Twitter and Facebook) to report their activities to the public. This information may enable de-anonymization of the anonymized crime map.

7.4. Reduce the level of detail of the mapping data (map generalization)

In Belgium, the Data Protection Agency argued that only at an anonymized level of detail of 1:50,000 or less, mapping data would not be considered personal data while mapping data at a level of detail of 1:10,000 would always be personal data (CBPL, *Commissie voor de bescherming van de persoonlijke levenssfeer* (Flemish Data Protection Agency), 2006b). A similar approach is followed in Germany: mapping data at a 1:10,000 level of detail are considered to be anonymous (see *Deutscher Bundestag*, 2008; see also Karg, 2008).

However, whether or not individual buildings will be shown on a map depends on the choices made in the mapping process and not the scale as such. One cartographer may decide to include the buildings while another may very well decide to leave the buildings out. Fig. 1 exemplifies this in two 1:50,000 maps of the same area provided by the

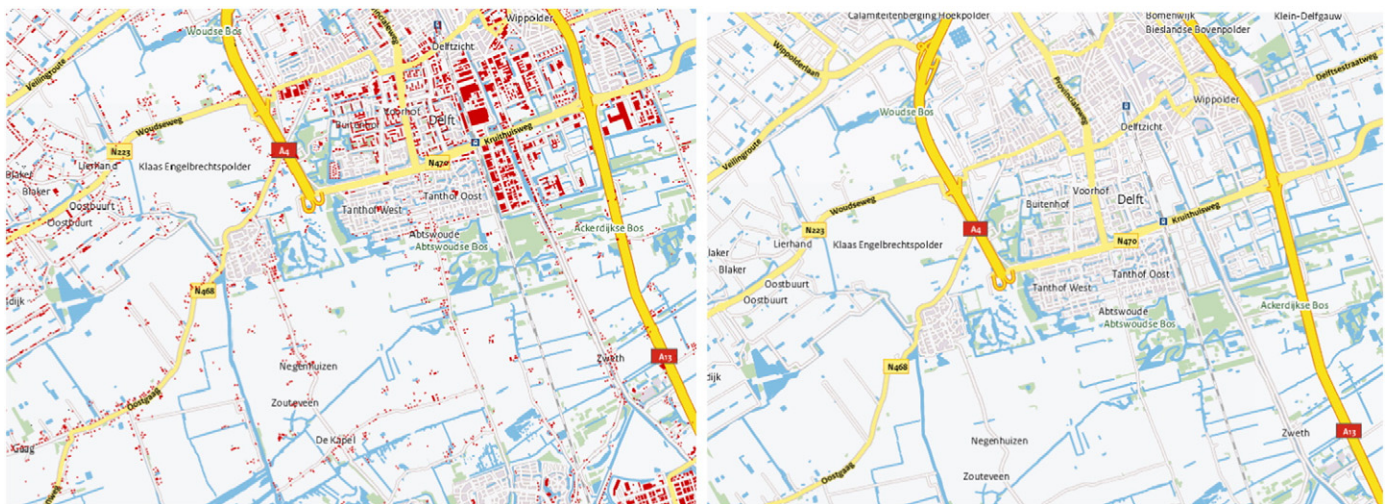


Fig. 1. Example of a 1:50,000 scale map with individual buildings (left) and without individual buildings (right). Courtesy of the Dutch Kadaster.

same organization (the Netherlands Kadaster). Existing services, such as provided by Nieuwsinkkaart,¹ allow the combination of different mapping data. With a few mouse clicks the individual addresses of a house can easily be identified on a 1:50,000 map.

These examples show that fully anonymous data are not so easy to create. Even aggregated data may turn out to be indirectly identifiable by adding data from other sources. The risk of de-aggregation of data will only grow as computing power increases, data analysis techniques to re-identify individuals advance, and more data becomes available as open data (see de Montjoye, Radaelli, Kumar Singh, & Pentland, 2015; Pandurangan, 2014; de Montjoye, Hidalgo, Verleysen, & Blondel, 2013; Koot, 2012; Simpson, 2011; Ohm, 2010; Narayanan & Shmatikov, 2008; Golle, 2006; Barbaro & Zeller, 2006; Sweeney, 2006). The European Data Protection Supervisor notes that “unless full anonymization can be completely ensured, data protection requirements continue to apply” (EDPS, European Data Protection Supervisor, 2012; see also Article 29 Working Party, 2013b). We conclude that ‘total anonymization’ can only be fully guaranteed at very general levels. This is possibly at the levels where the data is of no or very limited use (cf. Cavoukian & Castro, 2014).

7.5. Personally Identifiable Information 2.0

Another option may be to change the concept of personal data. The evaluators of the Data Protection Directive acknowledged in 2009: “The application scope of the Directive depends too strongly on whether or not the data processed can be defined as ‘personal’ data. It is all or nothing: there is no room for ‘more or less personal’ data (and accordingly “more or less protection”). [...] Strict application of the Directive’s concepts sometimes leads to unpredictable or counterintuitive results” (Robinson et al., 2009; see also Schwartz & Solove, 2011).

To address the flaws in the United States and European Union concepts of personal data, Schwartz and Solove (2011) introduced the concept of Personally Identifiable Information 2.0 (PII2.0). This concept calls for a different regime for identified and identifiable data (see Schwartz & Solove, 2011; see also El Emam, 2010; Karg, 2008) based on the risk of identification. When the risk of identification is high, there is a significant probability that a party can link the data to a specific person. Therefore, high risk identifiable data should have more strict data protection requirements than data with lower risks of identification. The result is that the necessary legal protections should generally

be different for categories (high, nominal, and low) of identified and identifiable data (Schwartz & Solove, 2011; cf. Karg, 2008). Schwartz and Solove (2011) argue that: “Information refers to an identified person when it singles out a specific individual from others. An individual is identifiable when there is some non-remote possibility of future identification. The risk level for such information is low to moderate.” A sub-category of identifiable information is the nominally identifiable information. This information has not been linked to a certain person yet, but there is a significant probability that such a link will be made. This information should be treated the same as identified information. In the risk assessment, Schwartz and Solove propose that “the means likely to be used by parties with current or probable access to the information, as well as the additional data upon which they can draw” should be taken into consideration.

However, since many mapping data are easy to link to persons through an address and the information available in the land registry, they would fall into the, easy to link to an individual, ‘high risk’ category. Therefore, the PII2.0 concept will not change anything for such mapping data. Moreover, since much data are in the end identifiable regardless of the risk of identification, PII2.0 as such does not provide a solution to our problem (see also Cuijpers & Marcelis, 2012).

However, the concept of PII2.0 may work very well if a different approach is applied. The current distinction between sensitive personal data, directly identifying data, indirectly identifying and non-personal data may be maintained. And, similar to PII2.0, for each category, different data protection rules should apply. Instead of categorizing data based on the risk of identification, it may be decided that certain data types in the indirectly identifying category, such as mapping data, are subject to less strict data processing rules. The current requirements may be kept, but for indirectly identifying mapping data the requirements may be weakened.

7.6. Move the responsibility from the data controller to data user

Typically, providers of open mapping data do not make the link to an individual themselves. It is the re-user combining the base data with other datasets that may create the personal data. Take for example the Cadillac in Section 3. The provider of the map does not link the mapping data to the driver of a Cadillac nor to his income. It is the user of the mapping data that does plot the Cadillac drivers on a map and identify a person. This suggests limiting the scope of the Data Protection Directive to only those users that *do* combine datasets that could result in data that identify individual persons. In such a case, government data

¹ <<http://nieuwsinkkaart.nl/>>.

providers of mapping datasets can provide their data as open data while users need to adhere to the data protection rules.

This approach has several advantages. It makes the very abstract and vague test what exactly is personal data very concrete: if the processed data identifies an individual, the data protection rules apply. This allows for clear boundaries between personal and non-personal data. Moreover, it designates responsibility to the end user, the user who identifies or has the intention to identify individuals to exploit information about them. The proposed approach indemnifies mapping agencies, which only provide single mapping datasets from liability claims of interfering with the right to privacy. It allows them to provide the mapping data as open data from which society will benefit.

One way of implementation in the European Union, is by rephrasing Recital 26 “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” (Recital 26 of the EU Data Protection Directive) into “account should be taken of all the means used by the controller to identify the said person” (cf. [Article 29 Working Party, 2013b](#)).

We assess that the data protection for the individual remains identical to the situation as it is today. However, one may argue that by moving the responsibility from the data provider to the data user, the data provider will not control the data use and/or data user anymore and the knowledge about the use of the data is lost. The assumed lack of control and transparency of the use may impact upon the data protection rights of individuals. However, should it be the role of a data provider to safeguard the data protection of an individual in the instance the provider does not identify or does not have the intention to identify natural persons from his data? Although this direction seems promising, it needs further development and thought to be implemented in law.

7.7. Licensing the use of personal data

The European Data Protection Supervisor (EDPS) and the Article 29 Working Party have suggested that in order to allow re-use of personal data, personal data should be provided with a licence specifically prohibiting the re-identification of individuals and the re-use of personal data for the purposes that may individually affect the data subjects ([Article 29 Working Party, 2013a](#); [EDPS, European Data Protection Supervisor, 2012](#); see also [Dos Santos et al., 2010](#)).

However, this suggestion does not solve the problem we identified in this article. If it is generally possible to re-identify individuals from a dataset then the EU Data Protection Directive applies. Since personal data can only be processed for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes, a licence prohibiting the re-identification is doing the same thing as the law.

7.8. Limit the access to sources that provide directly identifying information

An important criterion in the assessment of whether a dataset is identifiable is the effort one makes to identify a natural person. For example, it is very easy to link an individual through publicly accessible information such as information available in the land registries. However, this information is not public in all EU Member States. In Germany, access to the information in the land registries is limited to only those with a legitimate interest (e.g., conveyancing lawyers, national revenue services, owners of real property). Therefore, those with a non-legitimate interest have much more difficulty to identify individuals from mapping datasets. Restricting access to the land registry data may be a solution to our problem (see [Berlee, 2015](#)).

However, the land registries are only one of the sources of directly identifying information. Other sources are widely available to the public, for example on social media. Therefore, this proposed solution can unlikely be sustained.

8. General data protection regulation

The world changed dramatically since the introduction of the Data Protective Directive in 1995. The fast developments in information and communications technology offer new possibilities, but also pose threats. These threats may undermine the trust of consumers in the online markets and therefore slow down economic growth ([European Commission, 2010](#)).

In order to address these new threats, the European Commission proposed a new data protection framework to replace the Data Protection Directive of 1995 (see [European Commission, 2012b](#); see also [European Parliament, 2014](#); [European Council, 2015a](#); [European Council, 2015b](#)). At the end of 2015, the European Parliament, Council and Commission reached an agreement to a consolidated text of the new General Data Protection Regulation (GDPR). This regulation, aimed at safeguarding the right to the protection of personal data in the light of the developments in the past decades (see Recital 5 GDPR), was politically agreed upon on 28 January 2016 (see [European Council, 2016](#)). The regulation is expected to enter into force in Summer 2016 and will then be applicable as of Summer 2018.

Although the new Regulation introduces new concepts such as the right to be forgotten, data portability, personal data breach notification, profiling, and easier access to their own data to strengthen citizens' fundamental rights (see, for example, [De Hert & Papakonstantinou, 2012](#); [Mantelero, 2013](#)), it did not fundamentally change the concept of personal data. Moreover, the data quality principles remain almost identical in the GDPR. Personal data must still be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes (Art. 5 (1b) GDPR). Moreover, personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (Art. 5 (1c) GDPR), and personal data must be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data are processed (Art. 5 (1d) GDPR).

Provided the almost unchanged concept of personal data and the data quality principles of the GDPR, our research findings remain unaffected by the new GDPR.

9. Conclusion

In this article, we argue that increased computer power, advancing data mining techniques and the increasing amount of publicly available data extend the reach of the EU data protection legislation to much more mapping data than currently assumed and acted upon. This could in effect obstruct the implementation of open data policies in the EU. Many of today's mapping authorities, supplying open mapping data, should expect a re-assessment of their role in identifying individuals and, as a result, mapping data may no longer be open in the near future.

In the context of mapping data, we showed that it is difficult to draw a clear and unambiguous line between when mapping should be considered personal data and when it should not. It depends very much on the circumstances of every case. A major issue is that the European data protection legislation requires that personal data can only be processed for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. However, mapping data processing often lacks an explicit and specific purpose. Therefore, the EU data protection legislation is a much more serious precondition for mapping data processing both in public and commercial sectors than currently assumed. And this directly impacts other interests of our information societies: promoting the free sharing of government information including open (mapping) data.

Our conclusion is that there are two options to restrain the ‘very hungry’ data protection legislation: the legislator should either adapt the Personally Identifying Information 2.0 concept with a personal data category having less restrictive data processing requirements, or

move a part of the data protection obligations from the data provider to the data user.

The hunger of the EU data protection legislation remains a serious problem for the successful execution of the EU digital agenda. We will have to await the practical impact that the EU General Data Protection Regulation will have on the availability of open mapping data. Our suggestions may help to arrive at a middle way between the interests of data protection and open data. In this way, the current and the expected widening of scope of the data protection legislation may be overcome and the benefits of open data may still be realized.

Acknowledgements

The authors gratefully acknowledge the support of the BSIK innovation program Next Generation Infrastructures (09.07.OTB), the STW-Maps4Society program (13718), and the Dutch Open data breakthrough team commenced by the Dutch Ministry of Economic Affairs. In addition, the help of Dr. Martijn Meijers and Mr. Benny Onrust of Delft University of Technology, and Gillian AvRuskin is very much appreciated.

References

- Article 29 Working Party (2000). *Working document on privacy on the Internet – An integrated EU approach to on-line data protection*, WP 37.
- Article 29 Working Party (2005). *Working document on data protection issues related to RFID technology*, WP105.
- Article 29 Working Party (2007). *Opinion 4/2007 on the concept of personal data*, WP 136.
- Article 29 Working Party (2011). *Opinion 12/2011 on smart metering*, WP 183.
- Article 29 Working Party (2013a). *Opinion 03/2013 on purpose limitation*, WP 203.
- Article 29 Working Party (2013b). *Opinion 06/2013 on open data and public sector information ('PSI') reuse*, WP207.
- Barbero, M., & Zeller, T. (2006). A Face Is Exposed for AOL Searcher No. 4417749. *The New York Times*.
- Berlee, A. (2015). *Meer aandacht voor privacy in de openbare registers?* *Nederlands Juristenblad* 2015/1091.
- Cabinet Office (2013). *Policy paper: G8 open data charter and technical annex*. 18 June 2013.
- Cavoukian, A., & Castro, D. (2014). Big data and innovation, setting the record straight: De-identification does work. Retrieved from: <http://www2.itif.org/2014-big-data-deidentification.pdf>
- CBPL (Commissie voor de bescherming van de persoonlijke levenssfeer (Flemish Data Protection Agency)) (2006v). *Adviesaanvraag inzake het gebruik van satellietbeelden bij de opsporing en de vaststelling van bouwovertreedingen*, Advies nr. 26/2006.
- CBPL (Commissie voor de bescherming van de persoonlijke levenssfeer (Flemish Data Protection Agency)) (2006v). *Bijhouden van gemeentelijke registers van onbebouwde percelen waarvan sprake in artikel 62 van het Vlaams Decreet van 18 mei 1999 houdende de organisatie van de ruimtelijke ordening en hun bekendmaking op het internet via het toekomstig loket*, Advies Nr 40/2006.
- CBPL (Commissie voor de bescherming van de persoonlijke levenssfeer (Flemish Data Protection Agency)) (2008v). *Advies nr. 32/2008 van 24 september 2008 inzake het voorontwerp van decreet betreffende de Geografische Data-Infrastructuur Vlaanderen (A/2008/032)*.
- CBPL (Commissie voor de bescherming van de persoonlijke levenssfeer (Flemish Data Protection Agency)) (2010v). *Aanbeveling uit eigen beweging inzake Mobile Mapping (CO-AR-2010-007)*, Aanbeveling nr 05/2010.
- Cuijpers, C., & Marcelis, P. (2012). Oprekking van het concept persoonsgegevens beperking privacybescherming? *Computerrecht*, 6(November), 397–409.
- De Hert, P., & Papakonstantinou, V. (2012). The proposed data protection regulation replacing Directive 95/46/EC: A sound system for the protection of individuals. *Computer Law & Security Review*, 28(2), 130–142.
- Dekkers, M., Polman, F., te Velde, R., & de Vries, M. (2006). Measuring European public sector information resource. *Final report of study on exploitation of public sector information – Benchmarking of EU framework conditions*.
- Deutscher Bundestag (2008). *Beschlussempfehlung und Bericht des Ausschusses für Umwelt, Naturschutz und Reaktorsicherheit, Entwurf eines Gesetzes über den Zugang zu digitalen Geodaten (geodatenzugangsgesetz-GeoZG)*, Drucksachen 16/10892.
- Dos Santos, C., Bassi, E., De Terwagne, C., Salmeron, M., Fernandez, Tepina, P., & van der Sloot, B. (2010). LAPSI policy recommendation No. 4. Privacy and personal data protection. LAPSI Working Group 2 privacy aspects of PSI. http://www.lapsi-project.eu/lapsifiles/lapsi_privacy_policy.pdf
- Dutch Government (1998a). *Explanatory memorandum Data Protection Act (Memorie van Toelichting bij de Wet bescherming persoonsgegevens)*, Kamerstukken II 1997–98, 25892, Nr. 3.
- Dutch Government (1998b). *Regels inzake de bescherming van persoonsgegevens (Wet bescherming persoonsgegevens)*, Kamerstukken 25892 Nr. 6.
- Dutch Government (1999a). *Regels inzake de bescherming van persoonsgegevens (Wet bescherming persoonsgegevens)*, Kamerstukken 25892 Nr. 9.
- Dutch Government (1999b). *Regels inzake de bescherming van persoonsgegevens (Wet bescherming persoonsgegevens)*, Kamerstukken 25892 Nr. 13.
- Dutch Government (2000). *Regels inzake de bescherming van persoonsgegevens (Wet bescherming persoonsgegevens)* MEMORIE VAN ANTWOORD, Eerste Kamer vergaderjaar 1999–2000, 25 892, nr. 92c.
- Dutch Government (2006). *Explanatory memorandum Building and Addresses Registration Act (Memorie van Toelichting bij de Wet BAG)*, Kamerstukken II 2006/07, 30 968, nr. 3.
- EDPS (European Data Protection Supervisor) (2012). *Opinion of the European Data Protection Supervisor on the 'Open-Data Package' of the European Commission including a proposal for a directive amending Directive 2003/98/EC on re-use of public sector information (PSI)*, a Communication on Open Data and Commission Decision 2011/833/EU on the reuse of Commission documents.
- El Emam, K. (2010). Risk-based de-identification of health data. *IEEE Security & Privacy*, 8(3), 64–67.
- European Commission (2010). *Communication to the European Parliament, the Council, the European Economic and Social Committee, and the Committee for the Regions. A digital agenda for Europe*, COM (2010) 245 final.
- European Commission (2011). *Communication to the European Parliament, the Council, the European Economic and Social Committee, and the Committee for the Regions. Open data: An engine for innovation, growth and transparent governance*, COM (2011) 882 final.
- European Commission (2012a). *Impact assessment accompanying the document regulation of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) and directive of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data*. Commission staff working paper, SEC(2012) 72 final.
- European Commission (2012b). *Proposal for a regulation of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, COM (2010) 11 final.
- European Commission (2014). *Commission notice – Guidelines on recommended standard licences, datasets and charging for the reuse of documents*, OJ, 2014, C240/01.
- European Council (2015a). *Proposal for a regulation of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) [first reading] – Analysis of the final compromise text with a view to agreement, (15039/15 – 2012/0011 (COD))*, 15 December.
- European Council (2015b). *Proposal for a regulation of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, preparation of a general approach, (9565/15 – 2012/0011 (COD)), 11 June.
- European Council (2016). *Proposal for a regulation of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) [first reading] – Political agreement, ST 5455 2016 INIT – 2012/011 (OLP)*, 28 January.
- European Parliament (2014). *European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD))*.
- European Parliament and Council (1995). *Directive 95/46/EC of the European Parliament and of the council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data*, OJ L, 281, 31.
- European Parliament and Council (2003). *Directive 2003/98/EC of the European Parliament and of the council of 17 November 2003 on the re-use of public sector information*, OJ L, 345, 90.
- European Parliament and Council (2013). *Directive 2013/37/EU of the European Parliament and of the council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information*, OJ L, 175, 1.
- Gallo, C. (2012). *Masterclass Steve Jobs*. Business Contract: Amsterdam/Antwerp.
- GAO (2015). *United States Government Accountability Office, facial recognition technology, commercial uses, privacy issues, and applicable federal law. Report to the ranking member, Subcommittee on Privacy, Technology and the Law, Committee on the Judiciary, U.S.: Senate*.
- Golle, P. (2006). Revisiting the uniqueness of simple demographics in the US population. *Paper presented at the WPES '06 of the 5th ACM workshop on privacy in electronic society*.
- Graux, H. (2011). *Open government data: reconciling PSI re-use rights and privacy concerns*. European public sector information platform topic report No. 2011/3.
- Hof Amsterdam (2014). *Zaaknummer: 200.105.659/01 KG. Computerrecht 2014/116 met annotatie door F.C. van der Jagt*.
- IBM (2013). *What is big data?* Retrieved from <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
- Janssen, K. (2011). *The influence of the PSI directive on open government data: An overview of recent developments*. *Government Information Quarterly*, 28(4), 446–456.
- Karg, M. (2008). *Datenschutzrechtliche Rahmenbedingungen für die Bereitstellung von Geodaten für die Wirtschaft*. Unabhängiges Landeszentrum für Datenschutz. Schleswig-Holstein (ULD): Gutachten im Auftrag der GLW-Kommission.
- Koot, M. R. (2012). *Measuring and predicting anonymity*. Amsterdam: University of Amsterdam.
- Kulk, S., & van Loenen, B. (2012). *Brave new open data world? International Journal of Spatial Data Infrastructures Research*, 7, 196–206.
- Land Registry (2012). *Privacy impact assessment report*. Making price paid data available through publication in a machine readable and reusable format.

- Land Registry (2013). *Review of privacy impacts — Price paid data*.
- van Loenen, B. (2006). *Developing geographical information infrastructures: The role of information policies*. Delft: DUP Science.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2001). *Geographical information systems and science*. Chichester, England: John Wiley and Sons Ltd.
- Mantelero, A. (2013). The EU proposal for a general data protection regulation and the roots of the 'right to be forgotten'. *Computer Law & Security Review*, 29(3), 229–235 June 2013.
- de Montjoye, Y. -A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Nature scientific reports*, 3, 1376.
- de Montjoye, Y. -A., Radaelli, L., Kumar Singh, V., & Pentland, A. "Sandy" (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347(6221), 536–539.
- Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. *Paper presented at the IEEE Symposium on Security and Privacy*.
- Nebert, D. (2004). Spatial data infrastructure cookbook. Retrieved from: http://www.gsdidocs.org/GSDIWiki/index.php/Main_Page
- NSGIC (2013). This isn't private information. Retrieved from http://www.nsgic.org/public_resources/This_Isnt_Private_Information_082313_NSGIC_Outreach_Committee.pdf
- Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57(2010), 1701–1777.
- OMB (2013). Open data policy — Managing information as an asset. *Memorandum for the heads of executive departments and agencies*.
- Pandurangan, V. (2014). On taxis and rainbows. Lessons from NYC's improperly anonymized taxi logs. Retrieved from <https://medium.com/@vijayp/of-taxis-and-rainbows-f6bc289679a1>
- Pira International Ltd., University of East Anglia, & KnowledgeView Ltd. (2000e). Commercial exploitation of Europe's public sector information. *Final report for the European Commission Directorate General for the Information Society*.
- Poort, A. (2012). Statistiek saai? CBS-cijfers komen tot leven op de kaart. *NRC Handelsblad* (pp. 12–13) Retrieved from <http://www.nrc.nl/nieuws/2012/02/14/statistiek-saai-cbs-cijfers-komen-tot-leven-op-een-kaart/>
- Registratiekamer (Dutch Data Protection Agency) (1996). *Credit scoring database*.
- Robinson, N., Graux, H., Botterman, M., & Valeri, L. (2009). *Review of the European Data Protection Directive*. Santa Monica, CA: RAND Corporation technical report series.
- Scassa, T. (2013a). More on privacy and public gun permit data. Retrieved from http://www.teresascassa.ca/index.php?option=com_k2&view=item&id=120:more-on-privacy-and-public-gun-permit-data&Itemid=80
- Scassa, T. (2010). Geographical information as 'personal information'. *Oxford University Commonwealth Law Journal*, 10(2), 185–214.
- Scassa, T. (2013b). *Information maps*. Freedom of Expression and Privacy Retrieved from http://www.teresascassa.ca/index.php?option=com_k2&view=item&id=116:information-maps-freedom-of-expression-and-privacy&Itemid=81
- Schwartz, P. M., & Solove, D. J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *New York University Law Review*, 86(6), 1814–1894.
- Simpson, A. C. (2011). On privacy and public data: A study of data.gov.uk. *Journal of Privacy and Confidentiality*, 2011(1), 51–65.
- Sweeney, L. (2006). *Uniqueness of simple demographics in the U.S. Population*: Carnegie Mellon University.
- The Task Force Smart Grids Expert Group 2 (2011). *Essential regulatory requirements and recommendations for data handling, data safety, and consumer protection recommendation to the European Commission (final draft)*.
- UK Comptroller and Auditor General (2012). *Implementing transparency: Cross government review*. London: National Audit Office.
- Van der Sloot, B., & Zuiderveen Borgesius, F. J. (2012). Google's dead end, or: On Street View and the right to data protection: An analysis of Google Street View's compatibility with EU data protection law. *Computer Law Review International*, 4, 103–109.
- Vickery, G. (2011). *Review of recent studies on PSI re-use and related market developments*.
- VTEBG (Vlaamse Toezichtcommissie voor het elektronische bestuurlijke gegevensverkeer) (2011s). *Aanvraag tot machtiging van het ter beschikking van de mobile mapping beelden door het Agentschap voor Geografische Informatie Vlaanderen (AGIV) aan het AGIV zelf en de steden en gemeenten en de deelnemers aan GDI-Vlaanderen voor een aantal taken in het kader van de uitvoering van het CRABdecreet, Beraadslaging VTC nr. 15/2011 van 22 juni 2011*.
- Watts, M., Brunger, J., & Shires, K. (2011). Do European data protection laws apply to the collection of WiFi network data for use in geolocation look-up services? *International Data Privacy Law*, 1(3), 149–160.