Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection

Bart Custers* and Helena Uršić**

Introduction

Big Data consists of very large amounts of (often real-time) data that are preferably but not necessarily structured.¹ Large databases are available everywhere: almost every company, government organization, and non-profit organization collects personal data about (actual and potential) customers, suppliers, and employees.² On top of that, most organizations collect information about their business processes, including product information, payment information, shipping information, etc. With the introduction of the ‘Internet of things’ in which objects are equipped with interconnected embedded devices transmitting information about their status and location a whole new type of information is introduced, omitting human agents who register data in databases.³ Realizing the potential of Big Data calls for more data reuse from all sources.

There are, however, several practical, technological, and legal barriers preventing easy reuse of data. Power is knowledge, and therefore, many organizations are unwilling to share with others the data they collected, because they consider it valuable as a competitive advantage. At the same time, for many organizations, it would be very efficient if they would have easy access to databases of other organizations and could use available data for all kinds of purposes. Technological barriers may consist of databases in different forms that cannot be coupled easily, and legal barriers may consist of privacy issues and intellectual property issues. Such barriers obviously cause friction. There are situations in which

Key Points

- Current personal data protection requirements like data minimization and purpose specification are potentially inimical to Big Data as they limit the size and use of Big Data. Potential benefits of Big Data to discover novel trends, patterns, and relationships may not materialize. Substantial loss of economic and social benefits of Big Data may be the result.
- In order to realize Big Data benefits, the reuse of data could be encouraged. Data reuse, when done properly, may be both privacy preserving and economically and socially beneficial.
- In this article, we provide a taxonomy of data reuse from both the data controller’s and the data subject’s perspective that may be useful to determine the extent to which data reuse should be allowed and under which conditions. From the data controller’s perspective, we distinguish data recycling, data repurposing, and data recontextualization. From the data subject’s perspective, we distinguish data sharing and data portability.
- Forms of data reuse that stay close to the awareness and intentions of data subjects should be approached less restrictively (for instance, by assuming informed consent), whereas for forms of data reuse that are ‘at a distance’, ie, in which awareness and transparency may be lacking and data subject’s rights may prove more difficult to exercise, more restrictions, and additional protection should be considered (for instance, by requiring explicit consent).

organizations ask customers or citizens several times for the same information, simply because they need it for another purpose or another database. Organizations are dissatisfied because this is economically inefficient, and customers and citizens may be irritated about completing yet another form.

The benefits of Big Data lie in the possibilities to discover novel trends, patterns, and relationships by combining very large amounts of data from different sources. The developments in the area of Big Data call for new technological, economic, and legal models in which the reuse of data is encouraged rather than hindered. Among the many challenges that Big Data raises, data reuse is one of the most urgent. For some types of data, like research data, not much sharing is actually taking place. From a technological perspective, a minimum condition for data reuse is an adequate technological infrastructure. Obviously, standardized approaches in IT architecture and data formats further facilitate data reuse and data exchanges, but also aspects like scalability, aggregation, reliability, availability, and security are relevant in this respect. From a societal perspective, it may in some situations be desirable to reuse data for different purposes or in different contexts, but in other situations, public support may be lacking. From an economic perspective, it may be very valuable to combine data from different sectors, but it may also occur that companies prefer to refrain from data sharing, for instance, when it concerns their key business models or intellectual property.

From a legal perspective, on which we will focus in this article, the current personal data protection requirements like data minimization and purpose specification are potentially inimical to Big Data as they limit the size of datasets or intellectual property.

We provide a taxonomy of data reuse that may be useful to determine the extent to which data reuse is allowed under current data protection laws. The taxonomy may also be useful to determine the extent to which data reuse is privacy preserving on the one hand and economically and socially beneficial on the other hand. We will argue that forms of data reuse that stay close to the awareness and intentions of data subjects should be approached less restrictively (for instance, by assuming informed consent), whereas for forms of data reuse that are ‘at a distance’, i.e., in which awareness and transparency may be lacking and data subject’s rights may prove more difficult to exercise, more restrictions, and additional protection should be considered (for instance, by requiring explicit consent).

This article describes the first results of the EUDECO project, an EU funded project under the Research and Innovation Framework Program Horizon 2020 of the European Commission. This project is focussed on addressing the question of the use and reuse of personal data. In this article, we specifically focus on data reuse in the context of legal restrictions.

This article is structured as follows. Section 2 elaborates on the emergence of Big Data and its actual and potential benefits. Section 3 provides a taxonomy for data reuse from both the data controller’s and the data subject’s perspective. Section 4 provides conclusions.

## Benefits of big data

### Datafication

In 2011, McKinsey released its breakthrough report on big data, revealing that the big data revolution has finally reached all the sectors in the industry. The amount of data available in scientific fields such as biology and physics has grown far above our imagination. For instance, the genetic sequencing data stored at the EU Bioinformatics Institute have literally exploded. In past 5 years, it has increased exponentially up to 20 petabytes. Still, this data pile is just one-tenth the size of the...
tremendous amount of data stored at CERN, Swiss particle-physics laboratory.\(^{12}\) What is essential to the phenomenon of big data is that the vast amount of data does not lie dormant, but is being used in innovative purposes to release its power anew.\(^{13}\)

Today, the discussion on big data is mostly economically oriented. More than its collection, storage, and ownership, the burning question is how big data can create value, what its economic benefits are, and how can it help the leading companies outperform their peers.\(^{14}\) In a similar way that oil laid the foundation of the smoke-stack economy, big data exploitation has become a foundational building block of the information economy.\(^{15}\)

**Unleashing big data value**

The possibility of big data reuse has triggered a number of cutting-edge business models.\(^{16}\) After all, the world’s most successful and innovative companies—Amazon, Google, Walmart, and, last but not least, Facebook—\(^{17}\) have built their business model on the collection and exploitation of big data.

Based on value creation, Mayer-Schoenberger & Cukier differentiate three types of big data business models. The first group are the organizations that own the data but turn to independent firms to license it to others to use. An example is Twitter, which enjoys a massive stream of data flowing through its servers, but is not willing or not able to reuse it in new ways.\(^{18}\) The second group are the organizations that extract big data value by engaging their employees’ analytical skills. Agencies that offer strategic digital consultancy have been flourishing. It is a presumption that big clients from traditional industries lack the skills that would enable them to perform valuable analyses. Data analytics help clients tailor the data to draw useful actions and improve key performance indicators. Finally, there is a group of firms that not only own data and sufficient analytics skills but also the mind-set with ideas about original ways to tap data to unlock new forms of value. For instance, by analysing billions of users’ failed search attempts and their typos, Google managed to develop the world’s most complete spell checker in basically every living language. The novelty in Google’s approach was in showing that ‘bad’, ‘incorrect’, or ‘defective’ data can still be very useful.\(^{19}\)

Those three groups do not operate in an isolated way but do interact with each other as well as with the rest of business players. The need to liaise with numerous actors on the big data market has triggered the growth of intermediaries—platforms that enable access to data\(^{20}\) to those that lack the assets and that sell the collected data to those that appreciate it more.\(^{21}\) Big data benefits are now spreading across the global economy.

**Legal trade-off between benefits and risks**

The emergence of big data and its proven benefits have demonstrated the complexity of the legal discussion, since current regulation may not sufficiently respond to the challenges related to big data sets,\(^{22}\) in particular in cases of its reuse. On the one hand, there have been claims that our society should expect a substantial loss of benefits of big data, if it attempts to confine it within an obsolete legal framework.\(^{23}\) On the other hand, we cannot turn a blind eye to the grey side of the big data revolution and its numerous risks.\(^{24}\) Big data reuse could be an increasing source of consumer detriment in terms of privacy, security, and consumers’ rights.\(^{25}\)

The European Commission first addressed the questions of data reuse by introducing the Directive 2003/98/EC of the European Parliament and of the Council on the reuse of public sector information.\(^{26}\) Due to many positive side effects, various open data initiatives have been also been increasingly interesting for the for-profit businesses, but never really attracted their attention until

---

13 See above n 3.
14 See above n 13.
18 See above n 3.
19 Ibid.
20 Ibid.
21 An example of a big data intermediary is a Swiss platform collecting medical data and re-selling it to (mostly) pharmaceutical companies <https://www.healthbank.ch> accessed 27 July 2015.
22 See above n 6.
24 See above n 3.
26 [2003] OJ L 345. The directive includes the definition of data reuse: ‘[…] the use by persons or legal entities […] other than the initial purpose within the public task for which the documents were produced.’
fairly recently. What the EU has succeeded in, however, has been a more active cooperation between public and private sectors in the form of private–public partnerships.

Today, the European legislator has shifted its regulatory focus to unleashing big data opportunities. Most recently, the Commission has taken important steps towards a more data reuse friendly legal landscape by releasing its Communication on a data-driven economy and proposing changes to the existing legal framework of EU Directive 95/46/EC on the Protection of Personal Data under the umbrella of the Agenda 2020 programme. While the EU has proclaimed the digital improvement its main goal for the future, it has also stressed the importance of strong and genuine protection of consumer rights in the information society. A tough trade-off between data protection and business incentives has been well noticed in the lengthy and complex process of adopting the new data protection regulation.

The value of big data is not in the mere data collection but in the insights deduced from it. As the focus of the data industry is shifting towards its value-adding reuse, a number of organizations and business associations have called for legal protections to focus more on how data might be used rather than limit what data can be collected. In the era of big data, the motives let alone the identity of the data reuser may be hidden and individual expectations may be confused. Abstract language of the Data Protection Directive (95/46/EC) adopted 20 years ago does not adequately answer the opaqueness and vagueness of many big data practices, let alone its reuse. The taxonomy of data reuse explained in the following section may be useful to determine the extent to which data reuse is allowed under current data protection laws and to find the right balance between economic incentives and ethical and transparency issues of data reuse.

**Data reuse: different categories**

A prerequisite for developing a taxonomy for data reuse is addressing the question what exactly is data reuse. The term data reuse in its broadest sense suggests that there is initial (primary) use of data and subsequent (secondary) use of data, ie, the reuse of data. The distinction between use and reuse may imply different aspects, however. In this section, we distinguish different types of data reuse and provide a taxonomy for data reuse. We start with discussing what exactly is data use, since data reuse can only happen after data use. The section ‘Data Reuse from the Data Controller’s Perspective’ will focus on data reuse aspects for data controllers (particularly data recycling, data repurposing, and data recontextualization), and the section ‘Data Reuse from the Data Subject’s Perspective’ will focus on data reuse aspects for data subjects (particularly data sharing, data portability, and the right to be forgotten).

**What is data use?**

Before we discuss data reuse, it is necessary to explain data use. This is important since data reuse (secondary use) can only happen after data use (primary use). This may seem an irrelevant question, since everyone will have a basic understanding of data use. Data use is something like using available data for a specific purpose. However, the EU regulation on personal data protection complicates this perception. EU Directive 95/46/EC on the protection of personal data does not deal with the concepts of data use and data reuse, but uses the concept of processing of personal data (‘processing’), which is defined in Article 2 as: any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaption or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction.

Whereas data use and data reuse may be conceived as an action that takes place after collection and storage of personal data and before erasure or destruction of the personal data, the legal framework considers collection, storage, erasure, and destruction also as forms of data processing. As a result, it may be suggested that data processing always starts with data collection (the first form of processing and proves that a more systematic taxonomy could be beneficial.
of processing) and that subsequent actions like data storage, preparation, and analyses are all next steps in data processing and, as such, may be considered as reprocessing or reuse. Since data are always collected (and often stored) before it can be used for any purpose, we do not think data collection and storage, though the first steps in most data processing, should already count as data use. Similarly, we do not think erasure or destruction, if these actions take place at the end of a lifecycle of personal data, should be considered as data reuse. In line with the common understanding of data use and data reuse, we do not include the collection, storage, erasure, and destruction of personal data as forms of data use or data reuse. To avoid misunderstandings in terminology, we will not use the term data processing in this article, but stick to data use and data reuse.

Data reuse from the data controller’s perspective

There are several ways in which data controllers may like to reuse personal data they have collected. Here, we distinguish data recycling, data repurposing, and data recontextualization.

Data recycling

The most simplified form of data reuse is using the same data in the same way for more than once. A typical example may be a health insurance company that collects patient data in order to have a proper client database used for billing the insurance premiums that are due and to reimburse medicines, treatments, and therapies. When they use a client’s address for sending them a bill, they will do this monthly, quarterly, or annually. In that sense, they periodically reuse the address more than once for the same purpose. This is a form of data reuse that we will call data recycling. It is rather straightforward, since the data are used over and over again in the same way. There are no significant legal issues here as long as a data subject does not revoke his or her informed consent. For instance, when a data subject chooses for another health insurance company, the previous insurance company is no longer allowed to use the data for sending bills.

Data repurposing

The example in the previous subsection can get more complicated. In case the same insurance company starts using the data to assess risks of patients in order to determine risk-based insurance premiums (eg, higher premiums for people at risk or showing unhealthy behaviour like smoking, not exercising, etc. and lower premiums for people at low risks showing healthy behaviour), they are reusing the data for a different purpose. This is a form of data reuse that we classify as data repurposing.

Data repurposing happens a lot, since data are used for many purposes. Addresses may not only be used for billing purposes but also for advertisements. Data may be combined to find new groups of potential customers, for assessing credit scores, or for assessing medical risks. Using Big Data, all kinds of new insights and predictions can be made about people. Sometimes, such data analysis may even reveal information about people they did not know themselves, such as the risks they run to attract specific forms of cancer or their life expectancy.

It is important to note that data repurposing in a general sense has another meaning than when used in legal terms. In a general sense, data repurposing can be understood as using the same data for several different purposes. From a legal perspective, EU data protection directive 95/46/EC focuses on data repurposing in Article 6.1(b): personal data must be ‘collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes’. The principle that the purposes for which personal data are collected should be specified in advance and that the data may only be used for these purposes is called the purpose specification principle. The principle that personal data should not be disclosed, made available, or otherwise used for purposes other than those specified, except with the consent of the data subject or by the authority of law is called the use limitation principle.

In other words, for data repurposing typically (additional) informed consent of the data subjects has to be obtained or another legal basis (for instance, specific legislation such as criminal investigation laws that allow the use of personal data to help to solve crime) has to be available.


36 Loshin makes the distinction between data reuse and data repurposing, see D Loshin, The Practitioner’s Guide to Data Quality Improvement (Burlington, MA: Morgan Kaufmann OMG Press 2011).

37 These and other principles for the fair processing of personal data, often referred to as the “privacy principles”, were drafted by the OECD in 1980.

38 For a complete overview, see <http://oecdprivacy.org/> accessed 27 July 2015.

39 See Article 5 of Directive 95/46/EC on the Protection of Personal Data stating that each member state has to determine more precisely the conditions under which the processing of persona data is lawful. Note that also anonymization of data is a kind of further processing, see WP29 (2014) Opinion 05/2014 on Anonymisation Techniques, Article 29 Data Protection Working Party, 0829/14/EN/WP216 <http://www.cnpd.public.lu/fr/publications/groupe-art29/wp216_en.pdf> accessed 23 December 2015.
Data recontextualization

When the health insurance company in the examples above starts selling the data, other companies may also make use of the data, for instance, for marketing their products to particular target groups. The data are then reused in a (sometimes completely) different context. This may cause issues of contextual integrity, since data may have a different meaning or may be interpreted differently in another context. For instance, health data may be interpreted differently by a physician than by a health insurance company. This is a form of data reuse that we classify as data recontextualization.

From a legal perspective, there is no real difference between data repurposing and data recontextualization. Both types of data reuse are usually referred to as function creep and are not allowed unless there is a legal basis for it. We introduce the distinction, however, because data recontextualization may bring along different legal issues, for instance, regarding the expectations that data subjects may have regarding the ways in which their data are used (they may know and trust the health insurance company but may not know or trust the marketing company buying their data) and the ways in which they are able to exercise their rights (they may know which data they disclosed to the health insurance company, but may not be aware who has bought their data). In short, when data are used in a new context, the ‘distance’ between the data subject and the data controller increases and the awareness of the available personal data and for which purposes the personal data is used may decrease among data subjects. As a result, it may become more difficult for data subjects to exercise their rights. Furthermore, the likelihood of interpretation errors may also increase.

When personal data are transferred to a third party and reused, data subject rights like the right to information and transparency do not cease to apply. On the contrary, at this point, it becomes even more important that privacy policies nowadays contain information overload and lack a meaningful choice for the users, which leads to the situation where data subjects no longer make informed decisions but simply consent whenever they are asked to do so. Not only are data subjects unaware how their information is processed, and under which conditions, they also lack basic understanding of whether their data can be and will be reused.

In practice, neither of the two options is frequently used. Describing future uses of personal data may be difficult since it is often unknown for which other purposes collected data may prove to be useful a few years later. Sometimes, new tools for data analyses may also yield new insights like novel patterns and relations in datasets. Not collecting data that can easily be collected is increasingly regarded as a waste of economic resources. For these reasons, many corporations collecting personal data formulate the purposes of their data collection very broadly, so that concrete purposes do not necessarily have to be known at the time of collection.

As a result, there are differences in the legal interpretation of repurposing or function creep and the more common understanding of it by data subjects. As mentioned above, from a legal perspective data repurposing or function creep only exists in case the secondary use of personal data goes beyond the purposes specified in advance. Since these are in general formulated very broadly, this may rarely be the case from a personal data. There are two options how to ensure data subject’s awareness. First, data reuse activities that might happen in the future are described and communicated to the data subject before his personal data are collected. Secondly, the data subject renews his consent every time before the data are reused for a new purpose. Both tactics often prove to be difficult to apply. In the first case, it is hard to predict all the purposes for data reuse that may appear in the future. In the second case, it is almost impossible to get in touch with all data subjects and to secure their valid consent. Research has shown that privacy policies nowadays contain information overload and lack a meaningful choice for the users, which leads to the situation where data subjects no longer make informed decisions but simply consent whenever they are asked to do so.

For instance, IT company Cisco states at <http://www.cisco.com/web/siteassets/legal/privacy.html> (accessed 27 July 2015): “We collect personal information for a variety of reasons, such as processing your order, providing you with a newsletter subscription, personalizing your experience, or in connection with a job application. We will inform you of the purpose for collecting personal information when we collect it from you and keep it to fulfill the purposes for which it was collected or as required by applicable laws or for legitimate purposes. We may combine the information we collect from you with information obtained from other sources to help us improve its overall accuracy and completeness, and to help us better tailor our interactions with you. We may also collect information relating to your use of our websites through the use of various technologies, including cookies.”
legal perspective. From a more common understanding, however, people may consider function creep to be the case when personal data are used for purposes they did not expect their data to be used for.

Literature on expectations of data subjects on how their data are used and what they consider acceptable shows that there is a considerable gap between the practices of data controllers and the expectations of data subjects. For instance, US-based research has shown that race and ethnicity play an influential role in how people use social media and share personal information. Users show concern for privacy, although there seems to be an incongruity between public opinion and public behaviour: people tend to express concern about privacy, but when asked about it, they routinely disclose personal information because of convenience, discounts, and other incentives, or a lack of understanding of the consequences. These tensions between attitudes and practices were also found by Acquisti and Gross (2006). A possible explanation for this tension may be that users do not connect the disclosure of their data at a particular time with the use of their data later, for different purposes, by different organizations and in a different context. Such connections may not be transparent when there are long periods of time between the data collection and actions based upon the processing or sharing of such information and when data are sold to other companies or between the (primary) data use and the (secondary) data reuse. Also, the ways in which data are processed may not be transparent. For instance, when the information collected is used for profiling, such profiling techniques, by their nature, tend not to be visible processes for data subjects.

Data reuse from the data subject’s perspective

Above we distinguished several ways in which data controllers like to reuse personal data they have collected. However, data subjects themselves may also like to use and reuse their own data. Below we discuss data sharing, data portability, and the right to be forgotten.

Data sharing

The current models for personal data protection law are based on the concept of informational self-determination. This concept, introduced by Alan Westin, is based on data subjects that are considered to be autonomous human beings that control their own lives including their personal data. Westin describes informational self-determination in terms of control over information, i.e., as a person’s right to determine for himself when, how, and to what extent information about him is communicated to others. Apart from some situations in which people are obliged to provide personal information (for instance, for taxation or in criminal investigation procedures), data subjects are therefore the major locus of control for disclosing personal data.

People may choose to disclose their personal data for different reasons, including purchasing online products or services, financial profit (getting paid for personal data or receiving reductions), and building online reputations. We will call this type of disclosure of personal data data sharing or data disclosure. Data sharing relates to specific individuals (sharing data with Alice and Bob or with your health insurance company), whereas data disclosure may also be a more general and undirected form of data sharing (eg, putting personal data on the Internet without knowing who actually visits the website). Whatever reasons they may have, people may prefer to share or disclose personal data in a specific context or for specific purposes but may not prefer to share or disclose personal data in another context. Typical examples may be that people are more likely to share their medical data with their physician than with their employer, people may be more likely to discuss...
their sex life with their partner then with their parents and people may be more likely to show responsible behaviour when their children are around then when they are on a bachelor party. Every time data subjects choose to share or disclose their personal data, this is data use. If they share their data with other people or in a new context or both, this is data reuse.

Furthermore, people may change their mind. They may consent to the use of their personal data (data use), for instance, on Facebook or Twitter, but change their mind and revoke their consent. They may change their mind later and provide consent again (data reuse or data use in case new data is provided). Data subjects may change their mind arbitrarily (they have a right to do so), but it may also be because they took a different perspective on things or because the data controllers changed their terms and conditions or privacy policies.

Data sharing does not imply that a data subject surrenders his or her data subjects rights. When you sell your car, you no longer own your car. With personal data this is different, since each time a data subject chooses to share or disclose personal data, he still owns a version of the data (although it is difficult to dub this as the ‘original data’). Data sharing does not affect a data subject’s rights regarding the control over his personal data. However, as discussed above, in practice it may imply that it is less transparent for a data subject which data controllers process his personal data and for which purposes.

As discussed above, personal data protection legislation requires that people are properly informed about the purposes of personal data processing when asked for their (informed) consent. Data reuse in the context of Big Data may be complicated as it is difficult to inform people about complex types of data analyses, let alone about forms of data processing and knowledge discovery that do not yet exist.

Data portability

In the proposal for General Data Protection Regulation, data portability is defined as the right to obtain from the data controller a copy of the personal data undergoing processing in an electronic and structured format that is commonly used and allows for further use by the data subject (Article 18). From an individual’s point of view, data portability is seen as a safeguard to his information self-determination. It is a way to avoid consumer lock-in, to opt for a more secure or more developed provider and to take full control over personal data. In other words, data portability is the ability for people to reuse their data across devices and services. This reasoning brings us to the conclusion that data portability is in fact data reuse per se and fits well into our data reuse taxonomy.

In 2012, the European Commission kicked off the data protection reform by publishing its proposal of the EU regulation on personal data protection. The objective of the new legal framework was to strengthen data protection, unify national legislation across member states, and adapt to the changed circumstances in a globalized and intra-connected world. This was the first time that a requirement on data portability was included in data protection legislation. In March 2014, after intensive negotiations in the Parliament, the initial Commission’s proposal was significantly amended and its initial formulation was softened. The Commission, Parliament, and the Council have recently entered the trilogue process, the final stage of negotiations, which will result in a finalized text of the Regulation. As the outcomes are difficult to predict, our analysis mostly builds on the Commission’s proposal, which has been so far the only complete, publicly shared draft.

Based on the Commission’s text, the right to data portability can be split into two elements. Firstly, under Article 18(1) of the draft Regulation, individuals whose personal data are processed electronically and in a ‘structured and

56 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) 11 final, 25 January 2012.
57 The European Commission’s definition of data portability focusses on the human side of the right, although Commissioner Almunia has clearly acknowledged that data portability is also a measure of competition law. Graef, J Verschakelen, and P Valcke, ‘Putting the Right to Data Portability into a Competition Law Perspective’ (2013) J Higher School Econ Ann Rev, 53, 63.
59 Besides the arguments related to antitrust and human rights, data portability should also be promoted due to its positive influence to data interoperability. Standardized data format would boost innovation and growth; however, many authors warn this is not easily achievable. See more: P Swire and Y Lagos, ‘Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique’ (2013) 72 Md L Rev 335.
61 See above n 56.
62 See, for example, Swire and Lagos, n 59.
commonly used format’ are given the right to obtain a copy of that data for further use.\(^6^5\) Secondly, Article 18(2) provides individuals the right to transmit their personal data from one provider to another.\(^6^6\) The article applies generally to all types of electronic data processing, including cloud computing, web services, smartphone apps, etc.\(^6^7\) As mentioned in the recitals, the idea of data portability was introduced due to alleged lock-ins in case of the social networks.\(^6^8\)

The right to data portability differs substantially from the pre-existing right to access.\(^6^9\) It extends beyond mere accessibility and emphasizes the importance of further data use and data reuse. In other words, data portability transforms passive data subjects into active reusers. A right to data portability empowers consumers to take advantage of value-added services from third parties and lets them ‘share the wealth’ created by big data.\(^7^0\) The latter will not always be easy, however, given the barriers caused by non-interoperability and the complexity of big data. EU data protection regulation is indeed breaking new ground with the data portability requirement. No jurisdiction has experimented with anything resembling the proposed Article 18.\(^7^1\) However, the fact that this solution has not been proved adds to legal uncertainty and concerns data controllers.\(^7^2\)

The right to be forgotten

The right to be forgotten is a crystallization of the more fundamental wish for ‘control’ over one’s personal data and another safeguard of information self-determination.\(^7^3\) Contrary to data sharing and data portability, which are described above as user controls, the right to be forgotten does not imply data reuse but rather deals with blocking all further, secondary uses.

The right to be forgotten is a manifestation of the right to oblivion in the digital age. Originally, the right to oblivion was introduced to be invoked in cases where undesired public exposure is given to a person’s past, as a shield against disproportionate intrusion by mainstream media (papers, news broadcasts, radio plays, etc.) into the private life of people who have entered into the public eye.\(^7^4\) The right to be forgotten has no such tradition or connotation,\(^7^5\) since it primarily aims at protecting an individual’s digital reputation, but the fundamental interests it safeguards are similar.\(^2^6\) As Werro establishes, the right to be forgotten ensures that someone can preclude others from identifying her in relation to her (criminal) past.\(^7^7\) What is important is that Werro’s definition focuses not so much on deletion of data, but rather on regulating (blocking) the (re)use of data.\(^7^8\)

In 2012, when the EU Commission came up with the data protection reform, it proposed the right to be forgotten as an independent right. This has been one of the most attention-grabbing parts of the Commission’s proposal, although it falls short from being a new legal concept, like the right to data portability.\(^7^9\) The EU Data Protection Directive from 1995 already included the principles underpinning the right to be forgotten,\(^8^0\) but

\(^{65}\) Only data that are processed in a structured and commonly used format fall under the definition of data portability. This implies the controllers are able to avoid data portability requests if data would be processed in a non-structured and uncommonly used format. The Council has noticed this downside of the Commission’s proposal and has avoided the original formulation in its version. See the comparison of the Parliament and Council text on the General Data Protection Regulation <https://edri.org/files/EP_Council_Comparison.pdf> accessed on 27 July 2015.

\(^{66}\) In the amended text of the European Parliament, both aspects are joined in the same section of the article on the right to access, namely Article 15(2a).

\(^{67}\) See above n 63.

\(^{68}\) While social networking sites like Facebook and Google+-offer users the possibility to obtain a copy of their data, there are still considerable limits on the direct transfer of personal information to other platforms. Moreover, social network providers do not allow third-party sites to directly acquire the user’s information. For instance, Facebook blocks Google Chrome’s extension for exporting friends. Graef and others n 57. See also G Zanfir, ‘The Right to Data Portability in the Context of the EU Data Protection Reform’, (11 May 2012) IDPL 11.

\(^{69}\) See above n 63.


\(^{71}\) See above n 63.

\(^{72}\) The high level of uncertainty regarding the right to data portability was confirmed by the London Economics’ study conducted for ICO. See more <https://ico.org.uk/media/about-the-ico/documents/1042341/implications-european-commissions-proposal-general-data-protection-regulation-for-business.pdf> accessed 23 December 2015.


\(^{75}\) Ibid.

\(^{76}\) G Finocchiaro and A Ricci, ‘Quality of Information, the Right to Oblivion and Digital Reputation’ in B Custers, T Calders, B Schermer, and T Zarsky (eds), Discrimination and Privacy in the Information Society, Data Mining and Profiling in Large Databases (Springer Verlag, Berlin, Heidelberg 2013).


\(^{79}\) In the landmark case Case 131/12 Google Spain v AEPD and Mario Costeja Gonzales issued on May 13 2014, the Court of Justice of the European Union made it clear that the right to be forgotten is encompassed in the fundamental right to privacy. See also C Kuner, The Court of Justice of the EU Judgment on Data Protection and Internet Search Engines’ (2014) LSE Legal Studies Working Paper No. 3/2015.

\(^{80}\) De lege lata the right to be forgotten has been reflected in the right to objection and deletion. Factsheet on the ‘Right to be forgotten’ ruling C-131/12, European Commission (2014).
the proposed legal framework reaffirms and reshapes them to suit the modern information society better.

The formulation of the right to be forgotten in the proposal emphasized the importance of consent and the purpose limitation principle. Article 17 (1) explicitly allows data subjects to seek the deletion of data and block further reuse when the consent is withdrawn or when the data are no longer necessary in relation to the purposes for which they were collected or otherwise processed. Article 17 (2) goes even further, proposing that the right to be forgotten should follow the data when the data controller has made it public (eg, by publishing it on a website) or when publication is delegated to a third party. In the first scenario, the original controller only has to take ‘all reasonable steps’ to inform third parties about the data subject’s request for erasure. In the second situation, the original controller will be considered responsible in any case. This formulation radically shifts the burden of proof—it is now for the data controller and not for the individual to prove that the data cannot be deleted because it is still needed or is still relevant. Article 17 provides for a number of exceptions to the general rule—if there are competing interests such as freedom of exception or various legal obligation, the right to be forgotten cannot be enforced.

According to Van Hoboken (2013), the added value of the updated provision for data subjects that want to see their data deleted is relatively minor. Kuner (2012), on the contrary, believes the new provision is a significant one, in particular its reversed side, the duty of the controller to inform third parties about the data subject’s request to erase data. After realizing how little it took the CJEU to formulate the right to be forgotten from the existing provisions, we agree with Van Hoboken that the new provision is anything but a revolution. However, we do acknowledge that the use of clearer language in the proposal might add to better compliance and more legal certainty.

The Commission’s proposal emphasizes the importance of empowering a data subject with adequate tools to govern not only his personal data use but also reuse (eg, through the controller’s obligation towards third parties in Article 17 (2)). The idea of the right to be forgotten as the regulation (blocking) of data (re)use is what distinguishes it from the right to erasure and right to objection as stipulated in the current directive. In addition, this is how the right to be forgotten aligns the legal framework with the realities of our Internet era.

However, many dilemmas have remained unsolved. The socio-technical context of Big Data implies that data processing is based on vague purpose definitions to allow unforeseen future uses and that data are increasingly used for secondary purposes. This fundamentally challenges not only the purpose limitation principle itself but also the effectiveness of a right to be forgotten. For example, in an increasingly personalized Internet almost every bit of personal data can be argued to be relevant, and it will be hard to establish that the data should be forgotten on the ground of ‘no longer being necessary for the purpose for which it was initially collected’.

**Overview and discussion of the taxonomy**

Summarizing this section, the resulting taxonomy for data reuse is as follows. From a data controller’s perspective, the major types of data reuse are as follows:

- Data recycling—using data several times for the same purpose,
- Data repurposing—using data for different purposes than for which they were initially collected,
- Data recontextualization—using data in another context than in which they were initially collected.

The second and third meanings of data reuse are expected to be of most added value in the European data economy. The added value of Big Data and open data is likely to increase in large, aggregated datasets in which data from different sources (eg, from different social sectors and industries) are combined.

The distinction between several types of data reuse from a data controller’s perspective may be useful for further interpretation of the privacy principles, particularly the use limitation principle. Current legislation does not distinguish between data repurposing and data recontextualization. It may be argued that the use limitation principle is, in a way, not specific enough. Even though the use limitation principle restricts the use of personal data for purposes other than those specified,
and thus addresses function creep, we think there is a major difference (particularly from a data subject’s rights perspective) between reuse in the same context or in a different context (for instance, when a data controller sells the personal data to other, and when personal data are transferred to other countries). Consequently, it may be argued that asking consent for data reuse should be more much more explicit in cases of data recontextualization than in cases of data repurposing. Asking consent for data repurposing, in turn, should be more explicit than consent for data recycling, for which consent may be assumed in most cases.

From a data subject’s perspective, the major types of data reuse are as follows:

- Data sharing or data disclosure—data subjects have the ability to (directly) allow use and reuse of their personal data,
- Data portability—data subjects have the ability to use and reuse their personal data across devices and services,
- The right to be forgotten—data subjects have the ability to block data use and reuse.

Strictly speaking, the third item (the right to be forgotten) is not a type of data reuse, but a right to block data reuse. As such, data sharing/data disclosure and data portability are all ways to reuse data and may promote Big Data, whereas the right to be forgotten is an instrument that has exactly the opposite effect, i.e., it intends to limit the use of Big Data.

The differentiation of data reuse from a data controller’s perspective is of more obvious use (both from a practical and technological perspective and for the further interpretation of privacy principles) than the differentiation from a data subject’s perspective. In our opinion, this results from the fact that most privacy principles address data controllers regarding what they may or may not do with personal data. This assumes that restricting data processing for data controllers guarantees better privacy and personal data protection of data subjects. This assumption seems to be controversial in the era of Big Data, in which many Big Data firms have sound privacy policies (from a legal perspective), but data subjects still have many privacy concerns (from a social perspective). Since the privacy principles are intended to protect the personal data of data subjects, we think it is important to take the needs, interests, and preferences of data subjects into account more explicitly when it comes to data reuse. In some cases, data subjects may object to data reuse, even though it may be allowed from a legal perspective. In other cases, data subjects may not object to data reuse or actually prefer data reuse, even though it may not be allowed from a legal perspective.

**Conclusions**

The benefits of Big Data can only be fully exploited when it is possible to combine large amounts of data from different sources. The possibilities of discovering novel trends, patterns, and relationships may yield economically and socially valuable insights and knowledge. The reuse of data is an important factor when it comes to materializing the benefits of Big Data. The developments in the area of Big Data call for new technological models (regarding standardization and adequate IT infrastructure), new economic models (regarding corporate secrecy and IP rights), new social models (regarding public support), and new legal models (regarding personal data protection) in which the reuse of data is encouraged rather than hindered. In this article, we focused on the legal perspective, particularly EU personal data protection regulations, in which requirements like data minimization and purpose specification are potentially inimical to Big Data as they limit the size and use of Big Data.

In this article, we have investigated different forms of data reuse, from both the data controller’s and the data subject’s perspectives. From the data controller’s perspective, data recycling, data repurposing, and data recontextualization can be distinguished. Particularly data repurposing and data recontextualization may be forms of data reuse that are expected to be of most added value in the European data economy. From a legal perspective, data repurposing and data recontextualization are both considered a form of function creep that is, in general, not allowed. In practice, this is bypassed by data controllers by either drafting broad purpose specifications for data use in their terms and conditions or privacy policies or by simply ignoring the need to obtain proper consent or a valid legal basis for data processing. From a data subject’s perspective, data sharing and data portability can be distinguished as forms of data processing in which data subjects may want their personal data to be reused. Also relevant is the right to be forgotten, which enables blocking the use and reuse of personal data.

Considering that there exist many reasons for encouraging data reuse but also for adequately preserving privacy, it is important to strive for a balance. The suggested taxonomy may be helpful to distinguish different types of data reuse and modify the extent of reuse that is

---

89 See above n 44.
allowed. We argue that forms of data reuse that stay close to the awareness and intentions of data subjects should be approached less restrictively (for instance, by assuming informed consent), whereas for forms of data reuse that are ‘at a distance’, i.e., in which awareness and transparency may be lacking and data subject’s rights may prove more difficult to exercise, more restrictions, and additional protection should be considered (for instance, by requiring explicit consent). In other word, forms of data reuse like data recycling, data sharing, and data portability should have little or no restrictions (for instance, implicit consent may be assumed), whereas forms of data reuse like data repurposing and data recontextualization should have more restrictions and further protection should be considered (including explicit consent, etc.).

When considering data repurposing versus data recontextualization, it may be concluded that the legal conditions are the same, since both forms of data reuse are often included in the same terms and conditions or privacy policies of data controllers. However, for data subjects exercising their rights may be much more difficult in the case of data recontextualization, since in practice this may involve the sale or transfer of personal data to other companies. We argue that a legal distinction between these types of data reuse, including additional protection for data subjects in the case of data recontextualization, would better reflect the ‘distance’ between the data subject and the data controller and the awareness of data subjects regarding the availability of their personal data and for which purposes these are used and reused. Additional protection for data subjects in the case of data recontextualization may be justified because the likeliness of interpretation errors may also increase and because it may be more difficult for data subjects to exercise their rights.