Advancing the EU Data Economy: Conditions for Realizing the Full Potential of Data Reuse

Custers, B.H.M.; Bachlechner, D.

Advancing the EU data economy: Conditions for realizing the full potential of data reuse

Bart Custers[a,*] and Daniel Bachlechner[b]
[a]*eLAW, Center for Law and Digital Technologies, Leiden University, Leiden, The Netherland*
[b]*Fraunhofer ISI, Germany*

Abstract. The rise of big data promises a plethora of opportunities for a data-driven economy, including accelerated innovation and growth, and increased productivity and competitiveness. However, there are still many legal, technological, societal and economic barriers that inhibit the use and reuse of data. It's not enough to have large volumes of data to capitalize on the benefits of big data. Data, for instance, must also be accessible, reliable and up-to-date, and there must be clarity about the provenience of data and the right business model. Companies can collect data themselves, but with respect to many use cases, it seems to be more reasonable to reuse existing data, which was collected by third-parties, at least to some extent. However, this is only happening to a very limited extent. This paper presents and discusses conditions that are prerequisites to realize the full potential of data reuse. The viability of data reuse depends on (1) compliance with legislation, (2) technical possibilities, (3) public acceptance, and (4) business cases that yield added value. Data reuse and a thriving data economy rely on these conditions. Solutions are offered to policy makers and company leaders to facilitate data reuse and to advance the data economy.

Keywords: Big data, data economy, data reuse, privacy, public acceptance, standardization

1. Introduction

In its 2014 strategy on big data, the European Commission (EC) indicated how it intends to support and accelerate the transition towards a data-driven economy in the European Union (EU).[1] The EC aims to shape the data-driven economy, the data economy in short, by means of the European Free Flow of Data Initiative. The initiative is regarded as an important instrument to realize a digital single market strategy for the EU, one of the priorities of the EC [1]. The data economy is expected to yield accelerated innovation and growth as well as increased productivity and competitiveness. As such, the rise of big data promises a plethora of opportunities for the data economy. However, there are still many legal, technological, societal and economic barriers that inhibit the use and reuse of data.

Large volumes of data are needed to capitalize on the benefits of big data but large volumes alone are not sufficient. Data, for instance, must also be accessible, reliable and up-to-date. Moreover, there must be clarity about the provenience of data and the right business model. Companies can collect data themselves but with respect to many use cases it seems to be more reasonable to reuse existing data, which was collected by third-parties, at least to some extent. As this is currently only happening to a

---

very limited extent, the question arises how the reuse of data can be supported. This paper answers this question by presenting and discussing conditions that are prerequisites to realize the full potential of data reuse. The findings are based on more than two years of intense discourse with actors directly involved in or doing research on the EU data economy maintained within the scope of EuDEco,[2] a Coordination and Support Action funded by the EC under its Research and Innovation Framework Program Horizon 2020.[3] EuDEco has not only held numerous workshops focusing on data reuse co-located with major European events such as the European Data Forum, ICT and summits of the BDVA but has also been in contact with opinion leaders from industry and academia via social media. The key workshops featured not only plenary sessions but also breakout sessions focusing on legal, socio-economic and technological aspects of the data economy. The viability of data reuse seems to depend on (1) compliance with legislation, (2) technical possibilities, (3) public acceptance, and (4) business cases that yield added value. Data reuse and a thriving data economy rely on these conditions.

Once the conditions required for data reuse are understood, it becomes obvious where the major barriers to data reuse are. They may depend a lot on the specific situation. For instance, in some cases the main barriers may be legal issues, whereas in other cases the main barriers may be technological, societal or economic issues. In this paper, based on a thorough analysis of the conditions, potential solutions for both policy makers and company leaders are developed that may prove to have the potential to facilitate data reuse and advance the data economy.

This paper is structured as follows: In Section 2, it is explained why data, particularly big data, is considered the new oil. It is briefly defined what big data actually stands for and how it fuels the data economy. In Section 3, a taxonomy for data reuse is described that results from previous research. This taxonomy is necessary to distinguish different types of data reuse, such as data reuse for new purposes or in new contexts. In Section 4, the conditions required for data reuse to be viable are presented and discussed looking on the data economy from a legal, a technological, a societal and an economic angle. Each of the perspectives is examined in depth. In Section 5, potential solutions for policy makers and company leaders are offered. Section 6 summarizes the main conclusions.

2. Big data and the data economy

Although there is no well-established definition of big data, its most prominent characteristic is its sheer volume. That is why it is called big. There is no generally accepted definition of the size required for datasets to be called big data. It is obvious that big data can involve many terabytes or even petabytes of data [2]. But it is neither only volume nor the data itself that matters. In general, big data refers to data and analytics that involve the three Vs: volume, velocity and variety [3]. *Volume* refers to the large amounts of data and the fact that no samples are taken. *Velocity* refers to the fact that many data are real time or nearly real time. *Variety* refers to the different nature of various types of data, including text, numbers, images, videos and sound. These characteristics of data pose significant challenges to analytics. Typically, big data does not allow the human eye or human intuition to grasp the essence of the data and the patterns and relationships hidden in it. Such hidden knowledge can usually only be discovered using advanced methods for data analysis.

---

[2]http://data-reuse.eu.

[3]Grant Agreement No. 645244. This project is focused on addressing the question of reusing data from economic, technological, societal and legal perspectives.

Data is increasingly considered the new oil of the economy. Data has become a major raw material of production as well as a key source of economic and social value [4]. In other words, data fuels the economy, more specifically, the data economy. It reveals completely new opportunities. For instance, data enables creating and delivering a new range of products and services, and accessing new groups of customers. It is expected that big data will accelerate these developments [5]. Big data will become a key element of competition underpinning new waves of productivity growth, innovation and consumer surplus [6].

All this sounds promising, but concrete examples of big data applications and the big data revolution are not abundant. As not all large datasets qualify for big data applications, claims should be evaluated critically. However, particularly large companies in all types of industries are dealing with big data and making or increasing their profit with it. Examples include retail organizations such as Tesco, Internet companies such as Google and Facebook, banks and credit card companies such as BNP Paribas and MasterCard, insurance companies such as AXA and Allianz, and navigation and mapping product producers such as TomTom and Garmin. Also an entire industry has grown around offering services for big data analytics, including companies such as Acxiom and Palantir.

It seems only natural that large companies also have the large databases as well as the skills and technologies needed for big data analytics. However, when each company collects its own data, the potential of the data economy is far from being fully realized. Such a data economy does not only exclude certain potential participants such as small and medium-sized enterprises (SMEs) but also leaves advantages of specialization unused by being hesitant with respect to making data available to third parties and using third-party data. All companies should be able to build businesses and applications on big data, even when they do not have large datasets themselves [7]. Moreover, non-ICT companies should be able to profit from big data developments too. This is why the EC tries to also get SMEs and non-ICT companies involved in big data developments [8]. However, a lot of data is still not easy to obtain for companies, particularly, for SMEs. In most cases, this stands in close connection with the economic value of data. However, the problem is not necessarily that the value assigned by the data owner, and thus the price, is too high but rather that, due to difficulties with respect to the determination of an adequate price, data owners refrain from making them available.

Things are changing, however. Traditional data sources such as company databases and applications are now increasingly complemented by non-traditional sources such as social media or sensors embedded in physical world devices including mobile devices, smart meters, cars and industrial machines. Although the potential benefits of data reuse are widely undisputed, data is currently leveraged only to a very limited extent by others than the ones who produced it. The data market, where it exists at all, seems to be underdeveloped.

The presentation and discussion of conditions for data reuse can contribute to the further development of the data economy as it helps to identify and remove barriers to data reuse. After removing the main barriers, the opportunities provided by big data can be better leveraged. The content of this paper is useful for policy makers and business leaders embarking on big data projects, as it scrutinizes barriers for data reuse from different perspectives. Relevant and meaningful solutions are created based on this scrutiny.

3. Data reuse

In this section, we first explain why a focus on data reuse is important. Then, we present a taxonomy for different types of data reuse that builds upon prior research. Differentiating between several types of

data reuse is important, as they may entail different types of legal, technological, societal and economic issues.

3.1. Focus on data reuse

Large volumes of data are generated by people, for instance, via the use of social media, and by technology via sensors embedded in physical world devices. However, the data needed to leverage benefits do not necessarily have to be generated again and again. To some extent, it may be possible to reuse existing data, a source that is sometimes overlooked, to complement or substitute new data. We think that data reuse is at least in some cases a more reasonable source of data than generating new data. A first argument is that data that was generated once, contrary to tangible raw materials, can be reused many times.[4] This argument, we call the *cumulation argument*. Second, collective efforts create volumes of data that cannot easily be reached on an individual level. For single persons, organizations and sensor-equipped devices, it may be impossible to generate and store the same breadth of data as collectives of persons, organizations and devices may be able to. Of course, collective efforts can generate new data only when the data is used for new purposes or in new contexts, this is by definition a form of data reuse (see next section). This argument, we call the *aggregation argument*. A third argument is that data reuse allows going back in time, whereas generating new data can only start today or in the future. The result is that data reuse can build on and expand existing datasets. Therefore, once newly generated data is reused, it is included in the collection of reused data. This argument, we call the *growth argument*.

Of course, we do not underestimate the value of generating new data, which is indispensable in many big data applications. Companies should be aware of the options they have and evaluate which option is best for their specific situation. In this way, well-founded 'make or buy' decisions can be made. Also, combining generating new data and reusing existing data may be interesting for companies in some situations. In the era of big data, in which large volumes of data are created at a fast rate, it is important to note that some parts of such data (both new and existing data) lose their value rapidly. When data is abundant and ubiquitous, many parts will be volatile. An important development is that a lot of data is streaming data, implying that the data is not stored at the receiving end of data transfers. This effectively disables data reuse.[5] However, in such cases it may still be that the data is stored at the sending end of data transfers. Furthermore, new data is generated at exponentially increasing rates. This may imply that the amounts of new data are in fact larger than the amounts of data that are potentially available for data reuse.

Taking all these considerations into account, it can be argued that an important case can be made for focusing on data reuse. Even though several legal, technological, societal and economic issues need to be resolved, the reuse of existing data offers considerable opportunities.

3.2. A taxonomy for data reuse

Before discussing conditions for data reuse, it is important to explain what is meant by it. Understanding data reuse requires an understanding of what data use is. Data reuse (secondary use) can only

---

[4]If data reuse would be without limits, the generation of new data would have to grow exponentially over time if it were to keep up in size with data reuse.

[5]It is not uncommon that companies go for a hybrid approach where they maintain a real-time and a batch layer. Data is first processed by a streaming data platform to extract real-time insights and then persisted into a store where it can be transformed and loaded for batch processing use cases.

happen after data use (primary use). Based on previous research, this section provides a brief taxonomy for different types of data reuse [9].

Data use is understood as using data for a specific purpose within a specific context. This general perception of data use is more or less the same from a legal, technological, societal or economic perspective. However, these perspectives may have different interpretations of data use when looking at it more closely. For instance, from a technological perspective, collecting and storing data is considered something that takes place before data can be used. Also deleting data, from a technological perspective, would not be a form of data use, but rather something that takes place after data have been used. However, from a legal perspective, the collection, storage and erasure or destruction of data are considered as forms of data use, at least for personal data in the EU. EU Directive 95/46/EC on the protection of personal data does not explicitly deal with the concepts of data use and data reuse, but uses the concept of processing of personal data, 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." Hence, processing of data includes use, but is a much broader concept that also includes collection, storage and erasure or destruction.

Three types of data reuse can be distinguished, depending on whether the purposes or the context or both differ from the original purpose and context for which data were collected and used [9]:

– *Data recycling* – using data several times for the same purpose in the same context
– *Data repurposing* – using data for different purposes than for which they were initially collected, but still in the same context as the original purpose
– *Data recontextualization* – using data in another context than for which they were initially collected

These different types of data reuse can well be illustrated by an example. Many online stores require that customers register personal data such as name, address and financial data the first time they order something. When a customer wants to order something the next time, he or she logs in and the personal data automatically appear. This is called data recycling as the personal data are used for the same purpose in the same context (though not for the same occasion). When the same online store starts to use the data to build customer profiles, for instance, for the purpose of assessing credit default risks or for estimating customer preferences, this is data reuse beyond the original purpose. This is called data repurposing. When the same online store starts selling the data, for instance, to companies in other sectors that may be interested in the data for direct marketing purposes, the data may be reused in a (sometimes completely) different context. This is called data recontextualization.

4. A multi-perspective view: Appropriate conditions for data reuse

In this section, a multi-perspective view on conditions required for data reuse is presented. A legal, a technological, a societal and an economic perspective are taken. The discourse with data economy actors suggests that the viability of data reuse depends on (1) compliance with legislation, (2) technical possibilities, (3) public acceptance, and (4) business cases that yield added value. Data reuse seems to rely on these conditions. To illustrate the approach, for each of the four perspectives, exemplary questions and critical aspects are shown in Fig. 1.
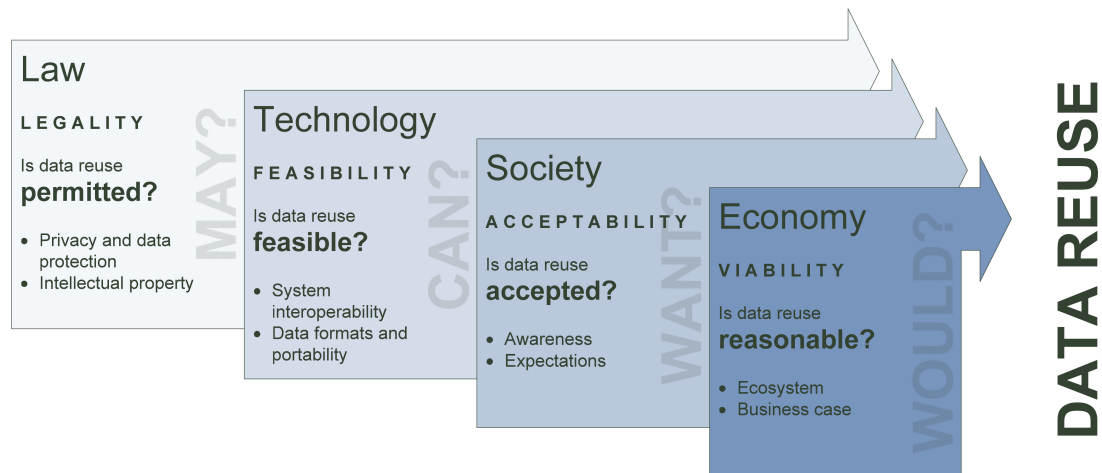
Fig. 1. Exemplary questions and critical aspects for each of the four perspectives.

Our approach implies that there are four different domains in which conditions for data reuse need to be met. These four perspectives in this paper are loosely based on [10].[6] The legal domain focuses on the *may question*. Is data reuse legally permitted? The technological domain focuses on the *can question*. Is data reuse technically feasible? The societal domain focuses the *want question*. Is data reuse socially accepted? Finally, the economic domain focuses on the *would question*. Is data reuse economically reasonable?

When one or more conditions are not met, the climate for data reuse is unfavorable and data reuse is unlikely to take place. It may be that a company has calculated a business case in which data reuse promises to add value or reduce costs, that social acceptance is likely, and that the required technological components are available and compatible. If the form of data reuse envisaged is not in compliance with the law, the conditions for data reuse are not sufficiently met. Similarly, it may be that there are no legal objections to the envisaged form of data reuse, that there is no social resistance to be expected and that there is a viable business case. If the technologies used by the parties involved are so different that they do not allow exchanging data, there will be no data reuse. The outlined situations show that it is necessary that the conditions are met for all four perspectives. For data reuse to happen, it must be legal, technically feasible, socially accepted and there must be a viable business case. Otherwise, nobody will take the initiative for data reuse and, as a result, there will be no data reuse. Below, the different perspectives are discussed in more detail.

4.1. The legal perspective

From a legal perspective, there are several reasons why data cannot be reused. The legal perspective for data reuse involves testing whether data reuse is allowed. A comprehensive overview of legal barriers and enablers to data reuse in the EU can be found in literature [13]. The main legal barriers to data reuse are rooted in privacy and data protection law as well as in intellectual property law. These are discussed below.

---

[6]Note that these perspectives may interact with each other. For instance, law may regulate technologies, but technologies may also be designed in ways that enforce or nudge specific behavior. See, for instance, [11,12].

Privacy and data protection

"We cannot share data because of privacy." This is a statement that is often heard. Knowledge is power and some people and organizations refuse to share their data because of this. However, in many cases, this is the exercise of power rather than the result of a legal analysis. When looking at privacy and the protection of personal data in the EU, data can be shared on several legal grounds (Art. 7 of EU Directive 95/46/EC). The most important legal basis is (unambiguous) consent of the data subject whose personal data is concerned. Other reasons are performing a contract to which the data subject is or wants to be a party, complying with a legal obligation, protecting a vital interest of the data subject, performing a task for the public interest or official authorities and serving the legitimate interests of the data controller as far as these do not interfere with the interests and rights of the data subject. As such, when data subjects consented to a contract with a data processor, for instance, by agreeing with the general terms and conditions or with the privacy policy, they usually agree with data reuse in terms of data recycling. When the terms and conditions include provisions that the personal data can also be used for other purposes, data subjects consent to data repurposing. When also provisions are included on selling the data, this may imply consent for data recontextualisation.

Obviously, a major issue here is that people do not read terms and conditions and privacy policies. It may take a lot of time to read privacy policies and to make a decision based on this information. If people would actually read all the privacy policies presented to them, it would take them 244 hours annually [14]. Even when people read these lengthy documents, they may not understand them. In many situations, the text is highly legalistic in nature or contains technical details beyond the comprehension of the average user. Finally, even when people actually read and understand the policies, they may not contain the preferred option and just offer a take-it-or-leave-it proposition [15].

From the perspective of EU data protection law, the purposes for which personal data are used should be specified, explicit and legitimate and not further processed (i.e., reused) in a way incompatible with those purposes (Art. 6.1b of Directive 95/46/EC). Stating explicit purposes is also referred to as the *purpose specification principle* and limiting the use of data to these purposes is referred to as the *use limitation principle*. The purposes of data processing are usually specified in terms and conditions or in privacy policies. In principle, the use limitation principle blocks any forms of data reuse except data recycling. However, many purposes are formulated very broadly and may include a whole range of purposes and even reuse of data in different contexts, for instance, when data subjects consent to selling their data to third parties.

In summary, from the perspective of privacy and data protection law, legislation does not block data recycling but it may block data repurposing and data recontextualisation. However, in practice, data repurposing and data recontextualisation are also included in the broad consent requests that people agree with via accepting terms and conditions and privacy policies. As such, privacy and data protection law are not always the showstopper for the reuse of personal data. In fact, in many cases, data protection law does not hinder data reuse as long as consent of data subjects is ensured. Obtaining this consent is not a big issue in many cases, although the value of consent may be controversial when people do not read or understand the policies. If the goal is to increase the possibilities for data reuse, the main challenge from a legal perspective is to ensure that consent requests represent a fair balance between broad purposes (enabling different types of data reuse) and specific purposes (ensuring that data reuse is in line with the preferences of data subjects).

Intellectual property

The other major legal area that may cause a barrier for data reuse is intellectual property (IP). IP refers to creations of the mind, including artistic works such as music and literature, but also discoveries and

inventions. IP law grants owners of IP some exclusive rights, called IP rights or IPRs. The most common IPRs are copyrights, database rights and trade secrets. Patents and trademarks may also provide protection in this respect but these do generally not relate to the actual data but rather to data products [16]. IPRs often represent (potential) monetary value. The main objective of IPRs is to promote progress. The basic idea is that those who invest in discoveries, inventions or creating works of art should also have the benefits of it. When companies have invested a lot of resources in obtaining large datasets, it is obvious that they want to extract the hidden knowledge and capitalize on the potential value of the data themselves. IPRs help them to protect their investments.

Within the EU, each country has its own IP laws. However, there are some international treaties. One of the first is the Convention of Paris in 1883, which was signed by 176 countries worldwide and is still in force.[7] This convention states a so-called priority right: the date that an applicant registers an IPR in one country is considered the filing date in other countries as well, provided that an applicant subsequently also registers the IPR in those other countries within 6 to 12 months (dependent on the type of IPR).

Of particular relevance with respect to big data is the legal protection of databases. On 11 March 1996, the Council of the European Union passed the Directive on the legal protection of databases,[8] giving specific and separate legal rights (and limitations) to certain computer records. These are called database rights, lasting for 15 years. Database rights do not have to be registered and are independent of copyright.

4.2. The technological perspective

From a technological perspective, the main concern regarding data reuse is whether reusing data is technically feasible. Thus, the technological perspective looks at whether a certain form of data reuse can be technically realized. This may depend on whether technical components such as information systems can interact with each other (interoperability) and whether data is understandable in different environments (portability). Both depend on the use of standards: Interoperability depends on the standards used to design and develop systems used for storing, processing and transferring data, whereas portability depends on standards used for describing data. Without system interoperability and data portability, data reuse may be hampered, but also vendor lock-in may occur, because switching from one vendor to another may yield significant switching costs.

Interoperability

The reuse of data heavily depends on the extent to which systems for storing, processing and transferring data are (or can be) connected to each other. For the tremendous amounts of data that make up big data, data management and data analytics systems with specific characteristics need to be available. For instance, the data management systems need to be scalable, which means that they should be able to handle the growing demands for effectively storing and retrieving data. This capability is sometimes referred to as elasticity [17]. Similarly, the analytic processes should be scalable [18]. Another challenge is the transmission of data, which requires protocols for communication, routing, network performance and security [19,20]. Data transmission is particularly important in a data economy because multiple parties are usually involved in the use and reuse of data. Obviously, each data transmission requires an

---

[7]http://www.wipo.int/treaties/en/text.jsp?file_id=288514.
[8]Directive 96/9/EC of 11 March 1996. See Official Journal of the European Communities No. L77, 27.3.96, page 20.

agreement on the data exchange format. Furthermore, other aspects, such as data quality and security need to be ensured during data transmissions. Data security is vital as security issues may damage trust and reputation [21]. Hence, also protocols for data quality, security and privacy may be required. Quality certificates may be helpful for users to determine the quality, the origin and the original context of data. Another requirement is that an adequate technological infrastructure is available in terms of bandwidth and hardware infrastructure. This still varies widely across the EU. For instance, many rural areas are still lagging behind in this respect [22]. Improved system interoperability or even full integration of systems may be achieved, among other approaches, by standardized structures (formats for developing systems), integration of data across enterprise applications and reduction of technology redundancy.

Data formats and portability

The availability of data formats and open data are key drivers of the use and reuse of data. To increase availability of open data, the EU initiated the EU Open Data Portal.[9] However, these data are made available in a variety of formats. For instance, in Germany, approximately 90% of the available datasets are in PDF format, creating a significant problem for data reuse [23]. Large parts of these data (about half of it, depending on the government institution) contain structured information that could be made reusable by converting it to other formats, such as CSV.[10] The PDF format hinders interpreting, combining and reusing the data across parties and countries.

In fact, no EU standard exists for data or at least for open data. Some industries have their own sectoral standards but this creates barriers for using data across industries. Because big data is a global phenomenon, the technologies used in the EU are global as well. The complexity of this modern technology has led to an increase in the number and variety of standards [24]. Because some standards are very common, they are to some extent de facto standards, not because they have been approved by standards organizations, but because they are widely used and recognized in practice. A typical example is the XLS format, for Microsoft Excel data files. Several years ago, the EDI format[11] used to be the de facto standard for exchanging data in the business context. With the rise of the Internet, XML quickly became a popular standard and is still used for a wide variety of data [25].[12] XML is the basis for other formats such as RDF and OWL, but it is often replaced by JSON,[13] which is perceived to be more lightweight and compact [26]. Research has shown that government data portals mostly use XML, CSV, PDF and JSON [27]. Also Keyhole Markup Language (KML) and raw text files (TXT) are often used [28]. Standard data formats not only enhance possibilities for data reuse, but also possibilities for database coupling.

As mentioned above, data portability prevents vendor lock-in, as it allows switching to other providers of products and services without immense switching costs. Data portability obviously depends on standards for data formats. Under the new EU General Data Protection Regulation (GDPR), people will have a right to data portability. According to Article 20 GDPR, data subjects shall have the right to receive their personal data in a structured, commonly used and machine-readable format and the right to transmit those data to another data controller without hindrance. The GDPR does not prescribe specific formats [29].

---

[9]https://data.europa.eu/euodp/en/data.

[10]CSV stands for Comma-separated Values. It stores tabular data in plain text where each line of the file is a data record and each record consists of fields that are separated by commas.

[11]EDI stands for Electronic Data Interchange.

[12]XML stands for Extensible Markup Language and was standardized by the World Wide Web Consortium (W3C).

[13]JSON stands for Java Script Object Notation. It represents data as text in attribute-value pairs.

4.3. The societal perspective

From a societal perspective, the main question is whether or not a certain form of data reuse is socially accepted. Social acceptance may be different on a case by case basis. For instance, data reuse may be acceptable when this is for the same purposes and within the same context (data recycling). However, when data is reused for purposes different from the initial purpose for which the data was collected (data repurposing) or when data is used in a different context (data recontextualisation), this may not be acceptable. A typical example is that people post family pictures or holiday pictures on Facebook. Within this online context, they feel comfortable to share the pictures. However, when Facebook starts using these pictures for advertising their social network with large posters in inner city public areas, people may start to feel uncomfortable with this new purpose.

Awareness

The question that immediately comes to mind is whether such forms of data repurposing or even recontextualisation are allowed. When users sign the terms and agreements, they may consent to different forms of data repurposing and data recontextualisation. However, even when people have waivered their rights, knowingly or without being aware of this, they may not accept the resulting form of data reuse.

Therefore, the societal perspective is closely related to the legal perspective, as laws can be changed when society wants this. Nevertheless, the two perspectives may not always be neatly aligned, because social acceptance may change when awareness about technological developments (and related benefits and harms) increases, whereas changes in the legislation reflecting perceptions in society may take (much) more time.

The societal perspective is also related to the technological perspective, as technological developments may change user attitudes towards technologies as well as user behavior. For instance, after some time people may get more used and adjusted to new technologies, which may make these technologies more acceptable to them. Or, after enthusiastic initial use of new technologies, people may start to see disadvantages of new technologies (e.g., long-term effects) that make them change their mind.

Expectations

Social acceptance strongly depends on the expectations that people have. When it comes to data reuse, it is relevant to assess which expectations people have regarding, for instance, their privacy. In the US, the scope of privacy as protected in the Fourth Amendment is tested with the reasonable expectation of privacy doctrine. For instance, at home, in a hotel room or public restroom people have a reasonable expectation of privacy. However, in the streets, train stations or parks, people may not expect privacy. Social networks are considered to be public spaces, not private ones, and any information shared there is covered under the so-called third-party doctrine. Users have no reasonable expectation of privacy and any information they share can be used by their service providers [30]. However, this reasoning does not imply that people accept unlimited use and reuse of the personal data they share online. When expectations regarding data reuse are in line with actual practices, it is likely that these are socially accepted. When expectations are different from actual practices, it may depend on the awareness of these practices whether they are accepted.

4.4. The economic perspective

From an economic perspective, the main question is whether a certain form of data reuse delivers added value. When there are no good business cases, it is unlikely that anyone will take the initiative

for (increased) data reuse. However, this does not mean that each time data is reused financial value has to be created. Data reuse may also deliver non-monetary benefits such as human capital and increased health, well-being or satisfaction levels. A business case is critical for all types of data reuse (i.e., data recycling, data repurposing and data recontextualisation). While calculating a business case is usually quite straightforward for data recycling and repurposing, it can become very tricky for data recontextualisation, particularly if a company lacks experience using data across different contexts. An absolute prerequisite for data recontextualisation is the existence of an adequate ecosystem that includes, for instance, suppliers and demanders of data as well as data marketplaces or data sharing platforms that bring them together. For data recycling and data repurposing, the ecosystems is less critical but still relevant.

Ecosystem

Data recontextualisation requires a particularly comprehensive ecosystem. From the perspective of a company that considers using existing data, it is crucial that it is able to identify, assess and compare potential data suppliers. What makes this a lot easier is the availability of intermediaries that bring suppliers and demanders together. In the context of the data economy, intermediaries typically are data marketplaces or data sharing platforms. The availability of adequate intermediaries that are successful in balancing supply and demand is much more typical for mature markets than for emerging markets. The European Data Portal[14] is an example for a data sharing platform that focuses on public sector information. The portal is funded by the EU and harvests metadata of open government data available on public data portals across EU Member States. The central point of access, which resulted from a significant public investment, has made identifying, assessing and comparing suppliers of specific types of open data in the EU easier. However, the situation is not as comfortable with respect to other types of data that might come into question for data reuse. Most public and private intermediaries in the EU data economy focus on very specific types of, for instance, government data, research data or business data, or offer only a rather small number of datasets. Examples for EU data intermediaries with a specific focus are Figshare, which focuses on research data and allows researchers to preserve, share and access research outputs including datasets, papers, figures, media and code, AMEE, which focuses on business and environmental data and makes data directly available to interested parties by means of free programming interfaces, and Mobility Data Marketplace, which allows interested parties to offer, search for and subscribe to traffic-related online data including readings from traffic and environment detectors, traffic management measures, information about parking spaces, road works information, hazard and incident messages, and petrol station prices. EU data intermediaries with a broader scope include qDatum, Dawex and DataExchange. None of the EU-based data intermediaries, neither in breadth nor in granularity, currency or accessibility, is comparable to the big US-based data intermediaries such as Microsoft's Azure marketplace that offer predominantly US data globally. Aspects of the ecosystem that are critical not only for data recontextualisation but also for data recycling and data repurposing include the availability of enabling technical infrastructure and suppliers of big data technologies (both remote services and tools run on site) as well as the active adoption and use of the technology by market participants, capital markets as well as policy makers that are open for the data economy, ongoing investments in research and development related to big data and data reuse, and adequate education and training to ensure that sufficiently skilled staff is available on the labor market. Public investments and market regulations have proven to be particularly effective means to bring the data economy forward. Apart from that, it seems to be critical to increase the awareness of data suppliers and demanders that data reuse, if done properly, leads to a win-win situation for the involved parties.

---

[14]https://www.europeandataportal.eu/.

Business case

Data reuse must deliver added value. Otherwise, it is unlikely that anyone will take the initiative. Possibilities for data reuse usually emerge during the development or implementation of a concrete business model. Such a business model may be highly data driven, but this is not always necessary. In any case, for data reuse to become an option, the business model must rely on data to some extent. Data-driven business models may focus on activities ranging from data acquisition, data manipulation and data exploitation to technology provision and consultation. A business model as a whole as well as many detailed decisions must be economically reasonable. To make sure the right options are selected, business cases are typically calculated in practice. A well-crafted business case explores all feasible approaches to a given problem (e.g., whether to use existing or newly generated data for a specific purpose) and enables business owners to select the option that best serves the company. A business case must compare all solutions that may potentially be the best option. The costs and benefits for all options must be described in details as well as any relevant assumptions such as changes in regulations or strategies of competitors, and any relevant dependencies between a certain option and other business decisions. The costs and benefits then have to be evaluated for all options, ideally, illustrated with data from similar situations. The cost-benefit analysis, the heart of the business case, should include the projected financial benefit to the company and a projection of when that payoff is expected. In the context of data reuse, doing this is perceived as challenging by many companies. Not only data from similar situations is typically rare but also costs and benefits in the context of data reuse are particularly difficult to assess ex ante. If companies have not worked with a specific dataset already, it is hardly possible for them to determine its value. Technology investments can be relatively low due to the prevalence of open-source tools or remote services but labor costs may easily be underestimated. The costs needed for planning, cultural alignment, process definitions and deployment may be bigger than expected at first sight. While in particular its deployment may be challenging, big data and data reuse activities typically deliver significant paybacks. Moreover, it is difficult to take non-monetary benefits of data reuse adequately into account. In practice, however, it is not unlikely that forms of data reuse do not directly lead to a financial gain but rather to investments into human capital and increased health, well-being or satisfaction levels.

5. The way forward: Solutions for policy makers and business leaders

The multi-perspective view presented in this paper is useful for understanding conditions that are prerequisites for data reuse. Data reuse requires that the conditions in all four perspectives are met. After having identified which conditions have not been met in a specific situation, it becomes clear in which domain solutions should be looked for to support data reuse and to advance the data economy. For instance, when assessing data reuse shows that it is hampered because of limited interoperability or data portability, it does not make sense to work on social acceptance via information campaigns. Or when the analysis shows that legislation and technology are not hindering data reuse, it may be a good idea to look for societal or economic barriers and solutions. In this section, we suggest solutions for increasing the opportunities for data reuse from each of the perspectives. Figure 2 shows exemplary solutions for each of the four perspectives.

5.1. Addressing legal barriers

When legal issues are prohibiting data reuse, broadening opportunities for data reuse depends on the area of law causing this barrier. As pointed out above, privacy and data protection law or IP law may be
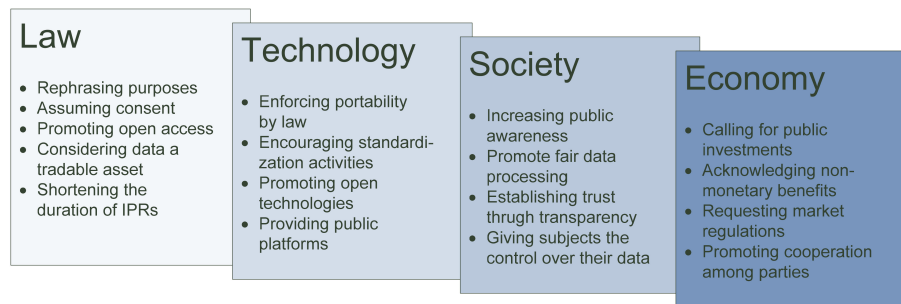
Fig. 2. Exemplary solutions for each of the four perspectives.

sources of limits to data reuse, although other areas of law, such as data localization laws, cybersecurity law, competition law and consumer protection law may also limit data reuse [13].

With regard to privacy and data protection law, several approaches may be possible to create more room for data reuse. A first approach may be to further investigate the balance between broad and specific purpose formulations. As explained, broad purpose formulations allow a lot of opportunities for data reuse but less legal certainty, whereas narrow purpose formulations create more legal certainty but limit options for data reuse to data recycling. Smart phrasing of purposes may create a better balance between these interests. For instance, listing a large amount of specific purposes might better meet both interests. In case users are not interested very much in the ways data is processed but are more concerned about the consequences of data processing, it may be useful to ask them which consequences are acceptable or not rather than explaining modes of data processing.

A second approach is related to this approach as it addresses assumed consent. Consent can be implicit or assumed in some cases. When people consent to processing of their sensitive data, like health records, for medical treatment, a reasonable approach might be to assume that they also consented to the use of non-sensitive data such as their address for the same purpose. When people indicated they are ok with receiving regular updates, it may be assumed that they also consent to the automated installing of these updates. In this approach, assuming consent for data reuse is inferred from large or general forms of consent to more specific forms of consent. People may disagree in specific situation to the assumed consent. However, using an opt-out system may result in fewer pointless consent requests than an opt-in system.

A third approach is using big data for extrapolating previously given consent. In other words, based on available data on consent decisions of a person, predictions can be made regarding future consent decisions. Based on these predictions, consent may be assumed in some cases. For instance, if a user has consent to data reuse of subscription data to three newspapers, it may be a reasonable approach to assume consent to other newspapers as well. Again, people may disagree with the assumed consent, as the consent predictions may be wrong in specific cases. Nevertheless, using the prediction to choose whether opt-in or opt-out should be the default may be useful.

A fourth approach may of course be to change legislation in order to broaden data reuse opportunities. However, from a practical perspective, it has to be mentioned that the EU legal framework for the protection of personal data was recently revised after many years of debate [31].[15] It is very unlikely that new changes to legislation will happen in the short term.

---

[15]Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

With regard to IP law, the protection of these interests is understandable and constitutes a good reason for limiting data reuse. However, there are some qualifications that can be made here. First, the protection of IP may be less appropriate when there were no private resources invested. For instance, when government institutions or publicly-funded research organizations collect data, it may be argued that this was done directly or indirectly with tax payers' money. Since society invested in the data, it could be argued that the data 'belong to' society and should, therefore, be publicly available. In 2003, the EU introduced Directive 2003/98/EC on public sector information.[16] The basic idea of this directive is that all content that can be accessed under national access to documents laws is in principle reusable beyond its initial purpose of collection for commercial and non-commercial purposes. The focus is on the economic aspects of the reuse of information rather than on the access of citizens to information, and Member States are encouraged to make as much information available for reuse as possible.

The main interest that IPRs aim to protect is ensuring that companies that invest in obtaining datasets have the opportunity to capitalize on their investments. However, using the data themselves and extracting useful knowledge from the data is not the only way to capitalize on the data. In some situations, it may be beneficial for companies to sell or hire the data they collected. Selling data provides immediate revenues and the advantage of data, in comparison with other goods, is that the same data can be sold many times. Hiring the data, for instance, by offering subscriptions that allow access to the data may also be a revenue model. For companies that sell or hire data, the advantage is that they can focus on collecting and preparing data, but have to focus less on data analyses, which requires different expertise.[17]

Another argument for data reuse and against IPRs is that combining datasets may yield results that otherwise would not be realized. The main advantage of big data is that it shows results that can only be found in large amounts of data. When two companies analyze their own, separate databases, they may not find results that can be found when these datasets are combined. Hence, sharing and exchanging data may be beneficial for both organizations. Contracts between two organizations that want to cooperate and set aside IPRs in order to create more benefits are not uncommon. However, arrangements between three or more parties are less common and more complicated in terms of what organization gets which part of the benefits. Nevertheless, such arrangements may be further considered as they may be valuable for all parties involved.

Finally, it may be discussed whether the duration of the protection that IPRs offer is still appropriate in an era in which new IP is generated at a very fast rate. The duration of IPRs should reflect a fair amount of time to provide sufficient opportunity to capitalize on investments. However, when the duration of the protection is too long, it will not be an incentive for companies to start working on and investing in new innovations. In such cases, IPRs will not achieve their goal, which was to incite and protect innovations and the investments required for this.

5.2. Addressing technological barriers

When technological barriers prohibit data reuse, it is mostly because systems cannot exchange data. This may be due to limited interoperability or issues regarding data portability. Both barriers can be

---

[16]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. See also LAPSI Policy Recommendation N. 2: The Interface Between the Protection of Commercial Secrecy and the Re-Use of Public Sector Information (n.d.). http://www.lapsi-project.eu/.

[17]Note that companies who sell or hire data have to know their clients well to be able to offer high quality products. To that extent, they have to have expertise in data analysis as well.

addressed by working with standards. Standards for interoperability may ensure that systems can communicate with each other and data transfers are enabled. Standards for data portability may ensure that data can also be processed in the environment to which they are transferred. Putting data in the right format is in some cases rather straightforward but ensuring system interoperability is a long term solution. Some companies may oppose system interoperability. For instance, Apple prefers to have a closed environment for its apps, among other reasons, to prevent viruses and malware entering the system.

Interoperability and data portability could be a legal requirement. As already mentioned, in the EU, the GDPR contains a provision on data portability. This provision aims to provide data subjects the right to receive their personal data previously provided to a data controller in a structured and commonly used and machine readable format and the right to transmit those data to another controller without hindrance from the controller to which the data have been provided. Thus, users have the right to obtain a copy of their personal data for further use and they have the right to transmit their personal data from one provider to another [32]. This applies to all types of data processing, including cloud computing and web services. The basic idea behind this right to data portability is to prevent vendor lock-in, to allow consumers to choose for a more secure, more developed or cheaper provider, and to strengthen competition in the market [33]. Granting data subjects the right to data portability strengthens their informational self-determination, as it provides them with more actual control over their personal data. At the same time, it further enables the transfer of data and opportunities for data reuse.

There are several organizations that develop standards for interoperability and data portability. One of the largest international standards organizations is ISO, the International Organization for Standardization, which is an independent, non-governmental international organization that drafts all kinds of international standards. There are several technical committees that draft international standards that are relevant for data reuse. The two most relevant are the (Joint) Technical Committee ISO/IEC JTC 1 on information technology and Technical Committee 184 on automation systems and integration. Each of these committees has several subcommittees. For IEC JTC 1, virtually all subcommittees are relevant for data reuse. When focusing on interoperability and data portability, subcommittee 32 (on data management and interchange)[18] and subcommittee 38 (on cloud computing and distributed platforms)[19] are relevant [34]. Important standards include ISO 9075 on database languages and ISO 11179 on metadata registries. For TC 184, most relevant for data reuse are subcommittee 4 (on industrial data) and subcommittee 5 (on interoperability, integration and architectures for enterprise systems and automated applications). The main standards of subcommittee 4 are ISO 8000 on data quality and ISO 10303 on industrial automation systems and integration [35].[20] Important standards of subcommittee 5 include ISO 11354, containing a framework for enterprise interoperability, and ISO 15745 on open systems.[21]

In the EU, standards created by CEN (European Committee for Standardization) are recognized as European standards. Via the Vienna agreement, CEN and ISO agreed to avoid duplication and potentially conflicting standards. CEN sometimes adopts ISO standards replacing the corresponding CEN

---

[18]For a list of ISO standards on data management and interchange, see ISO/IEC JTC 1/SC 32, at http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=45342.

[19]For a list of ISO standards on cloud computing and distributed platforms, see ISO/IEC JTC 1/SC 38, at http://www.iso.org/iso/home/store/catalogue_tc/catalogue_tc_browse.htm?commid=601355.

[20]While ISO 10303 is the standard that defines the neutral exchange format for design, manufacturing or production data, ISO 8000 is a more general standard concerned with the principles of data quality, the characteristics of data that determine its quality, and the processes to ensure data quality. ISO 10303 provides a representation of product information along with the necessary mechanisms and definitions to enable product data to be exchanged between applications as neutral, portable data.

[21]For a list of ISO standards on interoperability, integration and architectures for enterprise systems and automation applications, see ISO/TC 184/SC 5, at http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=54192.

standards [36]. It is important to mention that an issue with ISO standards is that some of them (like ISO 8000) are copyrighted and not freely available. That obviously does not contribute directly to increasing technological opportunities for data reuse.

The implementation of the EU open data policy [37] and the legal framework of Directive 2013/37/EU promote open data and interoperability and include some actions and requirements to promote open data and interoperability.[22] For instance, EU institutions are releasing their own documents as open data through the EU Open Data Portal[23] and they are supporting new open standards. For instance, the ISA program facilitates the use of common core data standards for national governments. It also promotes open-source software that can provide concrete solutions. A typical example is Apache Hadoop, which is a very common open-source solution for scalable and reliable distributed storage and processing of large datasets.[24]

5.3. Addressing societal barriers

In cases in which social acceptance hinders data reuse, the first question is whether people are fully informed about the data processing. Research has shown that many people do not know which data are available about them, to whom and in which ways and for which purposes these data are processed. As a result, people feel disconnected and not in control [38]. This may significantly decrease their trust in data processors and, in turn, decrease social acceptance.

Increasing awareness, for instance, by providing more transparency, may be helpful. Obviously, increased awareness will in many cases not directly increase social acceptance, as people may dislike particular forms of data collection or processing when they are informed about them. In these cases, it is important to seriously consider whether these practices are ethical and legitimate. If they are not, data controllers should refrain from these practices. The benefits of more transparency and user awareness for data controllers are more visible in the long term, when data controllers use fair information processing practices, are transparent about this and provide users with more options to control their data, this may increase trust among users and, in turn, increase social acceptance. Trust is hard to gain but easy to lose, so this approach requires a lot of time. To get people involved, practical first steps could be information campaigns, more transparency in terms and conditions and privacy policies, and providing users with more controls, like privacy settings.

5.4. Addressing economic barriers

Economic barriers may keep companies from initiating data reuse initiatives. The basic idea is that when a proper business case can be calculated in which companies can add value through data reuse, for instance, by collecting, aggregating, analyzing or selling data, someone will seize such an opportunity. However, when no viable business case can be made for it, data reuse will not take place.

When there are business cases that can be made in specific sectors, but no one takes these opportunities, for instance, because the market is immature, governments can decide to invest in creating more

---

[22]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.

[23]https://data.europa.eu/euodp/en/data. Note that countries like France and the UK have also started their own open data projects.

[24]https://hadoop.apache.org.

mature markets. Big data only yields results beyond a critical mass of data. Hence, investments may be required before tipping points are reached.

However, it may also be possible that business cases cannot be made on a solely financial basis. As explained above, sometimes benefits of data reuse are non-monetary, such as investments in human capital and increased health, well-being or satisfaction levels. Some companies may decide to invest in such activities as part of their portfolio and for their reputation but such activities may also be relevant for government agencies that have specific non-monetary tasks. Hence, governments do not only have a role in regulating and improving the maturity of markets but they may also take up some big data activities themselves in order to support data reuse and advance the data economy.

A third approach is to look at different types of revenue models. Large companies may prefer to use their own datasets and extract useful knowledge from the data themselves but this is not the only way to capitalize on data. For instance, for SMEs it may be beneficial to sell or hire data. Selling and hiring data usually provides immediate benefits and the same data can be reused many times. Also companies may combine their efforts in coupling their datasets, which may yield results that otherwise cannot be realized. The main promise of big data is that some results can only be found in large amounts of data. Hence, sharing and exchanging data may be beneficial for organizations. Furthermore, developing (especially multi-party) standard contracts for organizations that want to cooperate and set aside IPRs in order to create more benefits, may be helpful.

6. Conclusions

The reuse of data is an important factor when it comes to realizing the full potential of big data including accelerated innovation and growth as well as increased productivity and competitiveness. Still, data reuse is only happening to a very limited extent in the EU. To some extent, this is due to issues in the area of big data that call for innovative approaches to deal with legal, technological, societal and economic conditions. Each specific form of data reuse must be legally permitted, technically feasible, socially accepted and economically reasonable in order to take place. This means that company leaders and policy makers have to deal, among other topics, with privacy and data protection, IP, system interoperability, data formats and portability, public awareness and expectations, ecosystems and business cases. Solutions may include, for instance, rephrasing purposes and assuming consent for the legal domain, encouraging standardization activities for the technological domain, increasing public awareness and giving subjects the control over their data for the societal domain, and acknowledging non-monetary benefits and promoting cooperation among parties for the economic domain.

References

[1]   COM (2015) 192 final. A Digital Single Market Strategy for Europe. Brussels, May 6th 2015.
[2]   Gandomi, A., and Haider, M. (2015) Beyond the hype: Big data concepts, methods and analytics, *International Journal of Information Management* 35: 137-144.
[3]   Laney, D. (2001) *3D Data Management: Controlling Data Volume, Velocity and Variety.* Gartner. Stamford CT: META Group Inc.
[4]   Tene, O., and Polonetsky, J. (2012) Privacy in the Age of Big Data; A Time for Big Decisions, 64 *Stanford Law Review Online*, Feb 2, 2012, 63-69.
[5]   Mayer-Schönberger, V., and Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*, New York: Houghton, Mifflin, Harcourt Publishing Company.

[6]   Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburg, C., and Hung Byers, A. (2011) *Big Data: The next frontier for innovation, competition and productivity*, Seoul: McKinsey Global Institute. See http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation.
[7]   Burns, E. (2016) Why haven't SMEs cashed in on big data benefits yet? *TechTarget*. http://searchbusinessanalytics.techtarget.com/feature/Why-havent-SMEs-cashed-in-on-big-data-benefits-yet.
[8]   CORDIS (2014) *Helping SMEs fish the Big Data ocean*, Community Research and Development Information Service (CORDIS). http://cordis.europa.eu/result/rcn/93077_en.html.
[9]   Custers, B.H.M., and Ursic, H. (2016) Big Data and Data Reuse; A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection, *International Data Privacy Law*, 1-12. doi: 10.1093/idpl/ipv028.
[10] Lessig, L. (2006) *Code Version 2.0*, New York: Basic Books.
[11] Van den Hoven, J., Vermaas, P.E., and van de Poel, I. (2015) Handbook of Ethics, Values, and Technological Design. Heidelberg: Springer.
[12] Vedder, A., and Custers B.H.M. (2009) Whose responsibility is it anyway? Dealing with the consequences of new technologies. In: Sollie, P., and Duwell, M. Evaluating new technologies: Methodological problems for the ethical assessment of technology developments. New York: Springer. p. 21-34.
[13] Ursic, H., and Custers, B.H.M. (2016) Legal Barriers and Enablers to Big Data Reuse – A critical Assessment of the Challenges for the EU Law, *European Data Protection Law Review* 2(2): 209-221. doi: 10.21552/EDPL/2016/2/10.
[14] McDonald, A.M., and Cranor, L.F. (2008) The cost of reading privacy policies. *I/S Journal for Law and Policy for the Information Society*. http://www.aleecia.com/authors-drafts/readingPolicyCost-AV.pdf.
[15] Solove, D.J. (2013) Privacy self-management and the consent dilemma, *Harvard Law Review* 126: 1880-1903.
[16] Kemp, R. (2016) Legal Aspects of Managing Big Data (Kemp IT Law, 2014). http://www.kempitlaw.com/wp-content/uploads/2014/10/Legal-Aspects-of-Big-Data-White-Paper-v2-1-October-2014.pdf.
[17] Trivedi, N. (2014) Codename Elastic Storage: A Software Defined Storage-as-a-Service Offering on SoftLayer for High Performance Data Management in Cloud. https://www.ibm.com/developerworks/community/blogs/ibmsyssw/entry/codename_elastic_storage_a_software_defined_storage_as_a_service_offering_on_softlayer_for_high_performance_data_management_in_cloud?lang=en.
[18] Kaisler, S., Armour, F., Espinosa, J.A., and Money, W. (2013) Big Data: Issues and Challenges Moving Forward. In *Proceedings of the 46th Hawaii International Conference on System Sciences* (HICSS), p. 995-1004. IEEE.
[19] Guo, C., Lu, G., Li, D., Wu, H., Zhang, X., Shi, Y., et al. (2009) BCube: a high performance, server-centric network architecture for modular data centers, *ACM SIGCOMM Computer Communication Review* 39(4): 63. doi: 10.1145/1594977.1592577.
[20] Wang, G., and Ng, T.S.E. (2010) The Impact of Virtualization on Network Performance of Amazon EC2 Data Center. In *Proceedings of INFOCOM 2010*, p. 1-9. Piscataway, NJ, USA: IEEE.
[21] Ravikumar, G.K., Manjunath, T.N., Ravindra, S., and Umesh, I.M. (2011) A Survey on Recent Trends, Process and Development in Data Masking for Testing, *International Journal of Computer Science Issues* 8(2): 535-544.
[22] Stern, M.J., Adams, A.E., and Boase, J. (2011) Rural Community Participation, Social Networks, and Broadband Use: Examples from Localized and National Survey Data, *Agricultural and Resource Economics Review* 40(2): 158-171.
[23] Klessmann, J., Denker, P., Schieferdecker, I., Schulz, S.E. (2012) Open Government Data Deutschland, Eine Studie zu Open Government in Deutschland im Auftrag des Bundesministerium des Innern. https://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/ModerneVerwaltung/opengovernment.pdf?__blob=publicationFile.
[24] Tassey, G. (1999) Standardization in Technology-Based Markets.
[25] Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E., Yergeau, F., and Cowan, J. (2006) *Extensible Markup Language (XML) 1.1* (Second Edition): W3C Recommendation. http://www.w3.org/TR/2006/REC-xml11-20060816/.
[26] Severance, C. (2012) Discovering JavaScript Object Notation, *Computer* 45(4): 6-8. doi: 10.1109/MC.2012.132.
[27] Graves, A., and Bustos-Jiménez, J. (2014) Towards Visual Overviews for Open Government Data. In *Proceedings of Data Wiz* 2014.
[28] Moyle, K., Blackall, L., Hudson, F., and Mason, J. (2012) *Open Data and Open Data and Open Standards*.
[29] WP29 (2016) Guidelines on the right to data portability, Article 29 Data Protection Working Party (WP29), WP242/16, Brussels 13 December 2016.
[30] Goodman, M. (2015) *Future Crimes*, New York: Anchor Books, p. 94.
[31] Kuner, C. (2012) The European Commission's Proposed Data Protection Regulation: A Copernican Revolution in European Data Protection Law, *Privacy and Security Law Report*. 11 PVLR 06, 02/06/2012.
[32] Graef, I., Verschakelen, J., and Valcke, P. (2013) Putting the Right to Data Portability into a Competition Law Perspective, *The Journal of the Higher School of Economics, Annual Review*, 53-63.
[33] Swire, P., and Lagos, Y. (2013) Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique. 72 *Maryland Law Review* 335.
[34] ISO/IEC JTC 1 (2014) *Big Data, preliminary report*, Geneva: ISO.
[35] Benson, P.R. (2009) *Data Portability: It is about the Data; the Quality of the Data*. MIT Information Quality Industry

Symposium, July 15-17, 2009.

[36] Malcorps, G., and Quintana-Soria, I. (2007) *The Vienna Agreement – CEN as an International Partner*. CEN StandarD-ays, Day 2, Session 5, 2007. Brussels: CEN.

[37] COM (2011) 882 final, Open data An engine for innovation, growth and transparent governance, Brussels, 12.12.2011.

[38] Custers, B., Van der Hof, S., and Schermer, B. (2014) Privacy Expectations of Social Media Users: The Role of Informed Consent in Privacy Policies, *Policy & Internet* 6(3): 268-295.