Visions of Technology

Big Data Lessons Understood by EU Policy Makers in Their Review of the Legal Frameworks on Intellectual Property Rights, Access to and Re-use of PSI and the Protection of Personal Data

Hans Lammerant and Paul De Hert

Abstract This article’s focus is on how the advent of big data technology and practices has been understood and addressed by policy makers in the EU. We start with a reflection on how big data affects business processes and how it contributes to the creation of a data economy. Then we look at EU policy making on big data and its understanding of the role and impact of ICT in the economy. We study 3 major legal frameworks affecting data flows and uses: intellectual property rights, access to and re-use of PSI and the protection of personal data. We explore how these frameworks affect the use of big data and how this is perceived and dealt with in the policy documents. In order to widen our perspective, we also take a comparative look at similar legal frameworks and policies in the US.

Keywords Big data · Policy · Data protection · Copyright · Open data

1 Why Understanding the Technology Visions of Policy Makers?

This article looks at EU policy developments regarding big data and, second step, see how it interacts with legal frameworks regulating data flows. Instead of taking, for instance, data protection law as a point of departure to look at big data, we
want to take the inverse approach and start from the perspective of big data and the data economy resulting from it. How do legal frameworks affect the use of big data and the operation of a data economy?

A data economy is built upon establishing data value chains and dependent on the possibility to collect, aggregate and process data from diverse sources in an automated process. Legal frameworks affecting data flows have therefore an impact on big data processing and define the space for a data economy. Seen from this perspective, not only data protection, but also other legal frameworks like intellectual property rights and the regulation of access and re-use of public sector information (PSI) frame or influence the data flows on which big data operates. Although not developed specifically for big data, these frameworks regulate and condition the access to and processing of specific types of data and shape therefore the data economy.

Policy makers and legislators look for legal frameworks that allow an economy to flourish and capture the benefits of technological developments while balancing between all the values and interests at stake. To reach that ambition they try to understand the effects of technological developments and grasp what it means in terms of how an economy functions. In other words, they form a certain idea about the effect of ICT on the economy and society, and develop regulation based on that idea. The initial reaction of policy makers on technological developments often starts with adapting and patching legal frameworks, when its impact is not very clear yet or is considered not too profound. Only later more fundamental reviews will be made, dependent on the new understanding of the societal impact.

Therefore visions on how technology changes economical processes are important drivers of legal change. Policy documents are also a testimony of policy learning about the impact of technology.

This article will focus on how the advent of big data technology and practices has been understood and addressed by policy makers. We look into the development of the EU big data policy and how it interacts with the legal frameworks regulating data flows. First we give a short introduction of our understanding of how big data affects business processes and how it results in a data economy (Sect. 1). Next we describe the evolution of the EU policy on big data and what it says on the role and impact of ICT in the economy (Sects. 2, 3 and 4). Following that we consider 3 major legal frameworks affecting data flows and uses: intellectual property rights (Sect. 5), the protection of personal data (Sect. 6) and access to and re-use of PSI (Sect. 7). We explore how these frameworks affect the use of big data and how this is perceived and dealt with in the policy documents. In order to widen the analysis, these sections also look at similar legal frameworks and policies in the US. In a last section we present our conclusion that recent EU policy

---

1This article is the result from research done as part of the BYTE project (http://byte-project.eu/). The authors are solely responsible for the opinions expressed.
documents reflect a new understanding of the data economy, but that the translation of this vision into the legal frameworks shows mixed results (Sect. 8).²

2 Data Economy and the Big Data Value Chain

The term big data is vaguely defined and partly a buzz word which came into popular use only recently. Therefore it appeared only recently in EU policy documents. As a term it is absent in the main EU policy documents such as the one on the Digital Agenda,³ on cloud computing⁴ or related documents from 2010 and 2012. It only appears in the policy documents of 2013 and 2014.⁵ These policy documents bear witness of the learning process of the EU institutions concerning the economic and social impact of ICT developments.

Big data as a phenomenon is enabled by new developments in distributed computing like cloud technology, allowing to deal with very large amounts of data at much higher speed. However, big data cannot be equated with these technologies or cannot be limited to these aspects of volume and velocity. It also implies qualitative changes in terms of what can be done with this data: a variety of structured and unstructured data sources can be much easier linked with each other and analysed in new ways. New business models are built upon the capacity to capture value from data through the development of a data value chain along which data is transformed into actionable knowledge.

The concept of value chain was first introduced by Michael Porter and consists of a series of linked activities through which value is created.⁶ These linkages are relationships between the performance of one activity and the cost and performance of another. The construction of a data value chain implies a new way to

²The scope of this article does not allow us to be exhaustive and limits us to exploring the subject. We focus on regulatory issues concerning the access to, linking of and (re-)use of data and the legal environment in which this takes place. Other elements of policies, like concerning investment in infrastructure or research projects, we leave outside of our consideration. Also specific regulations, e.g. on law enforcement, remain outside the remit of this article.


optimize output or to create new products and services and a new configuration of
activities and actors to do so.

This data value chain has obtained a central role in a data-driven knowledge
economy and pushes organisations and administrations to open up their data
sources and business processes in order to reap the benefits, resulting in a new
‘data ecology’ consisting of a diversity of actors providing, collecting or analysing
data and acting upon the results. Old organisational barriers are penetrated by data
flows. Old legal frameworks regulating such data flows come under pressure. They
present barriers to this new data-driven economy or have difficulties to assure the
balance between interests and values embedded in them.

The reconfiguration of activities by the construction of a data value chain also
changes the role of the Internet. Where the Internet was first conceived as a sepa-
rate economic space alongside the traditional economy, it evolved into a market
place and distribution channel. In this vision economic actors remain units outside
the Internet but meet each other through it. With the construction of a data value
chain the Internet penetrates these economic units and becomes also the environ-
ment in which value creating activities take place and get linked to each other.
Economic activity over the Internet broadens from an information economy,
foccused on content and services for human customers, to a data economy where
data mostly flows between a range of non-human actors processing this data, often
in real time. The development of the Internet-of-Things will further augment this
evolution. Big data practices are of course possible in contexts outside a data econ-
yomy, like for data-intensive scientific uses, but a widespread commercial use is
correlated with the possibility to build data value chains. This evolution of the role
of the Internet is, as will be seen, reflected in recent policy documents as it brings
up new regulatory questions about e.g. the space for data mining in the context of
IPR, and about profiling in the discussions on the upcoming data protection regu-
lation (General Data Protection Regulation or GDPR).

Key element in the construction of data value chains is the interoperability of
datasets, or assuring that datasets can be combined and analysed together. The
European Interoperability Framework (EIF) provides a useful conceptual model
of interoperability levels: legal, organisational, semantic and technical interopera-

7European Commission, The Digital Agenda for Europe—Driving European growth digitally,
8European Commission, annex II of the “Communication: Towards interoperability for European
9Technical interoperability concerns the technical aspects of linking information systems.
Organisational interoperability concerns how organisations cooperate to achieve their goals. It
implies aligning business processes and the related data exchanges. Legal interoperability con-
cerns how to deal with differences in legal status. Datasets can have a different legal status and be
subjected to different legal rules, what can lead to obstructions linking them or to limitations of
data exchange. Semantic interoperability ensures that the precise meaning of exchanged informa-
tion is understood and preserved throughout the data exchanges. It involves developing descrip-
tions or other metadata and vocabularies concerning the exact format of information and the
meaning of data elements and their relations. Growing levels of semantic interoperability make it
easier to link otherwise isolated data sources.
doubt about the fact that interoperability of datasets is more than just technique. It is (also) influenced by legal frameworks, organisational structures and needs investment in data quality in order to capture potential benefits.\(^\text{10}\)

Our focus on legal frameworks regulating data flows implies a focus on legal interoperability, but we will touch upon other aspects when useful. Legal interoperability is affected by several legal frameworks, developed in contexts where big data or interoperability of data were still unknown notions. In later sections we will consider 3 major legal frameworks affecting data flows and uses: intellectual property rights, access to and re-use of public sector data (PSI) and the protection of personal data. We will look into how these frameworks affect data flows and how policy deals with them when confronted with the new big data practices. First we will have a look at some of the basic documents in the history of EU policy making on big data.

### 3 Digital Agenda for Europe (2010–2012): Overview of Major Relevant Actions

The European Commission developed since the 1990s broad policy documents concerning the information society. These include a strong focus on developing a stable legal environment for commercial activities over the Internet: regulation of e-commerce, adapting intellectual property rights, and so on. But also a strong focus on economic development. In the 1990s this is linked with the liberalisation of telecommunication services. From 2000 onwards the e-Europe 2002\(^\text{11}\) and 2005\(^\text{12}\) action plans and the i2010 strategic framework\(^\text{13}\) focus a lot on improving Internet access through broadband as a key enabler and the development of a rich content industry and services making full use of this potential. This includes making public services accessible over the Internet. All 3 legal frameworks we consider in this article are reconsidered and adapted to the Internet economy in this earlier period. The data protection directive 95/46/EC is adapted in its final drafting to take better account of the context of digital telecommunications networks\(^\text{14}\) and in 2002 the E-Privacy Directive addressed specifically the electronic

\(^{10}\)This conceptual model was developed for public services, but we use it here in a generalised meaning.


communication sector.\textsuperscript{15} Copyright law has been harmonised and adapted,\textsuperscript{16} while a new regime of database protection was introduced.\textsuperscript{17} Re-use of PSI got regulated in 2003.\textsuperscript{18} These policies, and the directives drafted in this period, are based on the information economy vision: the Internet as market place visited by human clients (see our discussion above). In the later part of this article we will consider their functioning in the newer context of a data economy. First we look into how the more recent EU policy digested the advent of big data.

The 2010 \textit{Europe2020} strategy (updated in 2012) sets out a vision on how the EU has to develop its social market economy. This vision functions as a coordinating umbrella vision for more specific policy initiatives. Part of the \textit{Europe2020} strategy were 7 flagship initiatives, one of them being the ‘Digital Agenda for Europe’. The main focus of this Digital Agenda is “a digital single market based on fast and ultra-fast Internet and interoperable applications” and builds upon the earlier action plans.

The Digital Agenda contains a comprehensive agenda concerning the digital economy. It identified a wide range of obstacles: fragmented digital markets, lack of interoperability, rising cybercrime and risk of low trust in networks, lack of investment in networks, insufficient research and innovation efforts, lack of digital literacy and skills and missed opportunities in addressing societal challenges. The actions defined in answer to these obstacles are as wide ranging.\textsuperscript{19}

A first relevant action in the Digital Agenda, within the aim to create a digital single market, concerns the opening up of access to content. The main problem is that the European digital market is still very fragmented, both concerning private and public data or content. Action points identified are simplifying copyright clearance, management and cross-border licensing.\textsuperscript{20} Part of this has been the review of the PSI Directive in 2013 and the adoption of Directive 2014/26/EU on collective rights management and multi-territorial licensing, but also the ongoing review of the data protection framework with the proposed General Data Protection Regulation (GDPR) and e-commerce related legislation.\textsuperscript{21} The Commission planned continued action on e-commerce related issues and


\textsuperscript{17}European Parliament and the Council, Directive 96/9/EC of 11 March 1996 on the legal protection of databases.


\textsuperscript{19}We will focus on the actions and resulting policy initiatives that concern the access, linking and use of data.


intellectual property rights. These plans were repeated in the intention in the 2012 update of the Digital Agenda to make proposals to strengthen the European data industry, specifically on “issues such as common licensing conditions and the implementation of charging rules to enable public data to fuel the development of online content”. Also problems concerning text and data mining were mentioned by announcing the structured stakeholder dialogue Licences for Europe held in 2013, which addressed cross-border portability of content, user-generated content (UGC), data and text mining, access to audiovisual works and cultural heritage institutions. At this point the attention was still limited to text and data mining for scientific research purposes. As part of this effort on content the Commission also focused on public data. It presented its policy in the Communication on ‘Open data. An engine for innovation, growth and transparent governance’. Public sector information is seen as a resource. With an active open data-policy this resource is made available for the European economy.

A second important action area linked to data policies in the Digital Agenda is the focus on interoperability and standards. This concerns a wide range of hardware, software, IT services and it can also concern data. Standardization has always been an important instrument in the single market and it also plays a key role in creating a functioning data economy. When content remains locked up in incompatible formats, licenses, etc., the data economy remains very fragmented. The focus on making data sources more interoperable through standardisation is mostly present in the effort to enhance the interoperability between public administrations.

Big data was not mentioned in the Digital Agenda, but the agenda nevertheless contained attention for cloud computing as part of its innovation strategy, with the development of “an EU strategy for cloud computing notably for government and science” as specific action. Cloud computing is an important enabling infrastructure for big data processing, and attention for cloud computing is the only element specific to big data in this Digital Agenda. The Commission next outlined a specific policy agenda on cloud computing in its 2012 Communication Unleashing the Potential of Cloud Computing in Europe. It presented 3 key actions: enhance standards and certification, establishing safe and fair contract terms and conditions (through model contracts and contractual clauses, and a code of conduct for cloud computing providers) and the launch of the European Cloud Partnership. Especially the action on contracts has an important effect on the access and use of data, even when it concerns infrastructure rather than big data processing itself. It

---

22Ibid., p. 6.
can create a more predictable and safe environment in terms of data security and data protection and prohibit cloud providers to abuse data on their servers for other purposes. The state of work and results are presented in the 2014 Report on Implementation accompanying the Communication on a data-driven economy.\textsuperscript{26}

The Digital Agenda for Europe (2010–2012) did not contain yet a fully developed vision on the data economy. It continued to consider the Internet as a market place where consumers and e-commerce enterprises meet. Most attention therefore went to ensuring this market functions properly both for enterprises and consumers and to integrate national markets into a single digital market.

Big data or a data economy as such does not enter the picture yet. Cloud computing does but mainly as an infrastructure delivering more flexible IT resources to companies. Attention therefore goes to market conditions between cloud providers and enterprises buying cloud services. Interoperability concerns mostly hardware and software, but not so much data apart from specific applications. Data protection is still mostly seen as an element to establish trust for consumers in e-commerce, which remains in the older vision of an information economy, while specific big data-related concerns are not considered.

Notwithstanding this several important elements linked to a data economy are already present in the Agenda. IPR-related problems for data and text mining appear in 2012 as a distinct issue.\textsuperscript{27} PSI and open data are present as policy issues. The e-government policy and the policies on research data and geospatial data (both are distinct areas of policy planning) appear to be driving areas from which new practices putting data central are developed.

4 Towards a Thriving Data-Driven Economy (2014): Four Regulatory Issues

The European Commission presented an updated version of its vision on the data economy in its 2014 Communication on a data-driven economy.\textsuperscript{28} It builds upon the ideas first formulated by Commission Vice-President Neelie Kroes in a strategic initiative on the data value chain in November 2013,\textsuperscript{29} in response to the European Council’s conclusions of its meeting on 24–25 October 2013, where big

\begin{enumerate}
\item European Commission, A European strategy on the data value chain, November 2013.
\end{enumerate}
data as a concept appeared on the EU policy agenda. This policy agenda aims to “provide the right framework conditions for a single market for big data and cloud computing”. It puts data forward as the central element in the future knowledge economy. Data-driven innovation is defined as “the capacity of businesses and public sector bodies to make use of information from improved data analytics to develop improved services and goods”. Improved data analytics are seen as key to more efficient business and production processes. With this communication a more profound understanding of the impact of big data on business processes appears in EU policy documents.

The Communication further points to the slow embracing of this ‘data revolution’ in Europe compared to the US. Among the causes the “complexity of the current legal environment” is mentioned. To reverse this the EU must “make sure that the relevant legal framework and the policies, such as on interoperability, data protection, security and IPR are data-friendly”. Other needs include an accelerated digitisation of public services and sharing and developing its public data resources. In order to develop a data-driven economy good quality, reliable and interoperable datasets, backed by an enabling infrastructure have to be present, as well as an adequate skills base and close cooperation between public and private partners. The action plan announces several initiatives to make progress towards such data-driven economy, including the development of open data policies and standards and several regulatory issues.

The first regulatory issue concerns personal data protection and consumer protection. After the adoption of the GDPR the Commission plans to work on guidance for issues important in the big data context, like data anonymization, data minimization and privacy by design. Further regulatory work concerns ensuring the application of consumer law on big data technologies. The second issue raised is data mining and its relation to the copyright framework. Thirdly, the Commission plans to explore the security risks related to big data technologies and propose risk management and mitigation measures. Finally issues concerning data ownership and data transfer will be considered. Mentioned are data location requirements, presenting a barrier for cloud computing and big data, and data ownership and liability in the context of the Internet of Things.

In the 2014 Communication the European Commission develops a new vision on the data economy and puts forward the central role of (big) data in the knowledge economy. Regulatory issues raised concern similar areas as before, like IPR and data protection, but now with attention to their impact on the data value chain. Similarly, the attention for open data includes more attention for interoperability and an investment in semantic interoperability.

However, its proposed actions clearly build on the earlier initiatives. The earlier importance of the Internet as a market place does not disappear, but gets supplemented with attention for specific issues linked to data value chains.

---


The Communication *A Digital Single Market Strategy for Europe* of 6 May 2015 is the first major policy document on the digital economy of the newly installed Commission led by Jean-Claude Juncker.\(^{31}\) Although still focussing a lot on the Internet as a market place, an in-depth vision on the data economy is now clearly integrated.

The first pillar focuses on market integration by the removal of obstacles for cross-border trade and of the differences between online and offline trade. Under this pillar the Commission envisions to review the copyright framework and make legislative proposals before end 2015. Again mostly focussed on audio-visual content, it also foresees creating “greater legal certainty for the cross-border use of content for specific purposes… through harmonized exceptions”. Purposes mentioned are research and text and data mining.

The second pillar aims at reform of the telecommunications and the media sector to enhance market integration and competition. The Commission also plans before end 2015 a comprehensive assessment of the role of platforms like search engines, social media, e-commerce platforms, …. These platforms have been innovators and early adopters in the creation of a data value chain and building new business models around it. The success of some platforms has now led to concerns over their growing market power. This assessment will therefore consider issues like transparency (e.g. in search results), how the platforms use the information they collect, the ability of individuals and business to switch platforms, and other issues connected to the bargaining power of the platforms. Further under this pillar the Commission foresees measures to improve trust and security in digital services and protection of personal data. This involves the continuation of earlier initiatives like the GDPR and the Network and Information Security Directive, which are proceeding through the legislative process, and a review of the ePrivacy Directive after the adoption of the GDPR.

The third pillar is more focussed on the data economy. It will propose in 2016 a European ‘Free flow to data’ initiative. Where the GDPR prevents member states to restrict the flow of personal data within the EU, with this initiative the Commission wants to tackle other restrictions to data flows and on the location of data for storage and processing (e.g. for security reasons). In this context it wants to address issues like ownership, interoperability, access to data and data portability. The Commission also plans a European Cloud Initiative involving issues like cloud services certification, contracts, liability, switching of providers and so on.

Further the Commission wants to put extra effort into interoperability and standardisation. The focus of the standardisation effort is now broadened from

hardware and software to the data component. The Commission points to the need to define standards “essential for supporting the digitisation of our industrial and services sectors (e.g. Internet of Things, cyber security, big data and cloud computing)”. This effort also has a e-government component. The Commission plans to review the European Interoperability Framework and further focuses on achieving cross-border interoperability. It will present a new e-Government Action Plan 2016–2020 with several initiatives to extend national e-government services across borders. These initiatives include the interconnection of business registers, an initiative to pilot the ‘Once-Only’ principle cross-border, extending and integrating European and national portals towards a ‘Single Digital Gateway’ and accelerating the transition of member states towards full e-procurement and interoperable e-signatures.

The 2015 Communication takes up unfinished initiatives from the former Commission, like the review of copyright law and the GDPR, and builds further upon earlier work. The data economy gets more attention alongside the digital market perspective. Together with the earlier communication on a data-driven economy this communication presents a clear policy agenda on big data.

In the following parts we take a closer look at the legal outcomes of these policy initiatives. We study three major EU legal frameworks that affect or have the potential to affect data value chains. None of them has been developed specifically for big data. To broaden the perspective we compare the EU and US legal frameworks, explore how they apply on data and limit its use. We further look at how the advent of big data was received in this context. The first legal framework we consider is the one of intellectual property rights (IPR).

6 The Intellectual Property Rights Framework (Framework 1): Adequate for Big Data?

6.1 The Application of Copyright on Datasets Is not Straightforward

IPR protect intellectual creations and reserve certain exclusive rights concerning their use and distribution to their creators or those to whom these rights have been transferred. Each regime defines what falls under its protection. Certain regimes can apply to data and datasets. Most relevant are copyright and database protection.

Protection by IPR of datasets is a major obstacle for access, linking and use of data and therefore also for big data processing. These limitations can be legitimate, but the framework is not well-adapted for a situation where data flows in large amounts between a broad range of actors and gets processed in real time. When the data is protected by copyright or database protection, authorization of the right holders is required. This can lead to large transaction costs or delays, and
is practically impossible in certain use cases, e.g. text mining on thousands of articles or webpages. Solutions can be developed in licensing schemes specifically adapted to data mining practices, but such solutions remain limited to right holders applying them. A more radical solution would be limiting the protection and allowing the specific data use without requiring authorization. This can be done on several levels: the subject matter of the protection, the extent of the reserved usage, the exceptions on these reserved rights.

Copyright can exist over the individual data as well as over the database as a whole. Copyright protection for databases results from the copyright for collections. The protection concerns the organisation and structuring of the data but does not extend to the individual data items itself. Copyright of individual data items grants exclusive rights on the individual item, but is independent from the copyright over the database structure as a collection. Both have to be checked separately and can belong to different right holders.

General principle of copyright is that it protects expressions, but not ideas in itself, nor procedures, methods or mathematical concepts.\(^{32}\) Aim is to protect products of human intellect and creativity. Trigger for the protection is therefore some sort of originality. Originality implies originating from an author, but also being the result of some intellectual or creative effort.\(^{33}\) Novelty is not required, but the mere investment of effort in copying information does not reach the threshold for copyright protection. Also, purely factual information is not protected under copyright. Basic idea is that facts are discovered and not the result of creativity. Copyright protection given to the expression does not extend to the underlying facts. This factual information can be used by others, as long as they do not reproduce it in the protected expression.

The application of copyright on datasets is therefore not straightforward: not all data is protected by copyright, but only those that meet the originality-requirement. For instance, maps have been subject to copyright controversies, as the factual geographical information as such lacks the element of creativity.\(^{34}\) The \textit{Infopaq}-decision of 16 July 2009 the European Court of Justice (EUCJ) concerned the application of copyright law on a search engine of newspaper articles, providing summaries of articles. It stated that the protection by copyright applies only when the data “is original in the sense that it is its author’s own intellectual creation”.\(^{35}\) This originality requirement also needs to be checked when reproduction in part is concerned. In this case a string of 5 words before and after the keyword were stored. The Court considered a word in isolation not to be the intellectual

\(^{32}\)WIPO Copyright Treaty, art. 2; TRIPS, art. 9 §2.


creation of the author, but that such creation could be achieved “through the choice, sequence and combination of those words”.\textsuperscript{36} Isolated words were therefore not covered by protection, but strings of 11 words could be and this needed to be checked by the national court. This decision clarified the originality requirement upon which copyright protection in the EU is based and made clear that the specific technical characteristics of text and data mining methods are legally relevant in the context of copyright law. Methods based on ‘bag-of-words’ sets, making a frequency distribution of words in a text and thereby taking all words out of their context, can avoid the applicability of copyright protection, but not those methods using longer strings.

The lack of clarity concerning the application of copyright on data has been resolved in divergent ways. Based on a similar economical reasoning the US courts refuse to extend copyright protection to claims purely based on investment, while in the EU the policy maker has generally chosen to strengthen the protection. The resulting legal environment strongly affects big data practices.

\section*{6.2 Striking Differences with the US Copyright Framework}

The US courts have seen a lot of legal battles on what can be protected by copyright and what not, including cases concerning several sorts of data and compilations of data.\textsuperscript{37} Main legal precedent is \textit{Feist}, in which the Supreme Court made clear that effort or investment is as such not protected by copyright. It took distance from court decisions which granted protection to ‘sweat of the brow’ or ‘industrious collection’, through which courts had earlier developed a protection for factual collections. Instead it reaffirmed that originality was an essential requirement, grounded on the objectives of copyright protection listed in the Constitution “to promote the Progress of Science and useful Arts”. Copyright also needs to allow others to build upon the ideas and information contained in a work, which is the rationale for only granting protection to the expression but not to facts. The case concerned the white pages of a telephone directory, consisting of an alphabetically ordered lists of names with their town and telephone number. The Court considered that such lists of facts lacked any originality and were not protected by copyright. Factual data in databases or other works are available for reproduction or extraction, even when this extraction is substantial.

The EU has on the contrary resolved the issue by introducing an extra legal protection on databases with a sui generis right. Database protection is provided by directive 96/9/EC of 11 March 1996 on the legal protection of databases. This

\footnotesize{\textsuperscript{36}Ibid., §45.}

\footnotesize{\textsuperscript{37}An overview of case law can be found in Leslie C. Ruiter and Gerald van Belle, Data Extraction: Beyond the Sweat of the Brow, \url{http://www.stokeslaw.com/uploads/pdf/data_and_the_law-gerald_van_belle_and_leslie_ruiter.pdf}.}
directive contains 2 forms of protection for databases, one as copyright, another as a sui generis right. These protections can coincide.

The copyright on databases protects databases where “the selection or arrangement of their contents” is a result of “the author’s own intellectual creation”. This protection does not extend to the contents. This remains similar to the copyright protection on databases in the US, derived by the courts from the protection of collections.

The sui generis-right protects the maker of a database who has made a substantial investment in the creation of a database. Only the costs associated with “obtaining, verification or presentation of the contents” as a whole are taken into account, not the cost associated with obtaining, creating or updating individual data items. No originality is required, protection is based on the investment. The maker of the database is given the right to prevent extraction and re-utilization of the whole or of a substantial part of the contents of that database. This right does not prevent lawful use, consisting of extracting or re-utilizing insubstantial parts of database contents. The substantiality can be assessed both quantitatively and qualitatively. Further may this use not conflict with the normal exploitation of the database or unreasonably prejudice the legitimate interests of the maker of the database. Result is that any big data processing involving a substantial part of a database will need permission from the right holder during the 15 years term of protection.

The EU introduced the sui generis protection based on the assumption that property rights attract investments and therefore stimulate the economy. In the US protection was refused on a similar economical reasoning. Ian Hargreaves pointed in his review of the intellectual property framework to the evaluation in 2006 by the European Commission of the Database directive. This evaluation shows less investment instead of growth, while the US market kept growing without such protection. Hargreaves sees this as an example of policy development inconsistent with the available evidence. The European Commission has kept the directive unchanged seen the large support of the concerned industry for the directive. It can be questioned if such large support shows the economic value of the directive in general or if it shows the value for a specific interest group. Would a data economy be better off with less protection of databases through IPR?

A second difference between the EU and US frameworks can be found in the exceptions to the reserved rights provided in these frameworks. The EU


harmonised copyright and adapted it to the digital environment in the Information Society or InfoSoc directive.\(^{42}\) The directive includes a set of quite precise exceptions, which are mostly optional. All these exceptions are limited by the ‘three-step test’: they can “only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rights holder”.\(^{43}\) Relevant is the exception on the right of reproduction for temporary acts of reproduction which are transient or incidental, are an essential part of technological processes like transmission or other lawful uses and have no independent economic significance.\(^{44}\) This exception was meant for caching and temporary storage during digital communication. In the big data context the question is if this can also be used for text and data mining. In *Infopaq* the EUCJ stated that the copies made of the newspaper for the search for keywords could be considered a temporary and transient act of reproduction that fell under the exception if those copies were indeed automatically deleted at the end of the process. This exception could not apply for the further storage or printing of the strings of 11 words, when these fell under protection. In a second decision in the same case the EUCJ broadened the exception by using the 3-step test.\(^{45}\) The exception for temporary storage did draw attention also in other decisions,\(^{46}\) but in general the scope for data mining remains quite narrow.

The US Copyright Act of 1976 does not contain a long list of specific exceptions, but grants an exception to the fair use of a copyrighted work.\(^{47}\) What constitutes fair use is illustrated with the purposes of “criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research”. Further the law provides four factors to consider in order to determine what is fair use. The ‘four factor-test’ involves the “purpose and character of the use”, the “nature of the copyrighted work”, the “amount and substantiality of the portion used” and the “effect of the use upon the potential market for or value of the copyrighted work”. The evaluation is made globally and no factor is more important than another, although the economic impact has in practice got more importance. As part of the evaluation of the purpose courts have looked to the transformative nature of the new use. The more a new use is distant from the earlier use and the less it can be conceived as a mere re-packaging and copying, the


\(^{43}\)Ibid., art.5 §5.


\(^{46}\)EUCJ, C-360/13, *Public Relations Consultants Association Ltd v. Newspaper Licensing Agency Ltd and Others*, 5 June 2014 (aka the Meltwater decision); EUCJ, C-403/08 and C-429/08, *Football Association Premier League Ltd*, 4 October 2011.

more chance it makes to be considered fair. The case law on this fair use-exception is very extensive and concerns a wide range of Internet-related practices like hyperlinks, copying or reproducing of images and text by search engines, … The fair use framework proved able to flexibly deal with new technical developments.

This comparison between the EU and US copyright framework showed some striking differences. First, the US has limited the protection of databases to copyright and has never extended IPR protection purely on grounds of investment. Secondly, its fair use regime proved to be much more technology neutral and adaptable to new technological developments. These 2 differences in the IPR regime lead to a large difference in playing field for big data processing.

Several countries did notice the problem the IPR regime poses for text and data mining and adapted the exceptions or are discussing a change. The UK added an exception for computational analysis for the purpose of non-commercial research. In its international strategy on IPR the UK government further included the aim “to secure further flexibilities at EU level that enable greater adaptability to new technologies”. Japan updated its copyright law with a new exception giving space for information analysis. The Australian Law Reform Commission recommended after a public consultation to adopt a general ‘fair use’-exception like in the US as a flexible and technology-neutral solution. A similar review took place in Ireland, leading to the recommendation to add exceptions for ‘content-mining’ for purposes of education, research or private study to both copyright and the protection of databases as well as a fair use-exception. Both recent reports have not led yet to legislative action.

IPR policy in the EU has been focused at strengthening the protection of right holders, motivated by a concern to develop a strong content industry. The sui generis protection of databases is an early witness, but recent policy documents have kept this focus. The recent shift in perception towards a data economy created the space to raise the problems a strict IPR framework poses for text and data mining. This issue entered the agenda in 2012 and the Juncker Commission plans to adapt the exceptions regime for text and data mining. On the other hand, no discussion did arise on the usefulness of the sui generis-protection of databases.

50 Triaille, Jean-Paul, Jérôme de Meeûs d’Argenteuil and Amélie de Francquen, Study on the legal framework of text and data mining (TDM), March 2014, pp. 10–11.
52 Copyright Review Committee, Modernising Copyright. The Report of the Copyright Review Committee for the Department of Jobs, Enterprise and Innovation, Dublin, 2013.
We can conclude that the EU still focuses on strengthening protection with IPR from information economy perspective, but has now more attention to its fine-tuning in the context of a data economy. If this suffices for an adequate big data policy can be questioned. The legal interoperability of data remains lower in the EU due to IPR protection.

In the US no such debate on IPR can be found. The US IPR framework proved to be open and adaptable for new legal developments thanks to its fair use-regime. This does not mean no conflicts between new technological applications and IPR did arise. But these have not been subject of policy making, but of legal disputes and court decisions.

7 Protection of Personal Data (Framework 2)

7.1 A Comparison Between the EU and US Legal Regime

The legal frameworks dealing with personal data are very different in their foundations and grounded in different constitutional cultures. This results in a quite different environment to deal with big data.

First we would like to clarify the difference between privacy and data protection and show how both get a very different place in the EU and the US. Protection of personal data is based on the fundamental right to privacy, but has evolved into a framework of rights and duties which exceeds the right to privacy and has acquired the status of an autonomous fundamental right in itself. Both rights do partially overlap, but function with a different logic. Both set of tools are used in very different ways in the EU and the US.

The general European data protection framework is provided by directive 95/46/EC, but it is rooted in earlier instruments like Convention for the Protection of Individuals with regard to Automatic Processing of Personal data (also known as Convention 108), adopted by the Council of Europe in 1981. The data protection framework had a profound impact on the fundamental right jurisprudence concerning the right to privacy. The right to protection of personal data developed into a fundamental right in itself, distinct from the right to privacy.

This data protection framework provides that all processing of personal data requires a legal ground. In other words, all processing of personal data is regulated and subject to a set of rules guaranteeing the accountability of the processor and the transparency of the processing. The European data protection framework

---


applies to all processing of personal data. Personal data is defined very broadly as “any information relating to an identified or identifiable natural person”. Also the range of activities to which the directive applies is very broad. Processing is defined as “any operation or set of operations that is performed upon personal data, whether or not by automatic means”. This means that whenever data in a big data-context contains information linked to an identifiable natural person, the processing has to be according to the data protection principles and mechanisms have to be implemented to allow data subjects to exercise their rights. The only possibility to escape this framework is by anonymisation of the data.

Directive 95/46 provides principles to which any processing of personal data has to conform, like legitimacy (several grounds for legitimate processing are foreseen, including the consent of the data subject), finality, proportionality and relevance, accuracy, transparency, data subject participation and control, data security. The directive further provides the rights of data subjects, like the right to information about the data processing, to access the data, to object and to rectification. It also specifies the obligations of data controllers, like assuring the confidentiality and the security of the personal data and notifying or prior checking of automated processing to the supervisory authority. The directive foresees a control mechanism through the establishment of independent supervisory authorities. These data protection authorities have powers to investigate, to intervene and to start legal proceedings against violations of the data protection laws. The Commission proposed a new General Data Protection Regulation (GDPR) in 2012 and the review is still ongoing. The draft versions contain generally the same principles, but provide more detailed implementations.

The US framework does not subject all processing of personal data to legal rules guaranteeing more transparency and control for data subjects. Personal data can be freely used unless it is forbidden. The basic structuring of the legal framework is based on opacity tools.

The 4th Amendment to the US Constitution protects people “in their persons, houses, papers, and effects” against the government. Searches are only allowed with a warrant and upon probable cause. This 4th Amendment protection only applies towards the government and not towards private actors. Outside this limited area processing of personal data is in principle allowed, except when specific laws forbid it or subject it to certain rules. Privacy law between private actors was first established through tort law. Four privacy tort actions are recognized in the Second Restatement of Torts and can be considered as opacity tools between private actors, but these have no practical relevance for big data.

This comparison shows a fundamentally different situation in which big data processing using personal data can take place. This US constitutional framework gives free space for such big data processing, as long as no other specific law

---

57 European Commission, 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, 5 April 2012.
provides constraints. The EU framework does only allow unconstrained big data processing with anonymized data. When using personal data, big data processing has to be able to fulfill the requirements of data protection law.

This does not mean that data protection has no place in US law. The growing use of computers and the surveillance scandals from the Nixon and FBI director Hoover-era led to the formulation of Fair Information Practice Principles (FIPP): 58

- There must be no personal data record-keeping systems whose very existence is secret.
- There must be a way for a person to find out what information about the person is in a record and how it is used.
- There must be a way for a person to prevent information about the person that was obtained for one purpose from being used or made available for other purposes without the person’s consent.
- There must be a way for a person to correct or amend a record of identifiable information about the person.
- Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuses of the data.

The first 4 FIPP have to do with transparency, while the last one sets an accountability standard. These FIPP are similar to the principles underlying data protection in Europe, but have only been put into law in specific areas. The Privacy Act of 1974 implements the FIPP in the government and is the main legal framework concerning the treatment of personal information by the federal government. It regulates and restricts the collection, retention and disclosure of personal data. Further does it grant individuals a right of information, access and amendment or correction. 59 Since the 1970s a range of laws containing privacy protection for specific sectors have been established. These laws implement the FIPPs fully or partially and make them applicable on big data practices with data regulated by these laws.

The Federal Trade Commission (FTC) plays an important role in regulating privacy in the private sector. It regulates and supervises market practices and has the authority to enforce trade law through investigatory and litigation powers. Basic consumer protection is provided by the FTC Act, which forbids “unfair or deceptive acts or practices in or affecting commerce”, while the FTC also has the authority to enforce other specific consumer protection laws, like the FCRA or COPPA, and the EU-US Safe Harbor Framework. The FTC has taken up the role

---


of the de facto data protection authority by enforcing privacy policies of companies. The legal status of privacy policies has been ambiguous and enforcement under contract law failed in practice. The FTC has treated violations by a company of its published privacy policy as such a deceptive and in several occasions unfair act. It developed through settlements a common law-like jurisprudence establishing norms concerning transparency, data collection and use, and data security. This jurisprudence evolved towards treating the disrespect of industry standards on these issues as a form of deceptive act. FTC settlements and opinions have therefore become an important source of law. This FTC practice has widely broadened the areas where processing of personal data is subjected to constraints.

7.2 Impact of Data Protection on Big Data Economics

The rules contained in the EU data protection framework as well as in the US FIPP have received heavy criticism from industry and a range of scholars for not being suited for big data. These critics consider it to be an obstacle for technical development and the scientific and economic advantages a wider implementation of big data can bring, or consider it broken and not effective any more to protect privacy in the age of big data. Criticism has been levelled at the notions of personal data versus anonymous data, principles like purpose limitation, data minimisation, and consent as base for legitimate processing of personal data. On the other hand, a range of scholars and the WP29 defend the application of the data protection framework in the big data context and refuse to see enough ground in the fruits of progress arguments in terms of economy, security or science to lower the protection of privacy given by the data protection framework.

The GDPR drafts show some attempts to limit the application of the data protection framework or to lower the obligations in certain circumstances, like the inclusion of pseudonymous data. These contested attempts for legal fine-tuning embody the plea by the critics of the current data protection framework to move

the attention from data collection to a risk-based approach based on the actual use of personal data. These proposals involve a scaled approach through which the application of data protection principles gets modulated.\textsuperscript{63} The WP29 has reacted to this plea with its statement on a risk-based approach\textsuperscript{64} and other recent recommendations. It points to the risk-based elements present in the data protection framework, while making re-interpretations of data protection principles like purpose limitation which are more compatible with this approach.

The Obama administration has taken the initiative to remedy the piecemeal privacy law by an overall consumer privacy regulation, called the Consumer Privacy Bill of Rights.\textsuperscript{65} This Consumer Privacy Bill of Rights gives a wider implementation of the FIPP in the digital economy.

These principles will be further developed through multistakeholder processes in order to develop enforceable Codes of Conduct. The FTC would enforce this Bill. This can happen through a new authority provided by law or through its authority to prohibit deceptive and unfair practices. The Obama administration takes a double approach towards the further development. It prefers to enact the Consumer Privacy Bill of Rights through legislation in order to increase legal certainty, but if Congress does not want to vote this proposal into law, the implementation can anyway go on through the development of codes of conduct.

The law proposal itself has not seen a lot of action the last 2 years in Congress. The privacy multistakeholder processes have resulted in a Code of Conduct for transparency in mobile apps,\textsuperscript{66} while such a process is ongoing concerning the commercial use of facial recognition technology.\textsuperscript{67} Although both affect specific big data practices, the results of this initiative remain limited. Where a Consumer Privacy Bill of Rights would subject commercial big data practices to the FIPP, the situation remains one of piecemeal regulation in distinct laws and FTC enforcement of privacy statements.


\textsuperscript{64}WP29, Statement of the WP29 on the role of a risk-based approach in data protection legal frameworks, 30 May 2014.

\textsuperscript{65}White House, Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy, 23 February 2012.


The last years a lot of policy debate has taken place concerning big data and privacy, reflected in several important reports. One focus were data brokers, the other was specifically on big data and privacy. The Government Accountability Office (GAO), the Committee on Commerce, Science and Transportation in the Senate, and the FTC have investigated the data broker industry and the problems it poses concerning privacy. The reports all conclude that consumers can not exercise rights foreseen in FIPPs towards this industry. The FTC recommends to subject the different branches of this industry to legislation similar to FCRA and to assure transparency, access and amendment for consumers.

Further president Obama launched a Big data review, focused on big data and privacy. It resulted in 2 reports. The first report was made by a working group of senior Administration officials led by John Podesta and resulted from a broad process with stakeholder consultations and academic workshops. This report of the Big Data and privacy working group gives an overview of big data practices in the public and private sector, and points to both the positive gains as the dangers involved. It notes several areas where big data presents challenges like the marketplace, schools, the danger of new forms of discrimination and using data as a public resource. The report makes recommendations like: advance the Consumer Privacy Bill of Rights, pass national data breach legislation, extend privacy protections to non-US persons, ensure data collected on students in school is used for educational purposes, expand technical expertise to stop discrimination at the lead civil rights and consumer protection agencies and amend the Electronic Communications Privacy Act.

Parallel the President’s Council of Advisors for Science and Technology (PCAST) conducted a study of the technological trends underpinning big data, in order to assess the technical feasibilities of different policy approaches. Also this report start with a broad sketch of uses of big data and the possible tradeoffs between privacy, security and convenience. PCAST states that a policy focusing on limiting data collection is not a broadly applicable or scalable strategy. Also because a lot of privacy problems arise after the collection with the fusion of data sources. It argues that the use of data is the place where consequences are produced and the technically most feasible place for protections. Further, some

---

71White House, Big Data: Seizing Opportunities, Preserving Values, 1 May 2014.
72President’s Council of Advisors on Science and Technology, Big Data: A Technological Perspective, White House, 1 May 2014.
techniques for privacy protection used in the past do not seem robust anymore in the context of big data, like anonymization, data deletion (as old data sources can later prove useful in combination with others) or distinguishing the treatment of data from metadata (as metadata can be as much a risk to privacy as the data itself). Also the notice and consent framework is considered unworkable. This framework places the burden of privacy at the individual, while this individual is placed in an unequal position in relation with the provider. The responsibility for using the personal data in accordance with the preferences of the data subject should better be shifted to the provider.

This assessment puts doubt to the robustness of FIPPs, underlying the proposed Consumer Privacy Bill of Rights and other privacy regulations. PCAST still endorses these principles as sound, but states that big data puts effective operationalisation at risk. It suggests several adaptations in line with its recommendation to focus on use of data instead of data collection. Concerning rights meant as consumer empowerments, PCAST recommends to recast these empowerments as obligations of the entity using the data whenever such empowerment has become practically impossible to exercise in a meaningful way.

On both sides of the Atlantic protection of personal data and its relation to big data has become a policy issue. The different starting situations influences the debate on both sides. Also for personal data the EU framework is more restrictive for big data than the US privacy laws. Except in certain sectors, personal data can be freely used in the US where in the EU this is only the case for anonymized data. Result has been the development of strong data economy in the US, with the development of specialised actors in the data value chain, like data brokers, and the development of such data value chains based on personal data, like targeted advertising. The lack of a general protection of personal data in the US incited the FTC to become implicitly a rule maker based on consumer law and the reasonable expectations of the consumer. The Obama administration has made attempts to make privacy protection more general as part of consumer law, but without result till now.

The policy debate shows a similar struggle with the practical implementation of data protection principles or FIPP on big data processing. The underlying question is if the legal mechanisms to ensure transparency about what data processors do with personal data, developed for the information market, are still effective mechanisms in a data economy and if they allow building data value chains. Although the space given to data protection principles vary much, policy makers on both sides of the Atlantic do not question the underlying principles of data protection, but are looking for a more ‘data-friendly’ implementation. This is not directly reflected in the EU policy documents discussed earlier, but rather in the WP29 statements and in the legislative process of the GDPR. The policy documents generally present data protection as an important tool to build consumer trust. However, in the Communication on a data-driven economy the Commission announces that after the adoption of the GDPR it will work on guidance
concerning big data-related problems like on such as data anonymisation and pseudonymisation, data minimisation.\footnote{European Commission, \textit{Towards a thriving data-driven economy}, COM(2014) 442, 2 July 2014.}

Of all 3 legal frameworks the tension between data protection and the realisation of economic opportunities with a data economy remains the most difficult to resolve. In the US the status quo remains the most ‘data-friendly’ solution from a commercial perspective and the federal government proves to be a too weak actor to force change. In the EU the outcome is less clear, but a clear demand from big data companies exists to soften the protection.

8 Public Sector Information and Open Data (Framework 3)


The underlying logic is on both sides the same. Older frameworks of passive transparency, or access to documents on request, get augmented with active transparency and open data policies. Where the passive transparency procedures were tools to enlarge governmental transparency, the active transparency policies are more economically motivated. They look at PSI from a market perspective and try to avoid that public bodies have a distorting effect. Open data policies present a shift to a data economy perspective and include more attention to data quality.

Main regulatory focus of EU policy on public sector information (PSI) has been the review of the directive on the re-use of public sector information or PSI-directive 2003/98/EC, which was realised in 2013. The directive concerns PSI held by public sector bodies in member states. EU-law differentiates access from re-use of PSI, as it has no competence to regulate access to PSI in member states, except
on environmental information. The right of access consists of a right to see and to
take knowledge of the content of documents, but does not imply automatically that
this information can be used without restrictions. Re-use is defined as: “the use by
persons or legal entities of documents held by public sector bodies, for commer-
cial or non-commercial purposes other than the initial purpose within the public
task for which the documents were produced”. In other words, re-use concerns
the further use of the information, after having received knowledge of it. When
public authorities make a further use of information outside their public task, it is
also considered re-use. E.g. the commercialisation of certain data in order to recu-
perate costs, like publishing maps. Specific sectoral rules exist on access and re-
use, like the Inspire-directive.

Aim of the 2003 PSI directive was to create an internal market of PSI. It wants
to assure that all private actors can use PSI in an equal manner. The PSI has to
be available for re-use both for commercial and non-commercial purposes under
the conditions stipulated by the directive. States are not obliged to give access or
to allow re-use, but once the permission for re-use is given it must be done under
equal conditions for all players and in a transparent manner. The conditions linked
to the re-use of documents have to be non-discriminatory for comparable catego-
ries of re-use.

The PSI directive also fitted in the vision of the European Commission on eco-
nomic development of the information society. PSI had to fuel a market of rich
content and therefore become available for such content producers. Therefore
charges for the re-use of PSI have to be limited to the marginal costs incurred for
their reproduction, provision and dissemination and may not include costs linked
to the original collection of data. The directive provides an exception when public
sector bodies are required to generate revenue to cover a substantial part of their
costs and for libraries. Secondly, monopolies by public sector bodies have to be
prevented. Non-discriminatory access and re-use concerns also public sector bod-
ies for activities outside their public tasks. Commercial activities by public sector
bodies outside their public task have to take place under the same market condi-
tions as for private actors.

Where the 2003 PSI-directive fits in the vision of the Internet as content mar-
ket, the revision in 2013 shows a shift towards attention for the data value chain.
The revised PSI-directive provides that when possible PSI is made available “in
open and machine-readable format together with their metadata. Both the format
and the metadata should, in so far as possible, comply with formal open stand-
rards”. This improves semantic interoperability and facilitates the use of the data

---

76 European Parliament and the Council, Directive 2003/98/EC on the re-use of public sector
information (PSI-directive), 17 November 2003, art. 2(4).

77 Janssen, Katleen, The EC Legal Framework for the Availability of Public Sector Spatial Data. An
examination of the criteria for applying the directive on access to environmental information,
the PSI directive and the INSPIRE directive, ICRI, Leuven, 4 December 2009, p. 65.

78 European Parliament and the Council, Directive 2003/98/EC on the re-use of public sector
information (PSI-directive), 17 November 2003, art. 5(1).
in automated and aggregated ways. Other recent policy initiatives concerning open data and e-government also represent this shift. On the one hand by making data more accessible through data portals. The Commission planned in its Communication on Open data\(^79\) in 2011 to set up 2 data portals: the European Union Open Data Portal\(^80\) to make available its own data resources and a pan-European data portal with data from the Commission, member states and public sector bodies.\(^81\) Further it shifted more attention to data quality by supporting projects to enhance semantic interoperability.

The open data policy is also driven by the Commission’s effort to enhance interoperability between public administrations as part of its e-government policy. Through cross-border exchanges of information between member state and EU public administrations it tries to enable European public services. Objective is to aggregate ‘basic’ public services and to make them Europe-wide accessible in cross-border services. The European eGovernment action plan 2011–2015 did set the objectives to have by 2015 a number of key cross-border services available online.\(^82\) Such increased interoperability between public administrations would not only lead to more efficient and effective public administrations, but also have strong impact on the data economy. Open data policies have limited effect when data cannot be linked easily and remains locked in incompatible formats. Interoperability between open data sources turns these sources into big data.

The European Commission developed a European Interoperability Strategy (EIS)\(^83\) and a European Interoperability Framework (EIF),\(^84\) and promotes now the adoption of national interoperability frameworks by member states in line with the EIF.\(^85\) The EIS combines a top-down approach through European policy development and coordination with a bottom-up, sectoral approach through projects. The practical implementation of this sectoral approach is found in the program on Interoperability Solutions for European Public Administrations (the ISA program),\(^86\) supporting activities to facilitate cross-border digital collaboration between public administrations from member states and EU institutions. The top-down approach is further developed in the EIF, which defines an agreed approach to interoperability. It sets principles of and a conceptual model for European


\(^{80}\)https://open-data.europa.eu.

\(^{81}\)http://publicdata.eu.


\(^{86}\)http://ec.europa.eu/isa/.
public services and describes interoperability levels, interoperability agreements and governance.

The revision of the PSI directive figured already in the 2010 Digital Agenda, but the shift in attention towards the use of PSI as resource for a data economy became more visible first in the Communication on Open Data in 2011. The attention for interoperability in the e-government initiatives was also an early sign of attention for data value chains. It shows that e-government initiatives also function as tools for policy learning.

In the US we see similar attention shifts in the PSI policy. The Obama administration has from the start in 2009 given a strong impulse for enlarging the availability and access to public sector information, building upon pre-existing legislation for passive and active transparency.

Passive transparency, the giving access to information on request, is provided by the Freedom of Information Act. Active transparency, the providing of information on the initiative of the government, is regulated by the E-Government Act of 2002 and the Paperwork Reduction Act. The Paperwork Reduction Act dates from 1980, but was strongly revised in 1995 and was also the convenient place to include the framework for an information management and dissemination policy. It prevents agencies to restrict dissemination by using exclusive distribution arrangements, to restrict use, resale or redissemination or to make it subject to fees or royalties, and to ask user fees exceeding the cost of dissemination. Another important element is that copyright protection is not available for the US government. This legal framework led an early foundation for private sector use of PSI and the development of an information market.

The Obama administration added to this legislative framework a policy initiative by the executive branch to give a stronger implementation of open government policy. Rationale behind this policy is on the one hand strengthening democracy by enhancing accountability towards the public and participation of the public. On the other hand the objective is to make the government more effective by strengthening cooperation within the government and with private actors.

This resulted in the Open Government Directive, presenting a policy road map for the implementation of open government by executive departments and agencies. It instructed agencies to make more government information available online in open formats. When deciding about publishing information, the presumption should be in favor of openness, that is “to the extent permitted by law and subject to privacy, confidentiality, security, or other valid restrictions”. The publication of information should preferably be in open formats. An open format
is defined as “one that is platform independent, machine readable, and made available to the public without restrictions that would impede the re-use of that information.” This definition contains elements of legal (no restrictions of re-use) and of technical interoperability (platform independent, machine readable). Objective of the open format is that the information can be “retrieved, downloaded, indexed, and searched by commonly used web search applications”. This focus on open formats shift the attention from information markets to making data useful for more developed data value chains.

A second policy initiative focused on digital government.\(^9\) It lists 4 main principles: an information-centric approach, a shared platform approach, a customer-centric approach and a platform ensuring security and privacy. All point again to making PSI useful in data value chains. The information-centric approach introduces an attention for semantic interoperability. It promotes a shift in thinking about digital information, away from the old approach focused primarily on presentation. An information-centric approach should focus on making data and content accurate, available and secure. It needs to turn unstructured content into structured data and to associate this structured data with valid metadata. Providing this data through web APIs enhances interoperability and makes the data assets widely available. It also supports device-agnostic security and privacy controls, shifting the focus from securing devices to securing data.

The evolution of policies on public sector information are similar in the EU and the US. Both developed PSI as a resource for information markets, but more recently focussed more on making PSI useful for data value chains. Their open data policies evolved from setting up data portals and a focus on quantity of data-sets to a focus on quality of data in terms of interoperability. Non-discriminatory frameworks and licenses improve legal interoperability, attention for open and standard formats and descriptions improve technical and semantic interoperability. The aim to improve public services is present on both sides, but e-government efforts differ due to the specific EU attention for cross-border interoperability.

We can conclude that both e-government and open data policies are important elements of big data policies in the EU and the US. They also present important areas of policy learning, especially on interoperability and on what the construction of data value chains involves.

### 9 Conclusions: Adapting Legal Frameworks to a Data Economy Remains Unfinished Business

In this article we looked at how policy makers digested big data in their ICT- or Internet-related policies. Big data practices depend on the possibility to build data value chains and are therefore very much affected by legal frameworks regulating

access, linking to and use of data. Such data value chains also change the role of the Internet from a market place where content and service providers meet human customers into a space where a lot of non-human actors exchange and process data in real time. We called this a change from an information economy to a data economy, as it changes the level of interactions over the Internet. This change puts stress on existing legal frameworks and can change views on the objectives which these legal frameworks have to reach.

We first looked at how the EU has adapted its policies to the advent of big data. On policy level the EU has developed in its 2014 Communication on a data-driven economy⁹² a profound vision on a data economy, overcoming and deepening the earlier focus on the Internet as marketplace. It has more attention to the role of data and the need to make data interoperable for the creation of data value chain as part of a data economy. This change in focus also raises the question if the legal frameworks affecting such data value chains are still adequate and if this change in policy vision is also translated into new objectives concerning the legal frameworks regulating data flows. In this respect we found a more mixed picture.

We looked at 3 legal frameworks: intellectual property rights, the protection of personal data and the regulation of public sector information. To give a broader perspective we also compared these legal frameworks and policy responses in the US.

The copyright framework in the US has less problems with big data. Comparing the European IPR framework with the US shows some striking differences. First, the US has limited the protection of databases to copyright and has never extended IPR protection purely on grounds of investment. Secondly, its fair use regime proved to be much more technology neutral and adaptable to new technological developments. Both improve legal interoperability of data sources.

Result is that copyright raised no policy debate in the US, while the courts are the main actors in dealing with new technologies in this context. In other countries we see reviews of IPR policies and recommendations to adapt the copyright frameworks with new exceptions. This debate also started in the EU, but remains embedded in and limited by a focus on strengthening the IPR framework. Active policy attention for a data economy by the EU has not led yet to a thorough revision of the IPR framework and more legal interoperability. A revision of the exceptions in the copyright framework to improve the space for data mining is planned, but the database directive remains outside the policy focus.

On both sides of the Atlantic protection of personal data and its relation to big data has become a policy issue. The different starting situations influences the debate on both sides. Also for personal data the EU framework is more restrictive for big data than the US privacy laws. Except in certain sectors, personal data can be freely used in the US where in the EU this is only the case for anonymized data. Result has been the development of strong data economy in the US, with the development of specialised actors in the data value chain, like data brokers, and

---

the development of such data value chains based on personal data, like targeted advertising. The lack of a general protection of personal data in the US incited the FTC to become implicitly a rule maker based on consumer law and the reasonable expectations of the consumer. The Obama administration has made attempts to make privacy protection more general as part of consumer law, but without result till now.

The policy debate shows a similar struggle with the practical implementation of data protection principles or FIPP on big data processing. The underlying question is if the legal mechanisms to ensure transparency about what data processors do with personal data, developed for the information market, are still effective mechanisms in a data economy and if they allow building data value chains. Although the space given to data protection principles varies a lot, policy makers on both sides of the Atlantic do not question the underlying principles of data protection, but are looking for a more ‘data-friendly’ implementation. Of all 3 legal frameworks the tension between data protection and the realisation of economic opportunities with a data economy remains the most difficult to resolve.

The evolution of policies on PSI are similar in the EU and the US. Both developed PSI as a resource for information markets, but recently focussed more on making PSI useful for data value chains. Their open data policies evolved from setting up data portals and a focus on the quantity of datasets to a focus on quality of data in terms of interoperability. Non-discriminatory frameworks and licenses improve legal interoperability, attention for open and standard formats and descriptions improve technical and semantic interoperability. The aim to improve public services is present on both sides, but e-government efforts differ due to the specific EU attention for cross-border interoperability. We can conclude that both e-government and open data policies are important elements of big data policies in the EU and the US. They also present important areas of policy learning, especially on interoperability and on what the construction of data value chains involves.

On the whole we can conclude that the recent EU policy documents reflect an improved understanding of big data and a data economy, but the translation of this vision into new objectives for legal frameworks dealing with data flows proves to be more difficult and shows mixed results.

Bibliography


European Commission, A European strategy on the data value chain, November 2013.


Ira S. Rubinstein, Big Data: The End of Privacy or a New Beginning?, NYU School of Law, Public Law Research Paper No. 12–56.
Janssen, Katleen, The EC Legal Framework for the Availability of Public Sector Spatial Data. An examination of the criteria for applying the directive on access to environmental information, the PSI directive and the INSPIRE directive, ICRI, Leuven, 4 December 2009.


Lokke Moerel. 2014. Big data protection: How to make the draft EU regulation on data protection future proof. Tilburg University.


President’s Council of Advisors on Science and Technology. 2014. Big Data: A Technological Perspective, White House, 1 May 2014.


