Digital Transformation Monitor

Big data: a complex and evolving regulatory framework

January 2017
Big data: a complex and evolving regulatory framework

Given the rapid development of active technologies based on artificial intelligence and big data, the existence of efficient and appropriate European standards and regulations is key to support the competitiveness of EU industries and enterprises. Regulations at EU level must therefore allow for a swifter and flexible uptake of innovative technologies.

A rapidly growing Big data market coming with a wide array of regulatory challenges

Big data: the result of continued investments in data analytics technologies

The accumulation of data both in businesses and on the Web, along with the growing number of open data initiatives have enabled the emergence of the concept of big data.

The Big data trend mainly constitutes the evolution of existing data analytics technologies. There is no clear difference between big data and traditional analytics. However, the following three characteristics are usually associated with the big data trend:

• Volume: the data generated are usually produced in large volumes
• Velocity: data is generated at a high speed pace
• Variety: the data generated usually takes various formats or types including text, video, picture, sound or websites.

With the ever-increasing amount of digital connections, both in terms of number of connections and time spent, the volume, velocity and variety of big data is only going to keep growing.

Big data: numerous opportunities, new investment flows

Recent studies all indicate that the big data market is undergoing an upward trend which is not expected to slow down in the near future. A joint survey conducted by DNV GL and GfK Eurisko provides an insightful glimpse into the big data-related projects of different organizations (see figure 1). The study does not account for a representation of big data adoption across all enterprises globally however it demonstrates that big data is seen as an opportunity for 52% of survey respondents (with less than 5% seeing it as a threat). In addition, it also shows that 51% of firms with more than 1,000 employees plan to increase their investments in big data in the next two to three years (with 76% of all organizations planning to increase or maintain their big data investment).

U.S. giants, the big winners of the data monetization trend

The adoption of big data is booming in the advertising industry. Business models based on big (mainly personal) data are developing rapidly throughout the world. The ecosystem of infrastructures and services enabling the target, collection, storage and processing of personal data is largely dominated by U.S. providers of OTT services, who have succeeded in the monetization of personal data. The personal data of European consumers are therefore largely processed by these global market players. Indeed, Google and Facebook are the largest beneficiaries from ads; their reliance on the use of personal data and thus advertising is evident, with Google and Facebook producing 90% and 95% of their revenues respectively from advertising in 2015.

U.S. giants, the big winners of the data monetization trend
Big data: a complex and evolving regulatory framework

An evolving regulatory framework

A fragmented and heterogeneous regulatory framework

A patchwork of 28 different laws and heterogeneous sets of rules define data protection across EU member states. This leads not only to unequal protection rights for citizens, but also represents significant administrative burdens for business, including the uptake of big data. Common and harmonized guidelines on data usage, rights and quality are crucial for an effective and sustained EU-wide uptake of big data and the digital transformation of industry and businesses.

The agreement on the Commission’s EU data protection reform announced on 15 December 2015 now opens the discussion on national implementation rules. Currently, each country’s national data protection agency is set to define how they will implement the new regulation into their national framework which will likely increase the risk of EU discontinuity and problems of cross border data exchanges.

Moreover, consensus has not yet been reached on the usage and exploitation rights attached to different types of data. Exclusive and non-exclusive approaches are central as they will determine the need for complex fair pricing and oversight mechanisms.

A clear need for a common framework regulating the use of big data at EU level

A lack of common guidelines also results in data being recorded and saved in different taxonomies, formats, and types depending on the entity or the country producing it. This lack of guidelines and shared common taxonomies for metadata curation and integration limits the development of analytical platforms and the digital integration of data flows.

Most public and private data sources also have to face data quality challenges. Hence, increasing data quality is a fundamental success factor for the uptake of big data.

Two set of actions linked to a better and clarified usage of data and to data quality in order to reap the full benefit of big data should be undertaken:

- EU guidelines for companies to make the most of data
- Promotion and implementation of data quality standards

Number of data records lost or stolen in:

- 2014: 1.02 billion
- 2015: 707.5 million

The privacy paradox: an increasing use of Internet services despite the increasing distrust in Internet services

Various surveys have shown that the general public is losing confidence in their ability to gain control over their own personal data. A 2014 study by Pew Research found that 91% of respondents believed they had lost control over their personal data collection and use, and 88% also believed it was very difficult to remove inaccurate data about themselves. Yet, as the figure below demonstrates, despite the lack of trust in Internet services, the population is still using them. Ultimately, online services have become such an integral part of our daily lives that even low levels of trust level cannot prevent their use. Stakeholders thus face the challenges when trying to regain consumer confidence in these services.

The adoption of Big data solutions provides many opportunities. However, the increasing number of alarming reports indicating significant data breaches and the leak of such data also raises many questions. For instance, the monetisation of personal data is key for advertising businesses but a balance needs to be reached to ensure privacy, rights and the secure storage of data.

Security and privacy issues: a growing concern

Figure 3: World’s largest security breaches, Oct 2016

Figure 2: Comparison between the use and trust levels of diverse online services: (2015 survey)
EU guidelines for companies to make the most of data

The development of common guidelines is an essential prerequisite for the uptake of big data. Companies and industry players must be provided with the certainty that they can safely and securely generate value from data without breaching data protection laws or stepping beyond the boundaries of public acceptance. Without clear guidelines EU companies will both face challenges to use data as a business driver and expose themselves to risks. This will severely hinder the EU's competitiveness, innovation drive and economic growth.

Appointing a Chief Data Officer per member state to oversee the implementation of open data initiatives would be a further step in the right direction. An independent advisory panel made up of national chief data officers could forge consensus around a cohesive vision and strategy for capturing the full benefits of data-driven innovation in Europe, guaranteeing similar and fair market conditions for all market players while protecting consumers, workers and business investment.

In the healthcare industry, the development of multipurpose EU consent templates should be promoted to enable the creation of pan-European data sets and to encourage the use and exchange of Electronic Health records (EHR).

Promote data quality

Promotion and implementation of data quality standards is a strong enabling factor across industries as quality is linked to credibility and trustworthiness. A new generation of common curation methodologies and technologies for complex usages such as clinical trials, anti-fraud or energy savings has to be rapidly promoted.

Expected impact:

Pan-European guidelines on data usage and quality will further support the Digital Single Market by ensuring a harmonised European framework. Data accessibility is the primary enabler of data aggregation. In order for companies to gather and use data they must overcome the current barriers linked to the lack of clear consent and exploitation guidelines.

In this sense, the origin, rights and consent attached to different types of data must be well defined. Otherwise, SMEs will have to depend on data gathered outside the EU and face higher costs for obtaining data.

A clear definition of data usage and rights will boost the potential of pan-European Data Lakes. A data lake is a subject-specific repository for large quantities and varieties of data, both structured and unstructured. The data lake accepts input from various sources and can preserve both the original data fidelity and the lineage of data transformations. The lakes could help resolve the issue of accessibility and data integration for European businesses and citizens in different industries and application areas.

This would enable stakeholders along the value chain in different industries to exchange data within a specific protocol in a rapid and secure manner. Ensuring high quality data will bring more competitiveness to SMEs. For example, in the healthcare industry, so far, only established players have the resources to clean the data and offer datasets of sufficient quality for clinical trials.

Promoting data quality and curation standards will offer SMEs the chance to compete with larger players and bring new innovations to the market.
Big data: a complex and evolving regulatory framework

**Figure 5: Key changes in the European legal framework for the GDPR**

| Territorial scope                                                                 | • The obligations apply to all providers operating in Europe.  
                                 | • If they don’t have a legal presence, they must have a dedicated and financially sound representative in Europe. |
|----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Profile and big data                                                             | • Personal data must be collected with a clear initial goal and only for this purpose.  
                                 | • The directive regulates the use and reuse of non-sensitive personal data.  
                                 | • Pseudonymised data is also personal data. |
| For the user (the ‘data subject’)                                               | • Implementation of the right to be forgotten has been further reinforced.  
                                 | • User’s access to their file, including:  
                                 | - duration of data retention  
                                 | - details of the data recipients outside the EU  
                                 | - details of applicable regulations  
                                 | • **Big data**: explain the logic, meaning and consequences of the decisions taken by the processing when it is automated (profiling) and its purpose is not obvious.  
                                 | • Data portability: users must be able to request their data and have it provided in a usable format |

**General Data Protection Regulation**

In the context of the new General Data Protection Regulation, it is necessary to closely follow its implementation and interpretation by Member States. National barriers may be set, should Member States differ in their understanding of the regulation. This would be a major setback to the uptake of Big Data.

In the case of autonomous car, a mean of transport that requires to go through borders, a common rules of understanding must be shared by all Member States to ensure sufficient common standards.

Similarly, in the healthcare industry, the societal challenge need cross-border collaboration to share patient and cohort data, R&D progress or clinical trials. Clarification of article 83 may be needed, as well as specific rules tailored to the characteristics of the healthcare and pharma industry. The implementation of the GDPR on file and data exchange must therefore be closely monitored at EU level to enable the creation of a true pan-European data field.

**The GDPR and implications for the Safe Harbour agreement**

The extraterritorial scope of GDPR implies that U.S. providers will have to apply European data protection rules whenever they use European consumers’ personal information.

This led the European Commission to question whether the Safe Harbour provided adequate protection for EU citizens. In October 2015 the final judgment of the Court of Justice of the European Union (CJEU) ruled invalid the Safe Harbour data protection agreement between Europe and the United States.

**The Privacy Shield: Replacing the Safe Harbour**

**Stronger obligations on the U.S. to protect Europeans’ personal data**

Adopted by the EC: 
**June 2016**

Scheduled for review: 
**Summer 2017**

Regulations as the key to pave the way for more innovative solutions

The lack of clear regulations and legislations at EU level is without a doubt a barrier to growth.

This lack of an appropriate framework leads to the lack of trust of European citizens in big data solutions and therefore prevents its further adoption. The question of trust is a central challenge to all the stakeholders involved in the industry and is crucial for the effective development and uptake of Big Data.

As the backbone and underlying element at all stages of digital development, trust in the use and management of data needs to be built in and reinforced among all stakeholders and industry players. The creation of a common European framework regulating the use of big data would certainly constitute a big step helping to build trust and accelerate the uptake of the Big data trend.

**Moving Forward**

1. Ensure smooth transposition of the GDPR into EU Member States
2. Strengthen initiatives to regain consumer trust on personal data
3. Set in motion a plan to consider a global regulation framework
About the Digital Transformation Monitor

The Digital Transformation Monitor aims to foster the knowledge base on the state of play and evolution of digital transformation in Europe. The site provides a monitoring mechanism to examine key trends in digital transformation. It offers a unique insight into statistics and initiatives to support digital transformation, as well as reports on key industrial and technological opportunities, challenges and policy initiatives related to digital transformation.


This report was prepared for the European Commission, Directorate-General Internal Market, Industry, Entrepreneurship and SMEs; Directorate F. Innovation and Advanced Manufacturing; Unit F/3 KETs, Digital Manufacturing and Interoperability by the consortium composed of PwC, CARSA, IDATE and ESN, under the contract Digital Entrepreneurship Monitor (EASME/COSME/2014/004)

Authors: Vincent Bonneau & Soichi Nakajima, IDATE; Laurent Probst, Bertrand Pedersen & Olivia-Kelly Lonkeu, PwC

DISCLAIMER – The information and views set out in this publication are those of the author(s) and should not be considered as the official opinions or statements of the European Commission. The Commission does not guarantee the accuracy of the data included in this publication. Neither the Commission nor any person acting on the Commission’s behalf may be held responsible for the use which might be made of the information contained in this publication. © 2017 – European Union. All rights reserved.