

Since it served as the model for the provision, the Commission's practice in competition law can provide some idea of the extent to which the Commission will exploit the limits for fines given by the regulation. Council Regulation 1/2003 provides for even heavier fines – up to ten per cent of an enterprise's turnover (Art. 23 (2)) – for infringements of EU competition law. So far, however, the fines have not been even close to that threshold. Of the ten highest fines ever imposed, the highest (Pilkington) reached about 6.5 per cent of the enterprise's annual turnover, or 65 per cent of the maximum fine possible, and this was only after an earlier order to dissolve the same cartel was not followed. By contrast, six of the fines were less than three per cent of annual turnover.²⁵

Summary

Achieving harmonisation is perhaps the central goal of the ongoing reform of the European Data Protection Framework. Harmonisation requires not only the harmonisation of substantial law but also of administrative practices. This has been recognised by the Commission, which accordingly provided for a consistency mechanism in its proposed regulation. The consistency mechanism, in its current form, follows a carrot and stick approach.

²⁵ See the statistics available at <http://ec.europa.eu/competition/cartels/statistics/statistics.pdf>, section 1.6.

The EDPB gives national supervisory authorities a platform for voluntary cooperation. This is, however, backed by powers of the Commission which allow it to effectively intervene even in individual cases and not only to ensure the legality of national supervisory authorities' decisions, but also to determine what measures might be appropriate in an individual case. Though much criticised, such powers are necessary to achieve harmonisation, which must include the harmonisation of administrative practices. These powers can arguably be exercised by the Commission itself without compromising the existence of an independent supervisory authority at the European level, as the Commission is also subject to the scrutiny of the EDPS. Alternatively, the supervisory powers could be vested in the European Data Protection Supervisor or a new independent agency at the European level. The consistency mechanism is rightly flanked by measures to ensure that supervisory authorities have the necessary resources to discharge their duties.

Administrative sanctions are also needed. The scale of the sanctions proposed by the Commission is far less intimidating than many critics wish to portray. The sanctions are governed by the principle of proportionality and are modelled on the fines developed in competition law. Experience there shows that the maximum fine will be rarely, if ever, imposed. Overall, it is likely that the reform will be a big step towards harmonising data protection within the internal market.

Laurits Christensen and Federico Etro

Big Data, the Cloud and the EU Regulation on Data Protection

The recent increase in the creation and storage of ever larger quantities of data, referred to as "big data", is expected to enhance the productivity of the global economy. Indeed, it is as if a new factor of production had been constructed. Moreover, the use of big data is going to affect the way firms, academic institutions and consumers do business, make discoveries and interact with each other. It is estimated that people around the world generate more than 2 exabytes (i.e. quintillion bytes) of unstructured data (that is, data that lack a predefined model, such as search engine queries, posts on Twitter, "likes" on Facebook and so on) every day, and organisations generate and store even more exabytes of structured data every year. In most cases, people and organisations do not know what to do with all these data; consumers do not know how their own data are processed and protected; organisations need to develop

tools to process, store and transmit these data; and institutions must regulate data protection. All of this creates a set of interesting issues: what is the economic importance of big data? What is the appropriate role for cloud computing? What is the proper way to protect data?

From an economic perspective, it is important to understand how the accumulation of data, which represent a sort of new factor of production, can affect growth and generate added value across society. A typical example of the importance of big data is related to consumer behaviour online, whose importance has been rapidly increasing over time and is, in turn, a source of new business and trade online. This and other new possibilities are enhanced by the rapidly declining costs of storing data. Cloud computing has contributed more than any other technology to the diffusion

and use of large-scale datasets and to the spread of the benefits of the newly gained efficiencies across companies and to consumers relying on cloud computing solutions.

The use of cloud services to accumulate data implies that the volume of new data will increase exponentially in the years to come. Cisco forecasts that the annual volume of global IP traffic will exceed a zettabyte (a thousand exabytes) in 2015. The velocity of data collection is similarly increasing. Remarkably, 90 per cent of the volume of global data was generated in the last couple of years, and data creation is expected to be 44 times larger in 2020 relative to ten years earlier. This growth derives from a variety of unstructured data that are being accumulated today. The economy is going to benefit from this, with new businesses to be created based on big data and cloud computing. This is possible thanks to two main factors. The first is technological: novel computing methodologies help firms to understand and use data by means of machine learning and analytical tools. The second is economic: the hardware needed to store and process data has become incredibly inexpensive.

As a general matter, big data is expected to benefit the economy directly through more efficient marketing, more efficient pricing, more targeted product development and the development of new businesses (which is always a source of more competition and efficiency). There has been significant discussion of how big data will allow companies to become more efficient in their marketing efforts.¹ The ever increasing quantities of data are allowing firms to better understand what economists call “customer heterogeneity” (put simply, the fact that customers are not all alike and can differ in ways that are important to companies). The power of big data is increasingly allowing companies to deploy their marketing dollars in ways that increase the return on their advertising investment.

Big data is also allowing companies to price their products in ways that recognise customer heterogeneity. For example, there has been an increasing level of experimentation by companies in varying prices based on the outlet through which the product is sold (think Priceline vs. Expedia), the day of the week, or even the hour of the day. This variation in pricing allows companies and consumers to sort themselves according to their specific preferences. Those who are willing to invest time in searching for the lowest prices will be rewarded for their efforts.

Finally, big data holds the promise of improving product development by allowing companies to know more about consumer preferences without having to undertake survey

¹ See e.g. A. Ignatius (ed.): Spotlight on the Future of Advertising, in: Harvard Business Review, Vol. 91, No. 3, 2013, pp. 59-89.

research. The explosion of data allows them to gain this knowledge through careful analysis of consumer choices around real products and pricing. As firms learn more about consumer preferences within existing product lines, they can better target their future offerings in ways that cater to consumers’ revealed preferences. This is likely to be particularly true in the service sector where new product development time can be considerably shorter than in other sectors (e.g. automobiles).

These factors are likely to contribute to increased measures of productivity in traditional economic assessments. However, the full economic value of big data is hard to measure because it creates indirect benefits. The point can be explained with an example. The use of personal data for diagnostic purposes and to avoid duplicative testing in the healthcare sector clearly produces a real economic benefit. But because the value created does not involve explicit market transactions, attributing this benefit directly to data involves some inspired approximation. More than just a hypothetical, recent research from Microsoft Research Labs and Stanford and Columbia Universities has found that by analysing large volumes of data, they could identify previously unknown adverse pharmaceutical drug interactions. This research may be just the tip of the iceberg for improvements in health treatments. As the amount of data available to analyse health outcomes increases (either formally through drug studies or informally through big data), it will be possible to improve treatment regimens and make the healthcare sector more efficient, benefitting individuals and society more broadly. Of course, this example also points to complex issues concerning data privacy and data protection, which present new problems for regulation.

While the economic benefits from big data and cloud technologies are potentially large, privacy groups and consumer advocates have raised alarms over how personal data are collected and processed. They worry as well about the potential to learn details about people that they may wish to keep private. A recent analysis of Facebook data by a research team in Cambridge revealed that researchers could predict sexual orientation with a high degree of accuracy based on seemingly generic information revealed through Facebook pages. Bio-ethicists similarly worry that it will become possible to learn about individuals’ current (and probable future) health status by virtue of information revealed through the myriad interactions that create big data.

In response to these and other similar concerns about data privacy and protection, the EU has proposed a new Data Protection Regulation, with the aim of protecting individuals with regard to the processing of personal data. The new regulation seeks to create a single set of rules across the EU and introduces a number of new requirements for busi-

nesses. As the privacy debate continues, an important distinction will arise as to whether one should prevent the collection and analysis of data (the current approach favoured by the EU) or allow it and instead focus regulation on ensuring that the results are not used in an illegal or unethical manner. For example, should there be a prohibition on the collection of data that could, if analysed in particular ways, allow one to predict with reasonable accuracy which people are likely to develop cancer? Such a prohibition would preserve the privacy of the individuals and ensure that companies could not use that information inappropriately, but it would also prevent these individuals from knowing that they were at risk, which would allow them to take appropriate health precautions.

In the face of this debate regarding how to strike an appropriate balance between maintaining privacy and realising the economic benefits from these new areas of research, economic quantification of the potential benefits and costs can help inform the discussion. Recent research on the benefits from cloud computing and the costs of complying with the proposed regulations offers insights into how the benefits of new technology can be curtailed or eliminated through regulation. In prior research on the benefits of cloud computing, it has been shown that this new technology could generate benefits in terms of increased employment and economic growth.² This research showed the clear promise of the decreasing costs of computing and storage. But it is also in response to the growth of cloud computing and big data that the EU has proposed new data privacy regulations.

In related research with Greg Rafert and Andrea Colciago, we have shown that the cost of complying with the new data privacy requirements in the proposed data protection regulation is expected to have adverse effects on growth and employment, in particular among small and medium-sized enterprises (SMEs).³ This research analyses the compliance costs associated with the direct application of the new regulation as well as the indirect effects on job growth and business creation. Compliance with the proposed regulation poses a number of challenges for firms. The first challenge concerns the design of systems and procedures for data protection. In particular, under the proposed regulation, firms must develop data management systems that allow for greater flexibility, such as the right to data portability (i.e. the right to transfer data from one electronic processing system to another) as well as the right of data subjects

(identified natural persons) to obtain their personal data in a structured, commonly used electronic format. Additionally, data protection impact assessments must be incorporated into IT project management so that firms can identify and mitigate specific risks associated with the processing of personal data.

Another major challenge is the designation of a data protection officer (DPO). This obligation will apply to all public sector bodies and enterprises with 250 or more employees, as well as to firms whose core activity involves the monitoring of data subjects. The controller (the entity that determines the purposes, conditions and means of the processing of personal data) and the processor (the entity that actually processes personal data on behalf of the controller) will be subject to different obligations and possibly also to different supervisory authorities (which could create useless duplication costs).⁴ The controller and the processor will also have to ensure that the DPO is involved in all issues that relate to the protection of personal data and maintain detailed documentation on all processing operations. The compulsory notification of any data breach to the supervisory authority within 24 hours and to the data subjects without undue delay – which is extremely demanding, especially for non-serious data breaches – will lead to substantial compliance costs for firms. Several additional articles in the regulation will also result in additional costs, depending on the type and amount of information processed.

However, it should be noted that some of the proposed articles within the legislation will reduce costs for firms. For example, the “one-stop-shop” principle reduces some compliance costs by ensuring that data controllers and data processors that operate across countries are typically regulated by a single supervisory authority, though this is not the case for companies that happen to be both data controllers and data processors in different countries (for instance, cloud computing providers). Moreover, binding corporate rules will potentially reduce legal ambiguity surrounding data transfers, and joint operations on the part of supervisory authorities will reduce bureaucratic burdens. The ongoing effort to promote secure data transfers is important, as this is crucial for the development and the diffusion of cloud computing; however, more needs to be done, such as supporting and standardising the stronger and more transparent protection of data that are transferred outside the EU for cloud computing services. Both the costs and the benefits

2 F. Etro: The Economic Impact of Cloud Computing on Business Creation, Employment and Output in the E.U., in: *Review of Business and Economics*, Vol. 54, No. 2, 2009, pp. 179-208.

3 For technical details, see L. Christensen, A. Colciago, F. Etro, G. Rafert: *The Impact of the Data Protection Regulation in the E.U.*, Intertec Policy Paper, 2013.

4 Problems of duplication are reduced in many cases with this new regulation, but not for entities that happen to be both controllers and processors. A clearer distinction between the status of controller and the status of processor would be useful to identify which supervisory authority has jurisdiction. The controller should be the entity that determines the reason why data are processed and the processor the entity that determines how the data are actually processed.

Table 1
Summary of expected fixed and variable costs from the data protection regulation by article

Article	Fixed annual cost (euros)				Variable annual cost (euros)			
	Wholesale and retail trade	Hotels and restaurants	Transport, storage and communication	Real estate, renting and business activities	Wholesale and retail trade	Hotels and restaurants	Transport, storage and communication	Real estate, renting and business activities
23, 33					1,462	688	1,056	2,236
28, 35, 36, 37	0	0	539	2,763	0	0	56	254
31, 32	100	100	100	100	1,005	446	442	1,094
12, 17, 18	97	46	41	105	185	52	103	207
7	418	627	173	212				
11	219	103	92	237				
5, 19, 20					1,366	2,049	565	694
43					-1,128	-531	-471	-1,218
8	124	0	43	30				
22	0	0	0	0	0	0	0	0
24, 25	151	71	63	164				
34					156	73	65	168
55, 56, 57, 58, 76					-100	-47	-42	-108
77					303	143	127	327
79	0	0	0	0	0	0	0	0
Total	1,109	947	1,051	3,611	3,249	2,873	1,901	3,654
% of IT budget	6	5	6	20	18	16	11	20

Note: No firms related to manufacturing are thought to be impacted and all costs for this sector will be zero.

Source: Authors' elaboration.

of the proposed regulation have been taken into consideration in our analysis. Finally, it is important to note that we did not take into consideration the expected costs associated with the administrative sanctions, whose homogenous application to all companies (without distinction between intentional and unintentional harm) may create an unfair and disproportionate burden on SMEs that fail to comply with the regulation for reasons other than repeated negligence.

We have simulated the impact of the new regulation on the process of business and job creation, estimating first the likely costs and benefits created by the proposed regulation and then using a dynamic stochastic general equilibrium model. Our estimates of the average expected costs and benefits of compliance for SMEs are summarised in Table 1 (divided by macrosectors and expressed in terms of average net cost per firm). Our results suggest that the net costs are large – indeed, larger than what could be expected, for instance, from the evaluation of the Impact Assessment prepared by the European Commission. The percentage of firms impacted as well as the average expected costs and benefits were estimated for each of the 15 article groups deemed important in the EU Data Protection Regulation.

We calculated an average annual IT budget of €18,000 and then used this value to calculate total costs as a per cent of IT budget; these estimates are also provided in Table 1.

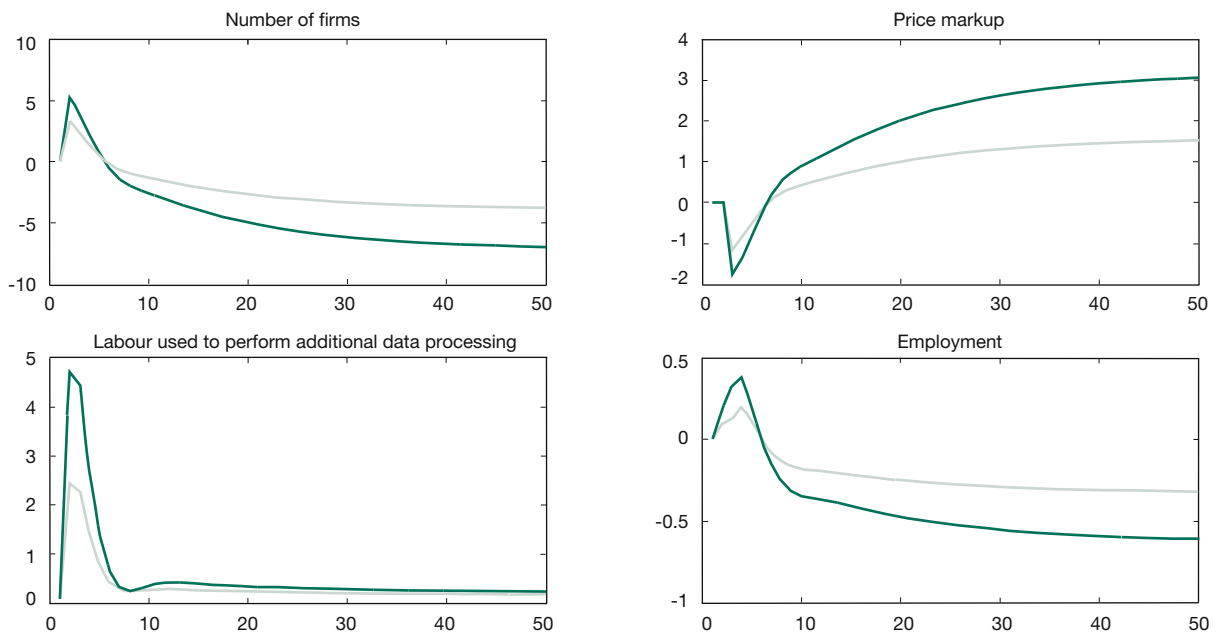
On the basis of the above estimates, we then simulated the impact in a macroeconomic model whose production structure is based on Etro and Colciago,⁵ augmented with a description of the labour market with search frictions and endogenous unemployment. The economy features four types of firms: the producers of intermediate goods, the final good producer, the producers of IT material and the providers of data management services. The intermediate goods industry features many sectors where the dynamics of the number of market competitors is endogenous. In this industry, firms face a sunk cost of entry into the market, which they decide to incur if sufficiently compensated by the expectation of future profits. Goods are produced using labour and IT, which can be interpreted as hardware but also as the stock of data stored at each firm. The industry

⁵ F. Etro, A. Colciago: Endogenous Market Structure and the Business Cycle, in: *The Economic Journal*, Vol. 120, No. 549, 2010, pp. 1201-33.

Figure 1

Simulated impact of the legislation on data protection in the EU 27: real estate, renting and business activities sector

Percentage deviation



Note: dark green line: baseline impact; light green line: low impact scenario.

Source: L. Christensen, A. Colciago, F. Etro, G. Raftoyiannis: The Impact of the Data Protection Regulation in the E.U., Intertec Policy Paper, 2013.

producing IT adopts physical capital as the only input, while in the industry providing data management services, the input is labour. The labour market is characterised by the frictions related to job search and matching. In the intermediate goods industry, both new firms and incumbent firms need to hire workers from the pool of unemployed agents who are looking for a job. They also need to set up a stock of IT before starting production. Similarly, the industry providing data management services faces labour market frictions.

The model counterpart of the introduction of the EU Data Protection Regulation can be illustrated as follows. In order to mimic the need to install a DPO, we have assumed that the intermediate goods producers will incur a period fixed cost. The designation of a DPO can be regarded as a fixed cost, since it does not scale with the size of the firm nor with the number of data records. On the other hand, the development of a data management system is a variable cost for the firm which depends on the amount of data processed or more generally on the number of projects currently being developed at a firm. For this reason, the model counterpart of this requirement is an increase in the units of data management services necessary to deal with each unit of information involved in the production process. Since the introduction of the EU Data Protection Regulation represents a permanent shock to the cost function of firms, we

computed a transition from the pre-reform steady state to the post-reform steady state of the economy we have just described.

The simulation under two scenarios shows a substantial negative impact of the introduction of the EU Data Protection Regulation on business creation and employment under both scenarios. Among the industrial macrosectors that we considered, the one most severely affected by the regulation is the real estate, renting and business activities sector (see Figure 1), which displays a long-run reduction in employment ranging from 0.2 to 0.6 per cent, together with a reduction in the number of market competitors ranging from three to five per cent. The reduction in employment and the number of operating firms is particularly severe in those sectors where compliance with the EU Data Protection Regulation will imply higher fixed operating costs for firms. For example, the effect is stronger in sectors in which a large fraction of firms will be required to designate a DPO.

These results suggest that much or all of the potential benefit from big data and cloud computing could be undone via regulation. If countries wish to benefit from the big data and cloud computing revolution, finding ways to balance privacy concerns against the expected economic benefits from both will be paramount.