

# Editorial

Everyone who has followed a course on statistics or empirical (quantitative) research methods knows how hard it is to put together a dataset that is representative, to design research that can be replicated, to set up a questionnaire or other tools for gathering data that are neutral and unbiased and to find correlations that are valid, significant and have meaning. Not for nothing, most academic papers based on empirical research in, for example, sociology, psychology and social (behavioural) sciences, often consist for more than half of a description of the research methods, the limitations of the research and the insecurities involved with the research findings.

AI, Big Data and profiling thrive on collecting, analysing and using large amounts of data and are often based on analytical tools that are grounded in basic statistics. When gathering data, the GDPR applies, at least when it concerns 'personal data'. When data-driven applications and technologies are used in practice, there is a variety of different legal instruments that apply, such as anti-discrimination law, tort law and the various human and fundamental rights instruments, at least when the applications have a direct effect on natural persons and their interests. But the methods for analysing data themselves are barely regulated and the computer scientists and programmers operating algorithmic data-analytics are not always aware of even the most basic standards of empirical research methods and statistics.

Consequently, actors involved with data analytics and profiling often believe that data themselves are neutral, while they are a human construct (*datum*, meaning *given* may be a somewhat misleading term); think that a dataset can be updated at will, while there is often no way of telling whether differences between old and new data arise from changes in reality or changes in the research methodology; perceive the categorisation of data purely from a utility perspective, while categories determine to a large extent how data are perceived and understood; describe algorithms as neutral analytical instruments, while they are decision trees based on human categorisations, attribution of weights and decisions; etc.

The General Data Protection Regulation (GDPR)<sup>1</sup> gives little guidance on this point. When data are gathered, Articles 5(1)(d) and 16 GDPR require that these are accurate, kept up to date and 'complete'. Article 22 of the GDPR specifies that, at least when profiling or automatic decision making has significant effects, there should be a human assessing whether the general profile, based on statistical correlations, applies to the specific situation at hand. Only recital 71 GDPR gives some guidance on the statistical research methodology that should be used:

---

DOI: 10.21552/edpl/2019/4/3

1 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) [2016] OJ L 119/1.

In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect.

No further guidance is provided in the GDPR on what may be considered appropriate mathematical or statistical procedures. A starting point for understanding this rule can be found in Article 338 of the Treaty on the Functioning of the European Union,<sup>2</sup> which holds:

1. Without prejudice to Article 5 of the Protocol on the Statute of the European System of Central Banks and of the European Central Bank, the European Parliament and the Council, acting in accordance with the ordinary legislative procedure, shall adopt measures for the production of statistics where necessary for the performance of the activities of the Union. 2. The production of Union statistics shall conform to impartiality, reliability, objectivity, scientific independence, cost-effectiveness and statistical confidentiality; it shall not entail excessive burdens on economic operators.

These principles are elaborated in the Regulation on European Statistics,<sup>3</sup> which specifies standards comparable to those set out by the United Nation's General Assembly<sup>4</sup> and by the European Statistical System Committee.<sup>5</sup> Such instruments specify principles such as:

1. *Independence*: There should be independence from external (policy or political) interference in developing, producing and disseminating statistics.
2. *Competence*: The competence of the head and employees of the department or team involved in statistical analysis should be beyond dispute.

2 Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C 326/47.

3 Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities (Text with relevance for the EEA and for Switzerland) [2009] OJ L 87/164.

4 Resolution adopted by the General Assembly on 29 January 2014 [*without reference to a Main Committee (A/68/L.36 and Add.1)*] 68/261. United Nations Economic and Social Council, 'Fundamental Principles of Official Statistics' (28 October 2013) E/RES/2013/21 <<https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf>> accessed 9 December 2019.

5 Eurostat, 'The European Statistics Code of Practice' (16 November 2017) <[http://ec.europa.eu/eurostat/documents/64157/4392716/Revised\\_CoP\\_Nov\\_2017.pdf](http://ec.europa.eu/eurostat/documents/64157/4392716/Revised_CoP_Nov_2017.pdf)> accessed 9 December 2019.

3. *Quality oversight*: Organizations should monitor the quality of their statistical analysis, evaluate their work and consult external experts when appropriate.

4. *Objectivity*:

- a. Data and statistics are compiled on an objective basis determined by statistical considerations.
- b. Choices of sources and statistical methods as well as the use of data are informed by statistical considerations.
- c. Errors discovered in the data or statistics are corrected at the earliest possible date and publicized.
- d. Advance notice is given on major revisions or changes in methodologies.
- e. Statistical outcomes and decisions are objective and non-partisan.

5. *Quality*:

- a. Procedures are in place to ensure that standard concepts, definitions and classifications are consistently applied throughout the organization.
- b. The methods used for gathering data are regularly evaluated and adjusted if necessary in order to ensure high quality statistics.
- c. When organisations share data about the outcomes of data analytics, they ensure that their systems, standards and methodologies are aligned.

6. *Validation*:

- a. Methods for gathering data are systematically tested prior to the data collection.
- b. Survey designs, sample selections and estimation methods are well based and regularly reviewed and revised as required.

7. *Reality and comparability*:

- a. Source data, intermediate results and statistical outputs are regularly assessed and validated.
- b. Sampling errors and non-sampling errors are measured and systematically documented.
- c. Revisions are regularly analysed in order to improve statistical processes.

8. *Accountability*:

- a. Statistics and the corresponding metadata are presented, and archived, in a form that facilitates proper interpretation and meaningful comparisons.
- b. Information on the methods and procedures used is publicly available.
- c. Access to microdata is allowed for research purposes and is subject to specific rules or protocols.

Although not all of these principles can be applied one-on-one to the Big Data context, it might be worthwhile to develop a number of statistical standards and principles for large scale data analytics. Although many of the approaches, methods and applications have changed radically in the Big Data context, at the same time, they have not. The basic statistical requirements, concerns and principles still hold sway over

these processes, or at least should. Many of the problems with biases, discrimination and ineffectiveness of Big Data technologies stem from applying inappropriate statistical methods and procedures. It would be good if the European Data Protection Board (EDPB) would issue an opinion on this point, setting out what should be considered appropriate mathematical or statistical procedures in the context of profiling, AI and Big Data.

The effectiveness of large scale data-driven processes is also the theme of this edition's opinions, penned by Susan Landau and Min-Young Choi. They discuss the efficacy of two data-driven technologies/applications in the United States and South Korea respectively. They are both sceptical that the applications they have studied are indeed effective, which I think may hold true for many more of the hyped data-driven technologies. One could only hope that Landau's simple but powerful statement 'If it isn't efficacious, don't do it' will be on the minds of policy and decision-makers both in the private and the public sector when deciding on the introduction of yet another data-driven technology or application.

For practical reasons, we have decided to postpone the publication of the best five submissions to EDPL's Young Scholar Award to next edition (1/2020). If you want to see the three best young scholars present their work, please come to EDPL's Young Scholar Award session at the CPDP conference, in Brussels on 23 January 2020. This edition contains three top articles by internationally renowned academics. Ronan Ó Fathaigh and Joris van Hoboken discuss the European Regulation of Smartphone Ecosystems, António Manuel Barreto E. Menezes Cordeiro analyses civil liability for unlawful data processing under the GDPR and Leon Trakman, Robert Walters and Bruno Zeller have done a comparative study on tort and data protection law in common law countries.

As always, special mention should be made of the reports section led by Mark Cole. Angelica Fernandez analyses EDPB's most recent opinion on Standard Contractual Clauses for Processors. In the latest entries in our GDPR Implementation Series, András Jóri discusses the GDPR application in the Hungarian context; Menezes Cordeiro sheds light on some controversial aspects of the 'implementation' in Portugal which will likely create legal problems in the future; and Matúš Mesarčík discusses legal changes in Slovakia and concludes, *inter alia*, that the strange scope of the Slovakian implementing law and the literal copying of GDPR provisions seem confusing for practitioners and render the majority of the Slovak Data Protection Act 2018, inapplicable to common data processing operations in many cases. Tobias Raab informs the reader about developments in Germany over video surveillance and facial recognition, Teresa Quintel discusses developments over facial recognition in Sweden and finally, in our Practitioner's Corner, Laurens Vandercruysse, Caroline Buts and Michaël Dooms discuss the problem of data management in smart cities.

Our case note section, led by Maja Brkan and Tijmen Wisman is packed with four contributions that are very much worthwhile reading, all regarding EU Court of Justice ju-

risprudence. Silvia De Conca discusses the *GC et al v CNIL* case for those interested in the continuing saga about the right to be forgotten; Susanna Lindroos-Hovinheimo analyses the *Deutsche Post* case, about the clash between access to data by governmental institutions and the right of data subjects to keep their data private; Paolo Cavaliere engages with the *Glawischnig-Piesczek v Facebook Ireland* case, about the liability rules specified by the e-Commerce Directive; and Primož Gorkič discusses the *Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW e.v.* case, inter alia giving further guidance on the notion of (joint) controllership in the GDPR.

The book review section led by Gloria González Fuster contains one book review by Professor Fuster herself, in which she discusses Guimaraes's *Global Technology and Legal Theory*, and another review by Gianclaudio Malgieri on Hildebrandt's *Law for Computer Scientists and Other Folk*.

Finally, we have incorporated two in-depth interviews with Rohit Chopra, Commissioner for the US Federal Trade Commission, and Alexandre Enaygues, Head of Data Privacy Europe at Novartis.

One final announcement. As most of the readers will know by now, Giovanni Butarelli passed away in Italy on 20 August 2019. From January 2009 to December 2014, he served as Assistant Supervisor of the European Data Protection Supervisor (EDPS); subsequently, he was appointed the European Data Protection Supervisor for a term of five years. Giovanni will be missed for his intelligence, for his global leadership in the field of data protection and most of all, for his warm, friendly and inspiring personality. On behalf of the board of EDPL, we wish his family, friends and loved ones all the strength they need in the coming period. At the same time, we are happy that Giovanni is succeeded by his friend and colleague Wojciech Wiewiórowski, former Assistant European Data Protection Supervisor and also a board member to EDPL. Although these are enormous shoes to fill, we are certain that Wojciech will strengthen EDPS's leadership in the data protection community of the EU and beyond.

For those interested in submitting an article, report, case note or book review, please e-mail our executive editor Nelly Stratieva ([stratieva@lexxion.eu](mailto:stratieva@lexxion.eu)) and keep in mind the following deadlines:

- Issue 1/2020: 15 January 2020;
- Issue 2/2020: 15 April 2020;
- Issue 3/2020: 1 July 2020;
- Issue 3/2020: 1 October 2020 (Young Scholar Award).

I hope you enjoy reading this edition of the European Data Protection Law Review!

*Bart van der Sloot*  
*Tilburg Institute for Law, Technology, and Society (TILT)*  
*Tilburg University, Netherlands*