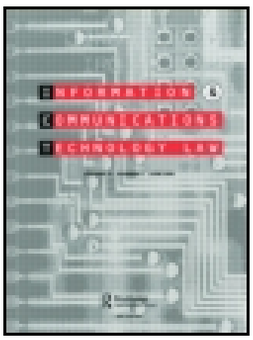

Routledge
Taylor & Francis Group

Information & Communications Technology Law

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cict20

Procedural law for the data-driven society

Bart van der Sloot & Sascha van Schendel

To cite this article: Bart van der Sloot & Sascha van Schendel (2021): Procedural law for the data-driven society, Information & Communications Technology Law, DOI: 10.1080/13600834.2021.1876331

To link to this article: https://doi.org/10.1080/13600834.2021.1876331

Published online: 20 Jan 2021.

INFORMATION & COMMUNICATIONS TECHNOLOGY LAW
https://doi.org/10.1080/13600834.2021.1876331

Routledge
Taylor & Francis Group

Procedural law for the data-driven society

Bart van der Sloot and Sascha van Schendel

Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, Tilburg, the Netherlands

ABSTRACT

Large-scale data applications are becoming an increasingly integral part of how both public and private sector organisations function. The transition towards a data-driven society means that processes within organisations will be organised structurally differently than they used to be and that decision-making will be based on profiles and algorithms more often than not. This change requires several adjustments to the legal regime, both to make the best possible use of the opportunities this change has to offer and to lay down safeguards against dangers and risks. To facilitate this process, a number of changes is needed to the current, individual-centred legal paradigm, such as laying down a protective regime for non-personal data, providing protection to public interests and societal harms and granting a bigger role for representative and collective actions and public interest litigation.

KEYWORDS

Big data; class action; privacy; collective interest; group representation; procedural rights

1. Introduction

Large-scale data applications are becoming an increasingly integral part of how both public and private sector organisations function. The transition towards a data-driven society means that processes within organisations will be organised structurally differently than they used to be and that decision-making will be based on profiles and algorithms more often than not. This change requires several adjustments to the legal regime, both to make the best possible use of the opportunities this change has to offer and to lay down safeguards against dangers and risks. The various reports and memoranda that have been published on this subject to date focus mainly on issues of material justice and granting strong material rights to citizens.[1]

Although these issues are and will continue to be of utmost importance, perhaps the greatest legal challenge spiralled by large data-driven operations is of a procedural nature. This contribution will argue that the biggest gap between the current legal paradigm and Big Data is in procedural law and concerns over access to justice and principles

CONTACT Sascha van Schendel S.vanschendel@tilburguniversity.edu

[1]Q Ai-min and P Jia, 'Right to Data, Data Sovereignty and the Basic Principle of Big Data Protection' (2005) 1 Jour. Soochow. Univ. 1. BJ Koops, 'Forgetting Footprints, Shunning Shadows: A Critical Analysis of the Right to be Forgotten in Big Data Practice' (2011) 8 SCRIPTed 229. Y McDermott, 'Conceptualising the Right to Data Protection in an Era of Big Data' (2017) 4 Big Data Soc. 1. S Wachter and B Mittelstadt, 'A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI' (2019) 1 Colum. Bus. L. Rev. 494. VN Gudivada, R Baeza-Yates and VV Raghavan, 'Big Data: Promises and Problems' (2015) 3 IEEE 20.

of procedural fairness. It is hard to underestimate their relevance, as citizens who have rights but are unable to successfully enforce them remain empty-handed. The main thesis of this contribution is that the gap between the legal paradigm and Big Data as a technology impacting society is that law is primarily concerned with providing protection to the private interests of natural persons by assigning them subjective claim rights, while many of the issues tricked by large-scale data processing operations transcend the individual. A legal regime that addresses incidental data harms only on an individual level runs the risk of leaving unaddressed the underlying causes, allowing structural problems to persist. That is why procedural law needs to be de-individualized.

It is impossible to describe the procedural law and the guarantees for access to justice that currently exist in general, because these differ from country to country and from jurisdiction to jurisdiction. Still, in particular in the Global North, there is a large emphasis on the individual. Although specific exceptions exist, some of which will be discussed in more detail in this contribution, serving as best practice or inspiration for a prospective legislative approach, the legal regime is mostly focused on attributing subjective rights to natural persons to protect their private interests. Criminal procedural law is focused on protecting the interests of the accused, administrative procedural law allows citizens to challenge decisions taken by governmental organisations when they have a direct, individual and substantial interest in the matter and civil procedural law grants claim rights to those that have suffered material or immaterial harm.

This means that by and large, under the current legal regime, in order to invoke a right, a citizen must be able to demonstrate an individual interest. This principle works relatively well for traditional legal matters, such as when a building permit is rejected by the municipality, when a person requests compensation for a defamatory publication or when the police wire-taps an individual's telephone communication. These are, however, matters that have an effect on specific individuals or small groups of people, making it relatively easy to identify and demarcate an infringement in time and place.

This is different, however, with many legal issues that revolve around large data collection programs. Big Data processes can hardly be demarcated in time and person, but form a structural and integral part of the actions and operations of governmental agencies, companies and citizens. For example, the cameras on the corner of almost every street in cities have no specific effect on one particular individual, they permanently film everyone who moves around in the city; an intelligence service that collects the communication data of an entire neighbourhood or a city does not affect anyone specifically or individually, but everyone equally; when a municipality, relying on smart city data input, decides to restrict the opening hours of shops in a deprived neighbourhood, such a policy is not targeted at specific individuals, but might reinforce societal inequality.

This contribution will take as starting point 10 tensions or legal lacunas, that have been abundantly described in literature,[2] and will be taken as non-controversial premises:

(1) Although citizens have a right to complain, it is often unclear to individuals whether their data is collected, processed and used and if so, by whom. This may be due to

[2]See inter alia: K Crawford and J Schultz, 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms' (2014) 55 BCL Rev. 93. NM Richards and JH King, 'Big Data Ethics' (2014) 49 Wake Forest L. Rev. 393. K Yeung, '"Hypernudge": Big Data as a Mode of Regulation by Design' (2017) 20 Inf. Commun. Soc. 118. I Rubinstein, 'Big Data: The End of Privacy or a New Beginning?' (2012) 1 Int. Data Priv. Law 12.

the fact that the operations of data-processing organisations are secret, because data collection is covert or is integrated in the environment.

(2) Estimates vary from a few hundred organisations to a few thousand that possess data about an average citizen. This number will most likely only go up with the advent of data technologies and applications. This means that it is increasingly unrealistic to ask from an individual to check with all these parties whether they have her personal data, whether they respect all relevant legal standards and, if not, to go to court.

(3) As more and more decisions are made on the basis of data-driven analyses, often with the use of self-learning algorithms, it will be increasingly difficult for ordinary citizens to understand the logic behind these decisions and attest that logic.

(4) The parties deploying Big Data are typically large multinationals and governmental agencies that not only have technical and practical expertise on the design and implementation of the data-driven processes, but also have the time and resources to sit out long and costly legal procedures, which does not hold true for most citizens.

(5) Most legal regimes require the person filing a complaint to demonstrate an individualizable harm that can be distinguished from the amorphous mass, while Big Data applications typically affect large groups or the entire population.

(6) It is difficult under the current legal regime to complain about positive things that did not happen due to data analytics. Suppose the police decide, on the basis of predictive policing, to patrol mainly in the southern district of a city and less so in the northern district. One question is whether an inhabitant of the southern district can file a complaint because she believes that the database, used for these predictions, is biased; another question is whether an inhabitant of the northern district can do so because she wants more surveillance in her neighbourhood. Most jurisdictions do not provide for such a possibility.

(7) An additional problem could emerge when the police decide to pay special attention to, for example, drug criminals and search available databases for clues, restricting the search queries to inhabitants of a particular neighbourhood, with a high density of people with a migration background. Suppose the results obtained, combined with additional evidence, subsequently lead to an arrest and criminal procedure. Can a person against whom irrefutable evidence has subsequently been found that she is dealing in drugs object to the evidence produced because the initial search was biased?

(8) Informational privacy and data protection regimes only protect data if they relate directly or indirectly to a person. When data are analysed in large datasets, however, data are mostly processed on aggregated level. This means that a large part of data-driven processes falls outside the scope of the current regulatory regime.

(9) Data analytics, to a large extent, relies on statistical analysis. While codes of conduct and guidelines have been developed for statistical authorities, organisations relying on Big Data processes are not bound by these codes and principles, meaning that many errors and biases in both the datasets, the algorithms and outcomes, persist.

(10) Big Data processes trigger societal questions that transcend the individual interest. To provide an example, the question of whether intelligence services should be

authorised to engage in so-called bulk interception is an abstract, normative question – narrowing this point down to the possible interests of a specific individual whose data might be collected diverts attention away from the bigger issue.

Taking these relatively uncontested tensions as a starting point, this article suggests that procedural law should be revised to make it ready for the twenty-first century. In essence, the argument will be that procedural law should centre less around individual rights and individual interests and more around public actions and societal interests. This argument will be developed in three steps. First, this article will show that data-driven processes should be subject to better procedural safeguards (Section 2). Next the article shows that the protection of non-individual interests should play a bigger role along with increased possibilities to address issues of legislation and systems (Section 3). The article advocates that the access to justice should be facilitated through more insight, less complex standing requirements and proceedings, and mechanisms to address litigation costs (Section 4). Finally, some conclusions will be drawn and a final reflection will be provided on the costs of legal procedures and awarding damages to claimants (Section 5).

This article is the result of a bigger study on the future of procedural law in the twenty-first century in light of the Big Data era. It was written at the request of the Dutch government, that can issue tenders via an independent organisation. WODC (the Dutch abbreviation for *Wetenschappelijk Onderzoek- en Documentatiecentrum*, in English: Scientific Research and Documentation Centre) is the Ministry of Justice and Security's knowledge centre. The WODC conducts independent in-house research or commissions external research. Such research is written independent from any governmental supervision or involvement. In this case, the WODC commissioned researchers from the Tilburg University to write a report on the following research questions.[3]

(1) How can Big Data and Big Data applications used by the government be conceptualised?
(2) Under what circumstances and conditions do citizens and interest groups currently have access to the courts in order to challenge Big Data applications used by the government?
(3) To what extent do such possibilities exist when there is no/minimal individual harm?
(4) What are the various possibilities to remedy the gaps identified under questions 2 and 3?
(5) What are the advantages and disadvantages of the various possibilities identified under question 4?
(6) To what extent can citizens and interest groups in other EU countries legally challenge Big Data applications used by the government?

For this study, the legal systems of Australia, Belgium, Canada, France, Germany, Israel, the Netherlands, New Zealand, the United Kingdom and the United States were studied, as well as the overarching European legal systems of the European Union and the Council

[3] <https://repository.tudelft.nl/assets/uuid:6f5ae910-be97-4860-9a25-0e1e386631ec/2900_volledige_tekst_tcm28-402015.pdf>.

of Europe. Belgium, France, Germany, the Netherlands and the United Kingdom were chosen because, of all European countries, these are the ones that have the most vital legal culture in terms of collective action, constitutional review and public interest litigation. The legal system of Australia, Canada, Israel, New Zealand and the United States were studied, in particular because of their Common Law tradition and the use of various forms of class actions and collective actions, as well as other legal figures that could be applied in the Big Data context, such as the use of sunset clauses and special advocates.

In addition to the legal comparative element of this research, several key players in the Dutch legal context were interviewed, such as lawyers involved in public interest litigation, civil society organisations, academics, governmental organisations using Big Data applications, the State's defence lawyer, the Ombudsman and a Supreme Court justice.[4] These parties were selected for an interview because they represent the various sides of the legal, societal and political spectrum at play in court cases initiated by groups and civil rights organisations against governmental organisations using Big Data applications.

Finally, to verify the results of the study, two workshops were held, in which civil servants, academics and policy makers were invited to discuss the preliminary findings of the study and give their opinion on sub-aspects of the report. The first was held in The Hague, where the Dutch government is seated and aimed at facilitating a discussion between civil rights organisations and public interest litigation lawyers on the one side and civil servants and policy makers on the other side. The second was held in Brussels, the place where the European Union is seated, and was aimed at facilitating a discussion between supervisory authorities, academics and civil rights organisations. At this second workshop, speakers included Ianika Tzankova (Tilburg University), Wojciech Wiewiórowski (European Data Protection Supervisor), Marc Rotenberg (EPIC) and Max Schrems (NOYB). Both were open workshops, but were only advertised in circles of experts and professionals in this field. Both attracted some 20 participants, which facilitated an interactive discussion.

Where this article mentions results obtained from interviews or workshops, reference is made hereto.

2. Procedural safeguards for big data processes

The prevalent legal regime on the collection, analysis and use of data in the jurisdictions studied is predominantly focused on the concrete effects data processing has on the interests of individuals. From the interviews with experts, discussions at the workshops and the literature studied, it emerged that the regulatory regime could be ameliorated when it would also lay down standards that provide protection to general, societal, interests. In particular, three improvements surfaced, which will be discussed in detail in the subsequent sub-sections, namely the regulation of non-personal data (Sub-section 2.1), applying standards for statistical analytics to Big Data processes (Sub-section 2.2) and

[4]Amnesty International Netherlands (Doutje Lettinga & Nine de Vries), Dutch Data Protection Authority (Aleid Wolfsen), Taxs Authority (Raymond Kok), Bits of Freedom (David Korteweg), Boekx lawyers (Otto Volgenant & Charlotte Hangx), Bureau Brandeis (Christiaan Alberdingk Thijm), The Institute for Human Rights (Jan-Peter Loof & Juliette Bonneur), DataUnion (Reinier Tromp), Supreme Court (Ybo Buruma), National Lawyer Pels Rijcken (Cécile Bitter), National Ombudsman (Reinier van Zutphen & Martin Blaakman), Privacy First (Vincent Böhre), Public Interest Litigation Project (Jelle Klaas), Radboud University (Roel Schutgens & Joost Sillen) en writer/philosopher (Maxim Februari).

evaluating the effectiveness of data-driven applications and terminating them when they appear to be ineffective (Sub-section 2.3).

2.1. Non-personal data

Legal regimes on informational privacy or data protection around the globe are based on the protection of individual interests. They either regulate 'personal data', defined as 'any information relating to an identified or identifiable natural person',[5] 'information' referring to 'data on the personality, personal status, intimate affairs, state of health, economic position, vocational qualifications, opinions and beliefs of a person',[6] 'records', defined as 'any item, collection, or grouping of information about an individual that is maintained by an agency',[7] 'personal information', meaning 'information about an identifiable individual'[8] or 'personal data', referring to 'any information relating to an identified or identifiable individual'.[9] Data are protected when they relate directly or indirectly to an individual. The problem with this approach is that large-scale data-driven applications and technologies do not necessarily thrive on personal data, they can be based on large sets of aggregated data.[10]

During one of the workshops, the following example was given, which was inspired by a real-life situation. Suppose a school registers the number of students per class that wear a hoody or a baseball cap. Such is used as an indicator for a-social behaviour and learning-problems. The school designs a policy in which teachers are attributed hours to fulfil their role as personal mentor; the more students that fall in the described category, the more hours teachers are assigned. Such is a relatively innocent and small data example, but it is easy to see how such a policy could evolve into something more problematic when applied on a bigger scale. What is important here is that no personal data is gathered, analysed or used. Attendees to the workshops expected in particular these types of processes to be commonly used in the future, as organisations could argue that as they are not processing personal data and consequently do not fall under the prevalent privacy and data protection regimes.

This debate has gained momentum in academic literature as well, in particular when the European Union decided to adopt not only the General Data Protection Regulation (GDPR), arguably setting the highest level of protection of personal data around the world, but also the Regulation on the transfer for non-personal data (RTNPD), which in many respects is the mirror-reflection of the GDPR. While the GPDR's aim is both to provide protection of the interests of natural persons and to stimulate the free movement of personal data,[11] the RTNPD only 'aims to ensure the free flow of data other than personal data'.[12] Consequently, the RTNPD provides no restrictions on the processing of non-

[5]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 (GDPR) art 4. (1).
[6]Israel Protection of Privacy Law, 5741–1981, art. 7.
[7]Records maintained on individuals, U.S.A. Privacy Act 2018, § 552a.
[8]Canadian Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, Assented to 2000-04-13.
[9]OECD Guidelines governing the protection of privacy and transborder flows of personal data, 2013.
[10]L Taylor, L Floridi and B van der Sloot (eds), *Group Privacy* (Springer, 2017).
[11]GDPR art. 1.
[12]Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union (RTNPD) art 1.

personal data; rather, it prohibits governments and discourages organisations from laying down restrictions.

The RTNPD recognises an explosion in the collection and production of non-personal data, which is attributed to the fact that

> expanding Internet of Things, artificial intelligence and machine learning, represent major sources of non-personal data, for example as a result of their deployment in automated industrial production processes. Specific examples of non-personal data include aggregate and anonymised datasets used for big data analytics, data on precision farming that can help to monitor and optimise the use of pesticides and water, or data on maintenance needs for industrial machines.[13]

These non-personal data are considered a huge asset for governments and companies and are believed to represent a high potential value for the EU economy, were it not that

> the effective and efficient functioning of data processing, and the development of the data economy in the Union, are hampered, in particular, by two types of obstacles to data mobility and to the internal market: data localisation requirements put in place by Member States' authorities and vendor lock-in practices in the private sector.[14]

That is why this Regulation prohibits data localisation requirements, unless they are justified on grounds of public security in compliance with the principle of proportionality, and encourages the industry to develop open data standards.

A number of academics have questioned this approach,[15] some have even claimed that the RTNPD would "undermine and even cause conflicts with the objectives of the GDPR and the fundamental right to data protection".[16] In general, two critiques against this distinction between personal and non-personal data have been put forward. First, that this distinction can hardly be made, because the status and nature of data is increasingly volatile. A dataset that contains ordinary personal data may be linked and enriched with another dataset and transformed into a set that contains sensitive data; the data may then be aggregated or stripped from their identifiers and become non-personal data; subsequently, the data may be deanonymized or integrated into another dataset containing personal data. These subsequent steps may happen in a split second. The second critique is that the underlying rationale for providing different regimes of protection to different categories of data is that the more directly data or datasets are linked to an individual and the more sensitive the data are, the higher the level of protection provided. However, whether data analytics is based on personal or aggregated data is increasingly irrelevant as it is possible to design and make policies that affect groups of people on the basis of general information that were never personal data and may not have an effect on specific individuals, but on large groups or everyone living in society.

Therefore, the first suggestion for the amelioration of the current privacy paradigm is to create legislation for processing non-personal data. Current privacy and data

[13]RTNPD rec. 9.
[14]RTNPD rec. 3.
[15]K Irion, Public Security Exception in the Area of non-Personal Data in the European Union' (2018). <https://pdfs.semanticscholar.org/bcf6/833a682291ef5227d2f77ec4f1129661daa6.pdf>.
[16]I Graef and others, 'Feedback to the Commission's Proposal on a Framework for the Free Flow of non-Personal Data' (2018) <https://ec.europa.eu/info/law/better-regulation/feedback/8922/attachment/090166e5b7f755e3_en>. D Broy, 'The European Commission's Proposal for a Framework for the Free Flow of Non-Personal Data in the EU' (2017) 3 Eur. Data Prot. L. Rev. 380.

protection legislation could serve as point of reference, though potentially, a 'data protection light' regime could be deployed when processing concerns non-personal data. Take the principles of purpose limitation, data minimisation and storage limitation. Even if a school would process no personal data, but only non-identifiable or aggregated data, would it be unreasonable to require it to think about its goal for gathering data and restrict its data processing to those data that are required in light of that goal? To provide another example, given the fact that data about categories or groups of people can be misused when they fall into the wrong hands, and could potentially do equal or even greater harm than when it concerned identifiable data, should not the principles of data security be applicable to organisations that processes non-personal data too? To provide a final example, current privacy and data protection regimes lay emphasis on transparency. In part, this is to inform persons of the fact that their personal data are being processed, but in part, it also ensures for public transparency and accountability. At least larger companies and governmental organisations deploying policies on the basis of aggregated data about groups or categories could be required to disclose on a public website which type of information they process, how and why.

2.2. Data analytics

An additional concern put forward, in particular by a number of interviewees, was the lack of standards for the analysis of data. Most standards that have been developed so far are not laid down in law, but take the form of principles, guidelines and codes of conduct. And although some of these do recognise both harm on an individual, group and societal level,[17] most concerns are over individual interests, such as individual control, lack of consent, informational privacy and human autonomy.[18] In addition, most scholars working on algorithmic accountability and most guidelines are focused on the question how the existing human rights and ethical principles could be safeguarded vis-à-vis algorithmic data-analytics.[19] As one report puts it: "An ethics framework for AI is not about rewriting these laws or ethical standards, it is about updating them to ensure that existing laws and ethical principles can be applied in the context of new AI technologies".[20]

In addition to applying traditional human rights regimes, such as the right to privacy, the non-discrimination principle and the right to a fair trial, to algorithmic decision-making and designing standards to bolster data subjects' capacity to have access to information about these processes, the core problem that was identified was the lack of knowledge about statistics and its limits by those that were involved in Big Data projects. A shift has happened from domains that have historically dealt with large-scale data analytics, such as social sciences, weather and climate forecasts and traditional governmental organisations providing public statistical data, to statistics deployed by computer

[17]European Parliament, 'A Governance Framework for Algorithmic Accountability and Transparency', PE 624.262 – April 2019 <https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf>.

[18]Government Use of Artificial Intelligence in New Zealand <https://www.cs.otago.ac.nz/research/ai/AI-Law/NZLF%20report.pdf>.

[19]See for example: Council of Europe, 'Algorithms and Human Rights' <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>.

[20]D Dawson and others, 'Artificial Intelligence: Australia's Ethics Framework' (2019) Data61 CSIRO, Australia. <https://static1.squarespace.com/static/52b5f387e4b08c16746b6b70/t/5d0f51600ff3ce000115884b/1561284978417/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf>.

programmers and policy makes. During one of the workshops, one person bluntly stated 'Big Data is statistics by non-statisticians' and one of the experts giving a presentation stressed: "Wild animals are subject to more regulation than data analysts". As one interviewee put it more mildly:

> Data has the emanation of neutrality, of objectivity. The word suggests that data are a given, a reality of nature. However, the opposite is true. Data are made by people, they are the product and the result of subjective choices. That is important because it is often naively assumed that data are the truth and data analysis yields the truth, without any further evaluation. The results of data analysis are all too often used as direct input for policy decisions.[21]

Obviously, some of the problems that result from this inexperience with data analytics might be addressed through the existing ethical and legal frameworks and additional ones may be tackled by the many suggestions now proposed by the scholars working in the field of algorithmic fairness, accountability and justice. But the problems encountered, as was confirmed by a number of civil servants present at the workshops, were of a much broader nature. When, for example, governmental agencies design general policies based on incomplete data, algorithms with false assumptions or invalid or non-significant correlations, such may create various problems that not necessarily lead to discrimination or unfairness. A very innocent example was used by way of illustration: suppose a governmental agency decides to do quality control for bridges constructed between 1970 and 1980 by a certain company only once every 20 years, because data has shown that these types of bridges are exceptionally strong, but the dataset on which these conclusions were drawn was biased, such may result in a public catastrophe. Or, when a minister decides to send books to households where children live, because data has shown that there is a statistical correlation between the number of books at home and children's performance at school, wrongly taking the correlation for a causal relation, such may lead to ineffective and inefficient expenditure of public money.

That is why one of the basic suggestions was to learn from of the standards that have already been developed in the area of statistics, such as,[22] but not limited to the Ethical Guidelines for Statistical Practice by the American Statistical Association, the European Statistics Code of Practice and the 10 fundamental principles of statistics by the United Nation's General Assembly.[23] Although not all of the principles developed for traditional statistics can be applied one on one to the context of Big Data analytics, in part because the statistical methods sometimes differ, these principles can serve as a source of

[21] <https://repository.tudelft.nl/assets/uuid:6f5ae910-be97-4860-9a25-0e1e386631ec/2900_volledige_tekst_tcm28-402015.pdf>.

[22] See also: GDPR rec. 71. Treaty on the Functioning of the European Union and Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation 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 No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities art. 338.

[23] American Statistical Association (2018). Ethical Guidelines for Statistical Practice Prepared by the Committee on Professional Ethics of the American Statistical Association. <https://www.amstat.org/asa/files/pdfs/EthicalGuidelines.pdf>; Resolution adopted by the General Assembly on 29 January 2014 (A/68/L.36 and Add.1)] 68/261. Fundamental Principles of Official Statistics; <https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7?t=1528447068000>.

inspiration for governments that want to ensure more effective and more qualitative data-analytics. Some examples that were discussed during the workshops where:

(1) *Professional environment*:
   (a) *Independence:* An organisation or unit should be independent from political and other external interference in developing, producing and disseminating statistics.
   (b) *Competence*: The capacity of employees engaging in data analytics should be beyond dispute and other arguments than their competence should play no role in their appointment.
   (c) *Clarity:* Statistical outputs are clearly distinguished from policy statements.
(2) *Recourses*: Adequate staff and resources are available to meet statistical needs.
(3) *Privacy*: Data are kept safely and confidentially.
(4) *Objectivity*:
   (a) *Objectivity:* Statistics are compiled on an objective basis; choices of sources and statistical methods are informed by statistical considerations. Statistical releases and statements made in press conferences are objective and non-partisan.
   (b) *Correction of errors:* Errors discovered in published statistics are corrected at the earliest possible date and publicised.
(5) *Reporting burden*:
   (a) *Egality:* Data gathering is spread as widely as possible over populations.
   (b) *Necessity:* The range and detail of the data is limited to what is necessary.
   (c) *No duplicity:* Existing sources are used to avoid duplicating requests for data.
(6) *Quality*:
   (a) *Relevant training:* There is a policy of continuous vocational training for staff.
   (b) *Consistency: S*tandard concepts, definitions and classifications are consistently applied and regularly evaluated and adjusted if necessary.
   (c) *Coherent:* Statistics are internally coherent and consistent (i.e. arithmetic and accounting identities are observed).
   (d) *Comparable:* Statistics are comparable over a reasonable period of time.
   (e) *Documentation:* Sampling errors and non-sampling errors are measured and systematically documented.
(7) *Validation*:
   (a) *Prior testing:* In the case of statistical surveys, questionnaires are systematically tested prior to the data collection.
   (b) *Monitoring:* Data collection, entry, and coding as well as editing and imputation methods are routinely monitored and revised as required.
   (c) *Designing:* Statistical authorities are involved in the design of administrative data in order to make administrative data more suitable for statistical purposes.
(8) Accountability:
   (a) *Metadata:* Statistics and corresponding metadata are presented, and archived, in a form that facilitates proper interpretation and meaningful comparisons.
   (b) *Transparency:* Information on the methods used is publicly available.

2.3. Sunset clauses

A third aspect that came to the fore is the perceived ineffectiveness of many of the Big Data applications and technologies, or, put differently, the lack of proof that initiatives such as predictive policing, data-driven fraud detection programs deployed by tax authorities, personalised advertising and mass surveillance actually work, that is, that they are more effective than non-data-driven applications. For example, in one of the countries studied, the Netherlands, several data-driven projects were evaluated and most were deemed ineffective. For example, with respect to a pilot with predictive policing, an evaluation performed by the Police Academy concluded: "We were unable to find indications that predictive policing ultimately leads to less crime".[24] Similarly, the Tax Authority, which was set to transform to a data-driven organisation a number of years ago, firing many employees and hiring data experts instead, has been in constant turmoil since. An independent institution concluded that more than half of the data-driven projects failed or were significantly delayed.[25] A commission of parliament evaluated all data-driven projects initiated in the public sector and concluded that many of those failed and were terminated after one or two years, and that the costs were often double of those initially calculated. In particular, the commission disapproved of the fact that when a project was initiated, more often than not, there was no evaluation of any kind of whether the project actually worked or not.[26]

For a longer time, scholars have cast doubts on the high expectations organisations in both the private and the public sector have of data-driven projects. An example may be the fight against terrorism and the use of mass surveillance by intelligence agencies. Many have suggested that mass surveillance is simply not the right tool for doing so.[27] This led Susan Landau, after having critically reviewed large-scale data programs operated by US intelligence agencies, to conclude:

> Efficacy should always come first. Asking the efficacy question requires looking hard at the value of a programme. What does it produce? Is the production worth it? Are there alternate ways of learning needed information that are less costly? What's the opportunity cost, that is, what resources – people, funds, expertise, engineering – are being used here that could be more effectively be spent elsewhere? Personnel and budgets are not infinite; asking these hard questions causes those deploying the tools to carefully focus on utility.[28]

Obviously, this is not to say that data-driven processes do not work per sé; many of them do or will do so in time. However, it is also clear that the effectiveness of data-driven processes in reality is often lower than what was expected or promised when the project started. Focussing on the effectiveness of data-driven processes has a number of benefits. First, it is uncontroversial; everyone agrees that when organisations use resources, in the private, semi-private or public sector, such should be done in the

[24]B Mali, C Bronkhorst-Giesen and M den Hengst, 'Predictive Policing: Lessen Voor de Toekomst. Een Evaluatie van de Landelijke Pilot. Politieacademie' (2017) <https://www.politieacademie.nl/kennisenonderzoek/kennis/mediatheek/PDF/93263.PDF>.

[25]Algemene Rekenkamer (2017). Tussenstand Investeringsagenda Belastingdienst <https://www.rekenkamer.nl/publicaties/rapporten/2017/10/11/tussenstand-investeringsagenda-belastingdienst>.

[26]Tweede Kamer (2014–2015), 'Parlementair onderzoek naar ICT-projecten bij de overheid' 33326(5) <https://www.tweedekamer.nl/sites/default/files/field_uploads/33326-5-Eindrapport_tcm181-239826.pdf>.

[27]See for example: B Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World* (W. W. Norton & Company, 2016).

[28]S Landau, 'If It Isn't Efficacious, Don't Do It' (2019) 4 Eur. Data Prot. Law Rev. 466.

most effective way possible. Second, effectiveness generally serves as a precondition, a question that should be answered even before turning to a potential 'balance' between private and public interests, between privacy and other human rights on the one hand and the need for public order, national security or fraud detection on the other hand. Even if, say, predictive policing would pass the balancing test, because public security is deemed more important than privacy protection, it still makes no sense to invest in predictive policing if tests show that such programs are not effective or, put differently, that investments in non-data driven police capacities yield more results in terms of reducing crime. Consequently, effectiveness places a potential restriction on data-driven processes in the form of public or shared interests, instead of the private interests of a citizen. Third, the burden of proof is typically on the organisation initiating the data-driven process to show that it is effective and more so than when the same investments were made in non-data-driven processes. This relieves citizens from the obligation to show damage or harm vis-à-vis these types of Big Data projects when they rely on their human or fundamental rights.

That is why it has been suggested to look for inspiration to the legislative practice, in particular in common law countries, such as the Australia, New Zealand, United Kingdom and the United States, to work with sunset clauses.[29] Such clauses are implemented in a law to ensure that a certain power or project is terminated after a specified period, unless the power or project is renewed by the legislature.[30] The use of these types of clauses allows the legislator, inter alia, to grant the executive branch exceptional powers in exceptional times, such as in the case of the fight against terrorism and the use of mass surveillance, or to allow an organisation to experiment with a technology or application, the effectiveness of which has yet to be proven.[31] Doing so, a city may be allowed, for example, to run smart city experiments for three years, after which their powers to do so expire, unless they are explicitly renewed.

> The central idea behind experimental legislation is that: (1) it concerns a temporary deviation of existing laws or regulations; (2) the scope of the experiment will be fixed in terms of time, place, and/or addressees; (3) the effects and side effects of the rules will be evaluated; and (4) in case of success, the regime will be broadened so that the experimental rules can also apply to other similar situations.[32]

That is why, when a new data-driven application or technology is introduced, at least in the public sector, but such could also be a best practice for the private sector, it should be the standard to work with sunset clauses. Before introducing a technology or application, a baseline measurement should be carried out, establishing what the situation is without the data-driven application or technology. Subsequently, it should be defined in concrete terms what goal the Big Data project should achieve in order to be deemed successful; i.e.

[29] AE Kouroutakis, 'Disruptive Innovation and Sunset Clauses: The Case of Uber and Other on Demand Transportation Networks' in S Ranchordas and Y Roznai (eds), *Time, Law and Change: An Interdisciplinary Study* (Hart, 2019). DH Schraub, 'Doctrinal Sunsets' (2019) 93 South. Calif. Law Rev. 3. AE Kouroutakis, *The Constitutional Value of Sunset Clauses* (Routledge, 2017). S Ranchordás, 'Sunset Clauses and Experimental Regulations: Blessing or Curse for Legal Certainty?' (2015) 36 Statut. Law Rev. 1.

[30] I Bar-Siman-Tov and G Harari-Heit, 'The Legisprudential and Political Functions of Temporary Legislation' in S Ranchordas and Y Roznai (eds), *Time, Law and Change: An Interdisciplinary Study* (Bloomsbury Publishing, 2020).

[31] MA Heldeweg, 'Experimental Legislation Concerning Technological & Governance Innovation – an Analytical Approach' (2015) 3 Theo. Prac. Leg. 2.

[32] R Van Gestel and G Van Dijck, 'Better Regulation Through Experimental Legislation' (2011) 17 Eur. Public Law. 3.

which added value should it bring in terms of effectiveness, efficiency or another value. After a year or two, an initial evaluation should be carried out to assess the extent to which that goal has been achieved. If such evaluation yields that it has not, the project should be terminated; if the evaluation suggests that the goal has not yet been achieved in full, adjustments and improvements should be made to the Big Data project at that point. A final assessment in terms of costs and benefits should be conducted after 3 years. It is then up to the legislator, or, in the case of a private sector organisation, to the board, to decide on the continuation of the project.

3. De-individualization of procedural rights

Most systems of procedural law focus on the individual. Standing is traditionally only granted when the applicant can demonstrate a special or individual connection to the matter complained of as well as some sort of material or immaterial harm. In order to accommodate for legal protection against the effects of Big Data on groups and society at large as well as the indirect impact on individuals, a de-individualization of the procedural law is necessary. Several directions of improvement are possible, which will be discussed in detail in the subsequent sub-sections. The opportunities to challenge the legality and legitimacy of laws and policies though constitutional and administrative law could be explored (Sub-section 3.1), the representation of individuals by groups and organisations in civil law could be expanded (Sub-section 3.2), and the possibilities for challenging Big Data systems, and their possible bias or discriminatory effects, through criminal law could be introduced (Sub-section 3.3).

3.1. Guarantees in administrative law

The first problem identified through the literature study, the interviews and the workshops with respect to the current legal regime in light of the data-driven environment is that often, individuals and organisations do not so much want to claim that they have been harmed individually and specifically and that such was unlawful or unjust, but that a law or policy as such is undemocratic or conflicts with the rule of law. Most legal systems studied only allow for such claims in exceptional circumstances. During one of the workshops, Max Schrems discussed a number of existing legal figures that could provide a partial solution on this point, such as that in certain jurisdictions, the Privacy Commissioner or Data Protection Authority has the capacity to challenge laws and policies as such before a court of law. Recent jurisprudence by the two supranational European courts, the European Court of Justice (ECJ) and the European Court of Human Rights (ECtHR), were also mentioned during the workshops as best practices that should be expanded beyond their current scope.

Although ever since its existence, the ECtHR has rejected claims that regarded the general validity of laws and policies,[33] in December 2015, it turned, stressing that an alternative approach was needed, in particular in cases revolving around mass surveillance by secret services. Because individuals often remain oblivious to whether their

[33]See e.g. ECtHR, Lawlor v VK, application no. 12763/87, 14 July 1988.

data have been collected by these organisations or not, and will thus seldom invoke their rights, the Court acknowledged that there

> is therefore a greater need for scrutiny by the Court, and an exception to the rule denying individuals the right to challenge a law in abstracto is justified. In such cases the individual does not need to demonstrate the existence of any risk that secret surveillance measures were applied to him.[34]

Since that case, the ECtHR has accepted two more matters in which it did not assess the harm inflicted on the applicant and whether that harm was outweighed by the interests over national security, but instead, assessed the 'quality of law'. Doing so, the Court validates whether the law or policy as such abides by what it calls the minimum requirements of law, such as that there is a clear attribution of power to the executive branch, subject to specific conditions and requirements, that there are sufficient mechanisms for parliamentary oversight and judicial scrutiny and that the law lays down limits to the duration and scope of data collection programs.[35] The European Court of Justice, similarly, allows organisations to file a complaint in the general interest, specifically when cases concern matters of data retention and mass surveillance, testing legislation on aspects of the rule of law.[36]

If legislators wanted to update their legal regime in light of the data-driven environment, such an approach could be implemented as a general legal doctrine, not just for mass surveillance programs and data retention cases, but for all data-driven projects within the public sector. This would allow citizens or civil rights organisations to have a judge scrutinise systems of predictive policing, data-driven tax evasion programs or smart city experiments in their entirety. When it concerns data-driven programs run by governmental organisations, an opening in administrative or constitutional law seems most appropriate, though several jurisdictions studied also seemed to favour such claims being made via tort law. Potentially, analogous claim rights could be introduced to private sector organisations, in particular in private-public partnerships; for example when they concern smart city projects and living lab experiments, citizens and organisations could be allowed to make analogous claims with respect to the Terms of Agreement of such partnerships.

Such an approach could be introduced not only vis-à-vis legislation and bylaws adopted by the national legislator, but also those of provinces, states, districts and municipalities. Under the legal regimes studied, three important aspects of data-driven projects are generally difficult to address (though some provided openings on one or two points). First, under the current legal regime, it is difficult or impossible for citizens and organisations to rely on principles such as discussed in Section 2.1 and in particular Section 2.2, for example, arguing that a dataset is biased or statistical methods are not

[34]ECtHR, Roman Zakharov v Russia, application no. 47143/06, 04 December 2015, para. 171.

[35]ECtHR, Centrum för Rättvisa v Sweden, application no. 35252/08, 19 June 2018. ECtHR, Big Brother Watch and others v the United Kingdom, application nos. 58170/13, 62322/14 and 24960/15, 13 September 2018.

[36]European Court of Justice, Digital Rights Ireland Ltd (C-293/12) v Minister for Communications, Marine and Natural Resources, Minister for Justice, Equality and Law Reform, Commissioner of the Garda Síochána, Ireland, The Attorney General, intervener: Irish Human Rights Commission, and Kärntner Landesregierung (C-594/12), Michael Seitlinger, Christof Tschohl and others, In Joined Cases C-293/12 and C-594/12, 8 April 2014. European Court of Justice, Tele2 Sverige AB (C-203/15) v Post-och telestyrelsen, and Secretary of State for the Home Department (C-698/15) v Tom Watson, Peter Brice, Geoffrey Lewis, interveners: Open Rights Group, Privacy International, The Law Society of England and Wales, Joined Cases C-203/15 and C-698/15, 21 December 2016.

adhered to, without claiming that this had a direct impact on her individual life. Second, under most legal regimes studied, it is particularly hard to address general policies as such, for example a decision by a municipality not to allow new shops in the southern (impoverished) district of the city. Although a specific entrepreneur that finds her application for a permit rejected might appeal such decision, a citizen living in the southern district is typically not allowed to appeal the policy in general. Third, as discussed in the introduction of this article, although many jurisdictions allow for citizens to appeal decisions that have a negative effect on their individual lives, there is less room to appeal decisions because certain positive effects have not materialised, such as when a citizen of the northern district had hoped to have more investments in social services in her neighbourhood, but most resources go to the southern district, or had hoped to have more police patrols in her area.

As one of the interviewed organisations, involved in strategic litigation in the field of human rights protection, stressed:

> It is the requirement of a personal interest / personal damage that is the main obstacle to litigation. This bottleneck applies in matters of privacy protection, where the individual interest in, for example, mass surveillance-like applications is not always easy to determine. This can be a problem in anti-discrimination cases as well; think, for example, of predictive policing or other Big Data applications, which may have a discriminatory or stigmatizing effect. Again, it is not always clear whether someone has been disproportionately affected individually by a particular policy nor is it always possible to determine whether there is a causal link between a factor that is included in the policy and the aim pursued by that policy.[37]

This legal obstacle might be removed by making the changes described above, both at the level of national law and on the level of lower administrative law.

3.2. Guarantees in civil law

Most systems of civil procedural law studied grant claim rights to legal subjects to protect their own personal interests. Individuals can bundle complaints when their claims concern the same breach or infringement. However, in privacy cases, courts are hesitant to accept complaints pertaining to the rights of groups or organisations, especially when compared to other rights such as the freedom of expression or freedom of religion, where courts are generally more open to complaints by organisations and groups, such as newspapers or religious minorities. This already poses a significant limitation to potential court cases vis-à-vis data-driven applications and technologies. In addition, there are difficulties with bundling complaints of the individuals affected by the same matter. First, people often do not know who else is harmed by a certain policy or algorithmic decision-making process. Suppose a bank, using a biased dataset, unreasonably disadvantages people with red cars that like music by Britney Spears, because data analysis has identified those two points as having a predictive value for whether they will repay their loans. Obviously, these people do not know each other and have no way of contacting one another. Second, even if they would, precisely because data-driven applications and technologies have an effect on large groups or society as a whole, the damage inflicted on a

[37] <https://repository.tudelft.nl/assets/uuid:6f5ae910-be97-4860-9a25-0e1e386631ec/2900_volledige_tekst_tcm28-402015.pdf>.

specific individual is often not significant enough to take action. Suppose a data breach occurred with Ikea, through which hackers obtain the purchases of customers with a family card over the last year. In most jurisdictions, low sums of damages would be awarded to individuals affected, if at all; often such gains do not outweigh the costs in terms of time, energy and expenses necessary to start and win a lawsuit.

Several models for collective and class actions have been analysed that could potentially provide a partial solution to this point. One general solution is to work with so called op-out representative claims, in which one party can submit a claim to the court that represents a large group of victims that have not explicitly joined the case. Perhaps the most well-known legal figure on this point is the American class action, which has been applied in the field of large-scale data-driven projects on multiple occasions. A class representative is the formal party in a class-action lawsuit, the court must approve the class representative when determining whether the lawsuit can be treated as a class action. The plaintiff must be able to show that her case is typical for that of the other class members. Although the plaintiff has to pay for the legal costs, the benefit is that she is the party that can negotiate the terms of a potential settlement. Still, many experts have pointed to a number of drawbacks. For example, in one of the workshops, Marc Rotenberg summarised briefly: "Settlements are often unfair; lawyers get rich and companies continue illegal practices. Problems with 'cy pres' awards: money goes to organisations aligned with interests of companies, not consumers". In one of his publications, Rotenberg elaborated:

> The theory is that it is more efficient to merge all of the individual suits that might otherwise be brought. But class action litigation has its own shortcomings. Attorneys who represent the class members frequently settle these cases with the companies and agree to terms that provide benefits to the company, such as eliminating the possibility of all future lawsuits, and sacrifice the benefits that the individuals who they purport to represent might otherwise achieve.[38]

Another model that was discussed was the framework provided by the General Data Protection Regulation.[39] Article 80 paragraph 1 outlines the conditions as following:

> The data subject shall have the right to mandate a not-for-profit body, organisation or association which has been properly constituted in accordance with the law of a Member State, has statutory objectives which are in the public interest, and is active in the field of the protection of data subjects' rights and freedoms with regard to the protection of their personal data to lodge the complaint on his or her behalf, to exercise the rights referred to in Articles 77, 78 and 79 on his or her behalf, and to exercise the right to receive compensation referred to in Article 82 on his or her behalf where provided for by Member State law.

According to paragraph 2, it is left to the Member States to determine in their national implementations of the GDPR whether an organisation has the right to submit a complaint to a court or oversight body without a mandate by data subjects to do so. As Wojciech *Wiewiórowski explained during one workshop, this provision leaves* "it up to Member States to adopt measures they feel are suitable within their jurisdiction". It is difficult to

[38]M Rotenberg and D Jacobs, 'Enforcing Privacy Rights: Class Action Litigation and the Challenge of cy pres' in D Wright and P De Hert (red), *Enforcing Privacy. Regulatory, Legal and Technological Approaches* (Springer, 2016).

[39]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).

analyse the extent to which this provision will turn out to be effective in practice, because the first cases based on this article have yet to go to court.

Perhaps the most promising best practices were identified in a number of European countries: a French and a British legal figure stood out. Under French law there is a strong tradition of group or collective litigation, under a sectoral approach. Collective actions are not regulated as such, but in different domains, such as environmental law, consumer law, health law, and so forth. Civil society organisations have the right to submit a complaint representing their members or those whose interests they aim to protect through, what is called, the *action des associations en défense des intérêts individuels de leurs membres*. There are procedures where a large number of plaintiffs can bundle their complaints in the *action en représentation conjointe* or the *action de groupe*. Lastly, there are procedures in which private organisations represent the collective or group interests, the *action en defense d'un intérêt collectif*. These group claims are also open to organisations operating in the field of privacy and data protection.[40] A claim can be brought before a civil or administrative court by associations that, according to their statutes, provide protection to the rights to privacy and data protection and have done so for at least five years; certified consumer organisations and trade unions can also bring cases in order to protect the interests of consumer or of employees. Biard and Amaro argue that this system has the advantage that individuals do not have to initiate the claim themselves and that there is a higher chance of winning a case, because these organisations have the expertise and the resources needed.[41]

In the United Kingdom, the Group Litigation Order (GLO) is used. Litigating parties can request a GLO with the court, the court is then responsible for all the complaints under the GLO and a register is kept. Citizens can register with a GLO that has been accepted by a court, if their claims concern the specific matter that is outlined in the GLO.[42] In literature, the GLO is described as a "(…) surrogate for a mature system of class actions. Group actions are different from class actions because each group litigant is a member of a procedural class as a party, rather than as a represented non-party".[43] There have been 108 GLOs to date.[44] The GLO is generally described positively. It is seen as a welcome addition to the other complaint models and as a way of bundling complaints that concern a similar issue. This makes the GLO (cost) efficient; also, it can be used to address a structural problems that affect large groups. Because parties themselves must explicitly decide to join a GLO, citizens retain procedural autonomy. The GLO also has advantages over normal legal representation.

> In some situations, it is better for there to be a measure of procedural discipline and for each claim to be carefully pleaded and registered as part of a group action. This allows the court to consider both common issues and individual divergences from that common ground within the same action.[45]

[40]Loi n. 2018-493 du 20 juin 2018 relative à la protection des données personnelles.
[41]A Biard and R Amaro, 'Resolving Mass Claims in France: Toolbox & Experience' (2016) 5 RILE- BACT 1.
[42]C Hodges, 'The Civil Litigation System' (2009) 622 AAPPS 1.
[43]N Andrews, 'Multi-Party Proceedings in England: Representative and Group Actions' (2001) 11 DukeJComp&IntlL 249.
[44]HM Courts & Tribunals Service (2015). Guidance Group Litigation Orders <https://www.gov.uk/guidance/group-litigation-orders>.
[45]Andrews, N. (2001) 249.

Nevertheless, a number of disadvantages have been identified. For example, the disadvantage of the GLO compared to representation is that "representative actions can promote better access to justice than group litigation because the latter requires positive steps to be taken by the alleged victim of a legal wrong".[46] In addition, because it is an opt-in procedure, a GLO sometimes represents only a very small percentage of the actual stakeholders, which in turn raises the question of the representativeness of the complaint and the legal process.[47] Therefore, it is sometimes suggested to introduce an opt-out GLO.[48] In some countries such models do indeed exist.[49]

Where they do not exist, an opt-out GLO could be introduced or a system similar to the French approach could be adopted. Regardless of the current possibilities under the various national laws of countries for groups to litigate in the group interest, interests of their individual members, or the collective interest, those possibilities should be flexible enough for the Big Data era; where possibilities for non-individual litigation do not exist in general, they could be introduced; where they exist outside of the privacy context, for example only in the field of consumer law, they could be expanded; where they are dependent on purely financial harms or monetary claims, they could be expanded to include claims for invalidating laws or Big Data processes or addressing non-financial harms in other ways.

3.3. Guarantees in criminal law

One of the fields in which Big Data has an impact on procedural rights is that of criminal law. Big Data analytics are used to predict locations of crime, the risk of recidivism, to determine sentencing, and analyse patterns for evidence and to pinpoint suspects. Examples include the various predictive policing programs and the recidivism risk assessment tool COMPAS.[50] Obviously, these systems come with design choices, such as which data to use to train the system and which model or algorithm to use in the analysis. Previous research has shown that in policing systems biased data is almost inevitable.[51] An additional challenge to automated systems is in the use of predictive systems, as predictions inherently rely on probabilities and statistics, increasing the chance that the outcomes of analysis are erroneous. Besides possible bias and errors inherently to the system, it is also possible that systems are used in a biased or discriminatory way. In the case of hot spot policing, patrols are sent out to areas, sometimes even in specific periods, where crime is expected to take place. In these 'risky' areas, more crimes will be registered and more individuals are arrested than without the extra patrols. That raises question about the residents of those areas

[46]ibid.
[47]Civil Justice Council (2008). Improving Access to Justice through Collective Actions, Developing a More Efficient and Effective Procedure for Collective Actions Final Report A Series of Recommendations to the Lord Chancellor, November 2008.
[48]G Pendell, 'Another Step Towards Class Actions in England and Wales?' (2008) 5 WLNR 16838114. MA Behrens, GL Fowler and S Kim, 'Global Litigation Trends' (2008) 17 Mich. St. J. Int. L. 183.
[49]R Mulheron, 'The Case for an Opt-Out Class Action for European Member States: A Legal and Empirical Analysis' (2009) 15 Colum. J. Eur. L. 409.
[50]An individual being sentenced to six years of prison based partially on the analysis of COMPAS challenged his sentence arguing that the use of such a system violated his right to due process.
[51]For example in: J Angwin and others, 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks' (2016) 1 ProPub. 1.

being submitted to more police surveillance than others living in adjacent areas, and are consequentially more likely to be arrested. Inadvertently this can lead to a surveillance bias towards certain groups of people, for example those with a certain socio-economic background.[52]

Under most current legal regimes, citizens are able to address potential flaws in data-driven systems only to the point that they have been directly and individually affected; biases or deficiencies in the preliminary stage of a criminal proceeding are difficult to address. One interviewee was a Supreme Court judge in a European country; he explained the potential lacuna as follows:

> One of the problems described in literature is that a discriminatory selection can be made in the preliminary phase of an investigation, while the ultimate evidence leading up to the arrest and possibly the prosecution of a person is not in itself discriminatory. Ethnic profiling is unlawful (e.g. searching cars merely because the driver has a dark skin). However, the legal situation is different when the selection is based on other characteristics. Suppose there is a database in which the police only search for people with a certain ethnic background (whether or not by means of indirect identifiers), and subsequently investigate the say 50 hits that follow from that search and ultimately decide to arrest 3 people on the basis of concrete evidence. The group that was subjected to scrutiny is determined on the basis of a discriminatory search, but the evidence itself, leading to the arrest of the 3 people out of the 50 hits, is sound and itself non-discriminatory. It can be compared to a traffic cop who would only look at cars with a Polish number plate, but only stops and search Polish cars when the police database shows that there is reason to do so, for example, because there are outstanding fines or an outstanding arrest warrant with respect to the driver. Yet again, the evidence leading to the officer stopping a car and a potential arrest is sound, but the group being looked at is limited to drivers of one nationality. Both cases are problematic, but the problem cannot be addressed through criminal law.[53]

In the infamous case against COMPAS, Loomis vs. Wisconsin, the issues raised in the petition were:

> (1) Whether it is a violation of a defendant's constitutional right to due process for a trial court to rely on the risk assessment results provided by a proprietary risk assessment instrument such as the Correctional Offender Management Profiling for Alternative Sanctions at sentencing because the proprietary nature of COMPAS prevents a defendant from challenging the accuracy and scientific validity of the risk assessment; and (2) whether it is a violation of a defendant's constitutional right to due process for a trial court to rely on such risk assessment results at sentencing because COMPAS assessments take gender and race into account in formulating the risk assessment.[54]

Ultimately, the Wisconsin Supreme Court denied the petition and due process claims, which led to criticism on the system:

> (…) the Wisconsin Supreme Court held that a trial court's use of an algorithmic risk assessment in sentencing did not violate the defendant's due process rights even though the methodology used to produce the assessment was disclosed neither to the court nor to the defendant. While the Loomis court provided a procedural safeguard to alert judges to the dangers of these assessments – a 'written advisement' to accompany [presentencing

[52] R Richardson, J Schultz and K Crawford, 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice' (2019) 94 NYUL Rev. Online 15.

[53] <https://repository.tudelft.nl/assets/uuid:6f5ae910-be97-4860-9a25-0e1e386631ec/2900_volledige_tekst_tcm28-402015.pdf>.

[54] Loomis v Wisconsin, docket no. 16-6387 <https://www.scotusblog.com/case-files/cases/loomis-v-wisconsin/>.

> investigation reports] – this prescription is an ineffective means of altering judges' evaluations of risk assessments. The court's 'advisement' is unlikely to create meaningful judicial scepticism because it is silent on the strength of the criticisms of these assessments, it ignores judges' inability to evaluate risk assessment tools, and it fails to consider the internal and external pressures on judges to use such assessments.[55]

Two potential best practices were identified that could strengthen criminal procedural law in the twenty-first century. First, possibilities could be offered to individuals being prosecuted to request that evidence will be excluded based on, or other procedural consequences attached to, the presence of a bias distorting results or application of analysis in a discriminatory way. It is up to courts to review whether the use of Big Data systems or the use of its outcome constitutes a violation of, for example, due process or the right to fair trial when it is shown that a system presents bias towards a suspect or is used in a discriminatory way. Allowing systems and their outcomes to be challenged in this way is in line with traditional views on criminal justice, such as excluding non-digital evidence gathered in violation of prevailing legal safeguards or standards. Allowing individuals to raise issues concerning, errors, bias or discrimination of systems in their defence enables a societal interest to be addressed and might require the police and prosecution to amend their systems if their outcomes are deemed impermissible by courts. Still, even if such a right would be embedded in criminal procedural law, it might be practically difficult for an individual to both defend herself against the criminal charges against her and to challenge the data-driven application at large that led to her arrest.

That is why a second, though more radical, solution might be considered, namely to expand the possibilities for NGO's, such as Amnesty International, to participate in criminal proceedings. Although in most jurisdictions studied, there is a monopoly for the public prosecution to commence criminal cases, in France, various organisations can, based on article 2-1 of the *Code de procédure pénale,* join in and even initiate a criminal procedure if an interest they aim to protect is at stake.[56] Such interest may concern, for example, combatting racism, sexual violence, crimes against humanity or animal cruelty. Although originally, only certified associations were allowed to initiate an action under criminal law in the collective interest, this criterion has been increasingly relaxed in French case law. This would allow organisations that represent privacy interests or interests in due process, to act against Big Data practices that are possibly unlawful. Inspired by this legal figure, workshop participants also discussed the possibility of allowing such organisations to join criminal cases. When the government has relied, for example, on predictive policing models, the defendant could focus her defence on the question of culpability in her specific case, while a civil rights organisation could join the case to challenge the predictive policing system and its potential biases or flaws as such.

4. Access to justice

The use of Big Data does not only pose challenges to the traditional focus of legal protection on individuals, but also creates burdens and practical problems that are difficult to

[55]<https://harvardlawreview.org/2017/03/state-v-loomis>.
[56]Biard and Amaro (2016).

overcome. Granting citizens rights, assumes that they are also responsible for enforcing them. This can be problematic for two reasons, as discussed in the introduction of this article. First of all, some Big Data processes are not visible to citizens. Second, there are so many data flows, it is practically impossible for an individual to keep track of each party that has her data, evaluate how they are processed and determine whether this accords to the prevailing legal standards. As explained, individuals often lack the expertise, finances and incentive to start or join collective actions or public interest litigation. That is why it is likely that civil rights organisations and specialised NGO's would have an important role in evaluating the extent to which Big Data projects stay within the relevant legal frameworks.

However, besides the legal obstacles discussed in the previous two sections, civil rights organisations that are regularly involved in public interest litigation and collective actions, point to what they see as perhaps the biggest hurdle: the costs. They stress that precisely because these types of cases do not revolve around the specific matter of one particular individual, but around e.g. the validity of a legal regime as such, societal interests and questions over legality and legitimacy, the legal proceedings are often lengthy. Moreover, the matters often concern very complex issues, both from a legal, technical and societal angle, so that either specialised lawyers or experts need to be hired in preparation of a complaint. Finally, a reason for the high legal costs is that the organisations experimenting with large-scale data-driven applications are typically the bigger organisations in the private and the public sector, that already have the technical expertise on the matter complained, have the resources to hire the best lawyers in the field and are able to sit out long trials. Delving deeper into the issues of access to justice and practical effectiveness of the current legal framework, transparency is discussed (Sub-section 4.1), measures to make procedures less lengthy and costly are evaluated (Sub-section 4.2) and several alternatives to litigation funding are mapped (Sub-section 4.3).

4.1. Transparency

One of the civil rights organisations interviewed stressed that one of the problems they face is that many of the large-scale data-driven processes function below the radar. That means that in order to know whether a technology or application exists, how it functions in practice and whether this accords to the prevailing legal standards often requires a legal proceeding of its own.

> More information should be available about policies and operations, in particular in the Big Data context. Currently, before even evaluating whether there is a case which you might want to bring to a court of law, a legal proceeding is often necessary to enforce transparency and obtain information. That takes up a lot of time, energy and resources.[57]

All legal regimes studied require some form of transparency, but the level of openness mandatory is typically limited to two elements. First, informational privacy and data protection regimes require that citizens are informed of the fact that their personal information is being processed, why, by whom and how. Second, most administrative legal regimes require governmental organisations to explain to citizens that are affected,

[57] <https://repository.tudelft.nl/assets/uuid:6f5ae910-be97-4860-9a25-0e1e386631ec/2900_volledige_tekst_tcm28-402015.pdf>.

how decisions are made. Some informational privacy and data protection regimes also lay down such a duty for private sector organisations when it comes to automated decision-making For example, the GDPR requires transparency in automated decision-making and profiling; the organisation using profiling has to provide meaningful information concerning the underlying logic as well as the importance of the processing and expected consequences for the individual in question.[58]

But when systems run on aggregated data or when decisions do not affect specific individuals, but larger groups or society as a whole, such an obligation generally does not exist. As explained, even when it concerns personal information or decisions on an individual level, organisations are not required to share general information, for example about the algorithm they use, the data in their possession and the factors taken into account in decision making processes. In addition, legal regimes allow for exceptions to transparency obligations, for example when it concerns trade or company secrets of private sector organisations, or matters of public interests in the case of public sector organisations. For example, one interviewee was the head of the data analytics unit of a national Tax Authority. He stressed that if the algorithms and data points used by the authority to scan tax forms on likelihood of fraud, it would drastically reduce the effectiveness of such operations.[59] Similar arguments have been made by police, intelligence agencies and other public sector organisations.

A partial solution could be found through the introduction of transparency requirements for processing non-personal data and the algorithmic analysis of data, as discussed in Sections 2.1 and 2.2. But especially when there are prevailing interests of the data-driven organisations not to disclose certain elements, additional safeguards should be implemented as this raises concerns over accountability, equality of arms and fair trial. A typical example is when information or evidence gathered by law enforcement authorities or secret services is used in a court case against a citizen, either in criminal law (e.g. a person suspected of terrorism) or administrative law (e.g. a person is denied a residence permit on the grounds that she is considered a danger to national security), but such evidence or the methods of obtaining evidence cannot be disclosed.

One solution, though a highly controversial one, is the use of a special advocate, already part of many Common Law systems.[60] A special advocate is a lawyer appointed specifically to review sensitive evidence and information on behalf of plaintiffs or civilians and subsequently to challenge that evidence or information where appropriate. The Attorney General appoints the special advocate; these advocates are specialised in the topic and have undergone a security background check. The special advocate is not allowed to share the information with any party, including the plaintiffs or civilians that the special advocate is reviewing the information for.[61] Introducing the mechanism of a special advocate enables an instance of oversight on processes of government agencies that are naturally opaque, but even more so enables due process or protection of interests of citizens submitted to such processes. Special advocates can be introduced in criminal litigation

[58]GDPR arts. 13 & 22.
[59]<https://repository.tudelft.nl/assets/uuid:6f5ae910-be97-4860-9a25-0e1e386631ec/2900_volledige_tekst_tcm28-402015.pdf>.
[60]Prevention of Terrorism Act 2005: J Ip, 'The Rise and Spread of the Special Advocate' (2008) 3 Pub. L. 717.
[61]A Boon and S Nash, 'Special Advocacy: Political Expediency and Legal Roles in Modern Judicial Systems' (2006) 9 Legal Ethics 1.

dealing with police practices as well as in litigation that deals with government practices in the social welfare or tax domain or other domains where full disclosure is impossible.

A problem is that special advocates will act in the public interest; they are not appointed by the defence and they do not receive instruction from them. The autonomy of the citizen is consequently limited. Therefore, a special advocate should only be considered legitimate when it is absolutely necessary to deny the individual access to the documents related to her defence. The starting point should remain transparency and the use of the special advocate an exception. This system requires that strict conditions are laid down for situations in which transparency towards a citizen would not be possible and the special advocate deployed. A criminal proceedings judge or external committee could perform review of the decision to deploy the special advocate. When special advocates would be appointed in Big Data litigation, it should be made sure that they have sufficient knowledge of statistics or familiarity with data science, if their role is to be meaningful. The special advocate would get access to the relevant algorithms, input data and outcomes of analyses that cannot be released based on national security reasons or to safeguard the working of the methods deployed by governmental agencies.

4.2. Expanding the role of amicus curiae and pre-judicial questions

A second point that emerged was that the length and complexity of legal proceedings in the Big Data context. This was one of the reasons why civil rights organisations indicated they would not start a court case; such organisations generally have very small budgets and are often just about capable of keeping their heads above the water. When public interest litigation or collective actions were brought before a court, most cases were in summary proceedings only, which typically means that courts only marginally assess the matter. When a normal proceeding was initiated, oftentimes, parties only had enough resources to litigate one instance, but not go to a court of appeal or the supreme court, meaning that only lower level judges scrutinised important matters of collective and general interest. Two solutions emerged: granting a bigger role to amicus curiae proceedings and/or to pre-judicial questions.

Many jurisdictions allow civil society organisations or groups to submit an amicus brief to the court, the so called *amicus curiae* intervention or participation. The amicus brief is mainly used in Anglo-Saxon jurisdictions and in international courts, such as the European Court for Human Rights. Through an amicus brief, organisations can provide courts with information and opinions on the matter of the case, without being a formal party in the proceedings themselves. Such opportunity is generally regarded favourable by civil rights organisations. For example, after the presentation of Marc Rotenberg at one of the workshops, the following was concluded:

> The potential for amicus briefs is actually quite effective in the US. It allows parties to submit their expertise to the court. The most effective briefs are the objective, factual ones; courts will cite and use that information when they feel it is neutral. That is why organizations such as EPIC try to let scholars and other experts sign on to the amicus briefs, in order to show that the claims made in the letter are supported by a vast part of the scientific community.

In the jurisdictions studied, there were significant barriers to amicus briefs. For example, courts had adopted a list of organisations that are allowed to submit such briefs, either in general or per case; in some countries, *amicus curiae* participations was only permissible in administrative law, but not in civil and criminal law proceedings; and in some countries, there was scepticism about applying this instrument to public interest litigation. One Dutch study concluded:

> With regard to the goals for which the amicus curiae can be used, we also note that several state councils interviewed for this study have pointed out that the amicus curiae is bound by certain limits. For example, a state council indicated that the use of the amicus curiae is less obvious in cases that are politically or societally sensitive, because this can put the administrative judge in an uncomfortable position.[62]

Because many Big Data projects do touch upon politically sensitive questions, constitutional aspects and principles of legitimacy and legality and because damage or specific consequences are often hard to pinpoint, this would mean that, following the rule discussed, amicus briefs would not be allowed in such cases. However, it may be these types of cases that would benefit the most from information and analysis provided by expert organisations. That is why allowing for a bigger role of amicus curiae participation might ameliorate the current legal regime.

A second model, that could be used to solve the problem that legal matters are currently not brought before a court and if they are brought, will often only reach the court of first instance or be submitted in summary proceedings, is to work with pre-judicial questions. A number of countries studied allow lower courts to submit questions on the legal interpretation of laws or decisions and their constitutional validity to the supreme court. In addition, both supranational European courts allow national courts to submit prejudicial questions when those concern, for example, the question whether a national law or practice is compliant with the European Convention of Human Rights or the EU legal acquis. The first article of Protocol 16 of the ECHR specifies each national state's supreme court

> may request the Court to give advisory opinions on questions of principle relating to the interpretation or application of the rights and freedoms defined in the Convention or the protocols thereto. The requesting court or tribunal may seek an advisory opinion only in the context of a case pending before it. The requesting court or tribunal shall give reasons for its request and shall provide the relevant legal and factual background of the pending case.[63]

Article 267 of the Treaty on the Functioning of the European Union specifies along similar lines:

> The Court of Justice of the European Union shall have jurisdiction to give preliminary rulings concerning: (a) the interpretation of the Treaties; (b) the validity and interpretation of acts of the institutions, bodies, offices or agencies of the Union; Where such a question is raised before any court or tribunal of a Member State, that court or tribunal may, if it considers that a decision on the question is necessary to enable it to give judgment, request the

[62] JCA De Poorter, LA Van Heusden and CJ De Lange, 'De amicus curiae geëvalueerd: Over de eerste indrukken van de inzet van het instrument van de amicus curiae in procedures voor de Afdeling bestuursrechtspraak' (2018) <https://www.raadvanstate.nl/publish/pages/2294/evaluatie_van_de_amicus_curiae.pdf>.

[63] Protocol No. 16 to the Convention on the Protection of Human Rights and Fundamental Freedoms Strasbourg, 2.X.2013 <https://www.echr.coe.int/Documents/Protocol_16_ENG.pdf>.

> Court to give a ruling thereon. Where any such question is raised in a case pending before a court or tribunal of a Member State against whose decisions there is no judicial remedy under national law, that court or tribunal shall bring the matter before the Court. If such a question is raised in a case pending before a court or tribunal of a Member State with regard to a person in custody, the Court of Justice of the European Union shall act with the minimum of delay.[64]

Although this possibility already exists in a number of jurisdictions, it is infrequently used in Big Data cases, for which there are several reasons. First, as one lawyer working not only in the field of privacy, but also in the area of freedom of expression and intellectual property, stressed, courts simply do not use the opportunity to pose pre-judicial questions in privacy and data protection related cases.

> Precisely because I have experience with other areas of law, it is striking to see how little jurisprudence there is on privacy and data protection law, how little judges use, for example, the possibility to ask pre-judicial questions and how little public interest litigation has been initiated with respect to privacy related cases.[65]

He and other interviewees have stressed that litigating parties often hope that a lower court will pose pre-judicial questions on the matter brought to it, because this allows them to obtain an opinion by the highest court, without having to cover the costs normally needed to exhaust all remedies. Second, only courts can decide to pose a pre-judicial question to the supreme court or one of the two European supranational courts, not the litigating parties themselves. Finally, in many jurisdictions, there are limits as to the types of cases that are eligible for pre-judicial questions, for example only allowing such questions when the matter before the court or the legal question posed represents an issue that is relevant for a multitude of other claims and potential court cases. Consequently, pre-judicial questions could be posed only with respect to a matter that concerns hundreds or thousands of people, such as when there is a data breach which affects all persons whose personal information were contained in it; this would allow one person or organisation to bring the matter of one specific individual to a court, which could then pose a pre-judicial question to the supreme court, which could issue a ruling that applied to all cases similar to the one before it. But that does not apply to public interest litigation and the types of questions discussed in Section 3 of this article, such as whether a law or policy as such accords to the principles of the rule of law, legality and legitimacy. In order for the legal regime to function optimally in the twenty-first century, jurisdictions could consider also allowing pre-judicial questions to be asked in these types of cases.

4.3. Mitigating costs

Finally, there are two opportunities for covering the costs endured by either citizens or professional civil rights organisations that would want to challenge potential harm or matters of public concern relating to large-scale data processes before a court of law. One is particularly helpful for collective actions or class actions, the other can be used to cover the costs of both these types of proceedings and public interest litigation.

[64] Consolidated version of the Treaty on the Functioning of the European Union.
[65] <https://repository.tudelft.nl/assets/uuid:6f5ae910-be97-4860-9a25-0e1e386631ec/2900_volledige_tekst_tcm28-402015.pdf>.

First, in a number of jurisdictions, standard damages were set for certain types of harm. Such were either laid down in law, developed in jurisprudence by courts or set by the public prosecutor. Typically, these are matters in which similar cases are frequently brought before a court and with respect to which it may be fairly difficult to substantiate harm and causality between the matter of complained of and the harm endured. A typical example may be car incidents, environmental damage or people who were detained and put before trial, but were acquitted.[66] To avoid long and tedious discussions about how much harm had been endured by the applicant and how to quantify immaterial harm in monetary currency, standards are set, e.g. that a person is awarded 150 dollar per day in unlawful detention. Such could be introduced to violations with respect to large data-driven projects as well and would relieve litigating parties from specifying how much harm, for example, a data breach had caused in a person's specific case or how negative effects endured due to malignant nudging should be quantified. Such a solution would be particularly helpful in class actions and collective actions, where, for example, 10.000 people would complain of the same incident, such as an unlawful smart city program used by their municipality in which their reasonable expectation of privacy had been violated.

Second, there are several funds available that could be used for either class actions, public interest litigation or both. On EU level, there is the fund called 'Capacity building for litigating cases relating to democracy, rule of law and fundamental rights violations'.[67] There are also initiatives such as the Digital Freedom Fund, which funds strategic litigation in the area of, inter alia, Big Data.[68] In the Canadian system, mechanisms are put in place to make use of third party funding, based on funding arrangements that are scrutinised by courts. This will mainly concern funding provided by private entities where the court has to assess whether the funding is in the best interest of the class.[69] Next to third-party funding there is also public funding in Québec and Ontario. In Ontario there is the Class Proceedings Fund, which assists the representative plaintiffs and is accessible through a formal application process. In Québec, the Fonds d'aide aux actions Collectives has been in place since 1978 and can also be called upon through a formal application by the class representative. The funding can, for example, include reimbursement for the payment of legal fees or expert fees.[70]

For Big Data related cases, litigation funds could be set up to finance a handful of public interest litigation cases per year. Such could be financed by the government itself, in order to ensure that laws, policies and practices accord to all prevailing legal standards. Citizens and organisations could then apply for funding; an independent party would decide upon where money would be awarded and if so, how much. Another possibility would be to devise a fund that receives seed money and afterwards, has to become

[66]Bowden v Caldor, Inc., 1998 Md. Lexis 407 (1998).
[67]European Commission (2019). Capacity Building for Litigating Cases Relating to Democracy, Rule of Law and Fundamental Rights Violations <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/just-pppa-liti-ag-2018>.
[68]Digital Freedom Fund: <https://digitalfreedomfund.org/about/>.
[69]Québec Act Respecting the Fonds d'Aide Aux Actions Collectives. Ontario Class Proceedings Act. Also see: <https://blg.com/en/News-And-Publications/Documents/A-Summary-of-Canadian-Class-Action-Procedure-and-Developments_-_SEP2018.pdf>.
[70]<https://blg.com/en/News-And-Publications/Documents/A-Summary-of-Canadian-Class-Action-Procedure-and-Developments_-_SEP2018.pdf>.

self-sufficient by receiving a percentage of damages awarded in cases it has funded. Such an approach is adopted in Canada. Such a mechanism only works for class actions and not for public interest litigation because in the latter type of cases, where for example a court is asked to declare null and void a law or policy, typically no damages are awarded. In addition, setting up a self-sufficient fund by investing seed money only, would currently only work for Common Law countries, because the damages awarded in those jurisdictions are typically substantially higher than in Civil Law countries, where damages for immaterial harm following from data-driven practices is often somewhere between 0 and a couple of hundred euro's.[71] To make such an approach viable for Civil Law countries as well, a combination should be found with setting standards for damages following from data-driven practices that are at least more than a thousand euro per incident.

5. Conclusion

This article started with a basic premise: the current legal regime is based to a large extent on the individual, such as by granting natural persons subjective rights to defend their individual interests, while data-driven projects often affect large groups or society as a whole. Most reports and articles have focused on material rights necessary in the age of Big Data and on issues of material justice. This article has focused on the other side of the coin, namely on concerns over procedural justice and fairness, procedural law and access to justice. It has signalled a number of problems that might result in legal lacuna's when legal regimes would not be altered or revised. These include, but are not limited to, the fact that considerable parts Big Data processes are currently left unregulated, namely to the extent that they are not based on personal information and do not directly affect the interests of a natural person, the fact that there are significant limits to addressing group and societal interests through administrative, civil and criminal procedural law and that it is practically difficult to have access to justice when it is tedious to merely find out which data are processed by which organisations and the courts cases that concern the general interests potentially affected by Big Data processes are often lengthy, complex and hence costly.

In order to identify a number of best practices, the legal systems of Australia, Belgium, Canada, France, Germany, Israel, the Netherlands, New Zealand, the United Kingdom and the United States were studied. In addition, several key players were interviewed, such as lawyers involved in public interest litigation, civil society organisations and academics. To verify the results of the study, two workshops were held, in which civil servants, academics and policy makers were invited to respond to the preliminary findings and several leading experts gave their opinion on sub-aspects of the report, such as Wojciech Wiewiórowski (European Data Protection Supervisor), Marc Rotenberg (EPIC) and Max Schrems (NOYB). An overview of the most important results has been provided in this article. The discussion of the current legal regimes was not aimed at providing in depth analysis on specific points, but rather to discuss their main features and general commonalities. Likewise, each of the best practices discussed, whether it be the regulation of non-personal data, the pros and cons of the special advocate, the application of sunset clauses and the

[71] <https://repository.tudelft.nl/assets/uuid:6f5ae910-be97-4860-9a25-0e1e386631ec/2900_volledige_tekst_tcm28-402015.pdf>.

merits of a criminal law system that allows civil rights organisations to initiate proceedings, deserves an article or even a book of its own. The goal of this article was not to add something to the existing literature and discussions on those specific points, but rather to present a broad palette of potential solutions that could provide partial solutions to the problems identified. Obviously, when national legislators would consider adopting one or more of these legal figures, they should evaluate their merits in more detail, in particular paying attention to whether the legal figure would fit their legal system and legal culture, as examples have been drawn from Australia to Canada and from France to the United States, from Common Law and Civil Law countries and from different fields and areas of law.

In general, three types of best practices or solutions have been discussed.

The first group of proposals concern shifting the focus in the legal regimes from individual interests by natural persons to general interests. Not only does this have the added benefit of allowing for the protection of the various general and societal interests at stake, it also means that citizens are relieved from the burden of proving individualizable harm and the causal relationship between that harm and the data-driven process. In addition, these types of interests are relatively uncontroversial; everyone agrees that when organisations use resources, in the private, semi-private or public sector, such should be done in the most effective way possible; respecting the rule of law and principles of legality and legitimacy are the foundations of constitutional democracies; having standards of care for data analytics ensures that algorithmic processes are more accurate and reliable. Not only are general concerns over legitimacy, accuracy and effectiveness uncontroversial, they generally serve as a precondition, a question that should be answered even before turning to a potential 'balance' between private and public interests. In this light, it might be important to introduce general rules for the processing of non-personal data and for data-analytics, to work with sunset clauses when introducing data-driven programs and to allow for legal claims in administrative, constitutional and criminal law that regard the legality and legitimacy of laws and policies as such and address potential biases in datasets or flaws in data analytics applied, for example by civil rights organisations.

The second group of best practices concerned the representation of individuals and individual interests. A problem is that individuals are ill-equipped to defend their personal interests through law due to a number of points: they are often unaware of the fact that their data are processed and by whom or are oblivious to the fact that a data processing operation has affected their life; it is practically undoable for an individual to keep track of all the organisations that have her data, assess how her data are processed and evaluate whether this accords to the prevailing legal standards; it is often very difficult to substantiate personal and individualizable harm that can be distinguished from general effects, which most legal regimes require; and it is difficult to establish a causal relationship between that harm and the data driven application or technology. That is why several legal figures that allow for the representation of individuals in their claims were discussed. The most controversial one is using a special advocate, that represents a person with respect to information or evidence that cannot be disclosed to that person for reasons of, inter alia, national security. In terms of collective actions, one problem under a number of regimes is that citizens have to take an action to join a case, for example about a data breach that has affected her and a high number of other citizens;

however, in addition to the problems over awareness and capacity, those affected do not know each other nor have easy ways to contact each other. That is why an alternative could be to grant specialised organisations the right to represent citizens in their claims or use models of opt-out class actions. Alternatively, human rights organisations could join proceedings in amicus briefs, where they could insert expertise and knowledge on specific aspects of a case. Finally, broad possibilities for public interest litigation by civil society organisations would allow the material principles discussed in the first group of best practices to be invoked before a court of law.

A third group of solutions aimed at improving access to justice. One of the biggest challenges for organisations starting class actions and public interest litigation are the costs. Because these types of cases do not revolve around the specific matter of one particular individual, but around e.g. the validity of a legal regime as such, societal interests and questions over legality and legitimacy, the legal proceedings are often lengthy. Moreover, the matters often concern very complex issues, both from a legal, technical and societal angle, so that either specialised lawyers or experts need to be hired in preparation of a complaint. A final reason for the high legal costs is that the organisations experimenting with large-scale data-driven applications are the bigger organisations in the private and the public sector, that already have the technical expertise on the matter complained, have the resources to hire the best lawyers in the field and are able to sit out long trials. That is why several best practices have been discussed. First, by actively providing more transparency on the working of data-driven processes, the datasets and the algorithms used, the 'pre-phase' of court cases would be removed, through which citizens or organisations need to make formal requests or go to court merely to know which data is gathered, how they are processed and whether there might be a reason to start a legal procedure. Second, an option for reducing the costs for claimants is to allow courts to rely more on pre-judicial questions, so that parties obtain an answer to their legal question from the highest court, without having to cover the costs normally needed to exhaust all remedies. Third, various ways for covering costs can be explored, either by setting fixed damages for types of data harms, by creating funds that support class actions by investing seed money for a fund investing in class actions in return for part of the damages awarded when the case is won or by establishing a fund that would cover the finances of a handful of public interest litigations per year.

Disclosure statement

No potential conflict of interest was reported by the author(s).