POLITECNICO DI TORINO
Repository ISTITUZIONALE

AI and Big Data: a blueprint for a human rights, social and ethical impact assessment

Original
AI and Big Data: a blueprint for a human rights, social and ethical impact assessment / Mantelero, Alessandro. - In: COMPUTER LAW & SECURITY REPORT. - ISSN 0267-3649. - STAMPA. - 34:4(2018), pp. 754-772. [10.1016/j.clsr.2018.05.017]

Availability:
This version is available at: 11583/2711554 since: 2020-05-05T12:35:10Z

Publisher:
Elsevier

Published
DOI:10.1016/j.clsr.2018.05.017

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Elsevier postprint/Author's Accepted Manuscript

© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/.The final authenticated version is available online at: http://dx.doi.org/10.1016/j.clsr.2018.05.017

(Article begins on next page)

12 October 2024

COMPUTER LAW & SECURITY REVIEW 34 (2018) 754–772

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

AI and Big Data: A blueprint for a human rights, social and ethical impact assessment

*Alessandro Mantelero**

Department of Management and Production Engineering, Polytechnic University of Turin, Torino, Italy

ARTICLE INFO

Article history:

Keywords:
Data protection
Impact assessment
Data protection impact assessment
Human rights
Human rights impact assessment
Ethical impact assessment
Social impact assessment
General Data Protection Regulation

ABSTRACT

The use of algorithms in modern data processing techniques, as well as data-intensive technological trends, suggests the adoption of a broader view of the data protection impact assessment. This will force data controllers to go beyond the traditional focus on data quality and security, and consider the impact of data processing on fundamental rights and collective social and ethical values.

Building on studies of the collective dimension of data protection, this article sets out to embed this new perspective in an assessment model centred on human rights (Human Rights, Ethical and Social Impact Assessment-HRESIA). This self-assessment model intends to overcome the limitations of the existing assessment models, which are either too closely focused on data processing or have an extent and granularity that make them too complicated to evaluate the consequences of a given use of data.

In terms of architecture, the HRESIA has two main elements: a self-assessment questionnaire and an ad hoc expert committee. As a blueprint, this contribution focuses mainly on the nature of the proposed model, its architecture and its challenges; a more detailed description of the model and the content of the questionnaire will be discussed in a future publication drawing on the ongoing research.

© 2018 Alessandro Mantelero. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Risk assessment models today play an increasing role in data protection, as recently confirmed by the EU General Data Protection Regulation (hereinafter GDPR).[1] Various types of assessment models can be adopted: they may be mandatory or voluntary, self-assessments or third-party/licensing schemes. They can only assess specific kinds of data processing or types of risk. They may be risk/benefit assessments or rights-based assessments. Finally, they may only focus on the legal issues or encompass societal issues as well.

Against this background, the first question we need to ask when defining an assessment model is whether the model is to be sector-specific or general. This is an important question, since data uses are not circumscribed by a specific domain or technology.

It hardly seems possible to adopt a technology-specific approach, for example, an IoT impact assessment, a Big Data impact assessment, a smart city impact assessment or an AI

* Corresponding author: Department of Management and Production Engineering, Politecnico di Torino, C.so Duca degli Abruzzi, 24, Torino 10120, Italy.
E-mail address: alessandro.mantelero@polito.it

[1] See Articles 25 and 26, GDPR.

https://doi.org/10.1016/j.clsr.2018.05.017
0267-3649/© 2018 Alessandro Mantelero. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

impact assessment.[2] All these technologies use data processing for decision-making: they differ in their methods but not in their scope. For this reason, and because the rights and values to be safeguarded are the same in these different contexts - regardless of the technology used, the model proposed here is not a technological assessment,[3] but a rights-based and values-oriented model.

In the context of data-driven applications, an assessment focused on a specific technology looks to be inadequate and only partially effective.[4] On the other hand, taking into account the various application domains (e.g. healthcare or crime prevention), different sets of rights, freedoms and values should be considered. So, a sector-specific approach focuses on the rights and values in question rather than the technology.

Thus, sectoral models concentrate their attention, not on the technology, but on the context and the values that assume relevance in that context.[5] This does not mean that the nature of the technology has no importance in the assessment process as a whole: a given technology determinates the most appropriate measures to take to safeguards the benchmark values.

Adopting a value-oriented approach, the assessment should focus on the societal impact of data use. This impact encompasses the potential negative outcomes on a variety of fundamental rights and principles and also takes into account the ethical and social consequences of data processing.[6]

In addressing these issues, this article builds on the results of previous research on data protection regulation in the context of data-intensive applications for decision-making processes. These works point out the criticisms affecting data protection in this context – which is dominated by an extensive use of Big Data analytics, algorithms and AI – and suggest the development of broader forms of data protection impact assessment, which also looks at the social impact and encourage a values-oriented use of data.[7]

In an initial approach, a mandatory multiple impact assessment was suggested to address these issues in an attempt to provide stronger safeguards for individuals.[8] However, a mandatory procedure encompassing societal issues was perceived as excessively burdensome and complex by business. This article therefore reconsiders the nature of the assessment and recommends a voluntary model,[9] which retains data controllers' freedom of decision, making this assessment a more acceptable solution than compulsory provisions.

Furthermore, a voluntary approach is more consistent with the existing legal framework, which seems to have difficulties in going beyond mere data protection in information use. In this sense, the GDPR – which provides one of the most advanced examples of regulation in this area – focuses on risk

[2] See AI Now Institute, 'Algorithmic Impact Assessments: Toward Accountable Automation in Public Agencies' (2018) <https://medium.com/@AINowInstitute/algorithmic-impact-assessments-toward-accountable-automation-in-public-agencies-bd9856e6fdde> accessed 4 March 2018.

[3] See Barbara Skorupinski and Konrad Ott, 'Technology assessment and ethics' (2002) 1(2) Poiesis & Praxis 95–122.

[4] In some cases, it is hard to define the borders between the different data processing fields and the granularity of the subject matter (e.g. the blurred confines between well-being devices/apps and medical devices).

[5] Specific impact assessments for Big Data analytics and for AI are not necessary, but we do need separate impact assessments for data-driven decisions in healthcare and another for smart cities, given the different values underpinning the two sectors. Whereas, for example, civic engagement and participation and equal treatment will be the driving values behind smart city technologies impact assessment, in healthcare freedom of choice and no-harm principle may play a more critical role. Differing contexts have different "architectures of values" that should be taken into account as a benchmark for the assessment models.

[6] See also Skorupinski and Ott (n 3) 101 ("Talking about risk […] is not possible without ethical considerations […] when it comes to a decision on whether risk is to be taken, obviously an orientation on norms and values is unavoidable").

[7] See Alessandro Mantelero, 'Personal data for decisional purposes in the age of analytics: from an individual to a collective dimension of data protection' in this Review (2016), vol 32, issue 2, 238–255; Alessandro Mantelero, 'The future of consumer data protection in the E.U. Rethinking the 'notice and consent' paradigm in the new era of predictive analytics' in this Review (2014), vol 30, issue 6, 643–660. See also Alessandro Mantelero, 'Regulating Big Data. The guidelines of the Council of Europe in the Context of the European Data Protection Framework' in this Review (2017), vol 33, issue 5, 584-602.

[8] See Mantelero, 'The future of consumer data protection in the E.U.' (n 7), p. 654–659. See also David Wright, 'A framework for the ethical impact assessment of information technology (2011) 13(3) Ethics and Information Technology 199–226; Paul M. Schwartz, 'Data Protection Law and the Ethical Use of Analytics' Data Protection Law and the Ethical Use of Analytics' (The Centre for Information Policy Leadership, 2011) <http://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/data_protection_law_and_the_ethical_use_of_analytics__paul_schwartzwhite_paper_2010_.pdf> accessed 18 December 2017.

[9] An important contribution in refining this proposal came from the Guidelines on Big Data issued by the Council of Europe in 2017, where a focus on the ethical and social consequences of data use was adopted by the members of the Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data. However, the discussion, which also engaged representatives of various stakeholders, outlined the difficulties in adopting a mandatory ethical and social assessment as part of the traditional data protection assessment. See Council of Europe, 'Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data', adopted in January 2017 and available at <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a> accessed 4 May 2017. Disclosure: the author had the privilege to be appointed as consultant expert in drafting the text of the guidelines and to follow the discussion of the proposal by the representatives of the Parties to Convention 108 in the Bureau of the Consultative Committee of Convention 108 and the Plenary Meeting. Concern about the mandatory nature of the proposed assessment and its consequences in terms of use of resources was also expressed by several commentators, some belonging to sectors of industry, during the presentation of my proposal on a mandatory assessment in the European workshop on "Algorithmic decision making and human rights implications" (Alexander von Humboldt Institute for Internet and Society - Hans-Bredow-Institute for Media Research, Berlin 2017), the Amsterdam Privacy Conference (Amsterdam, 2015) and the 9th International Conference on Legal, Security and Privacy Issues in IT Law (Lisbon, 2014).

assessment, but it is still far from a mandatory model encompassing societal issues.

The EU legislator recognises data processing risks such as discrimination and "any other significant economic or social disadvantage",[10] and the Article 29 Data Protection Working Party[11] and the European Data Protection Supervisor[12] suggest a broader assessment including analysis of the societal and ethical consequences of data use. However, despite these steps in the direction of an assessment no longer primarily focused on data quality and data security, Article 35 of the GDPR and the early assessment models from Data Protection Authorities (hereinafter DPAs) do not adequately highlight ethical and social issues.[13]

In this scenario, there is a clear tension between the increasing demand for ethically and socially oriented data use from citizens,[14] companies,[15] developers and computer scientists, on the one hand, and the lack of a regulatory framework to address these issues, on the other. Although this gap is partially filled by a variety of bottom-up initiatives,[16] corporate guidance[17] or ongoing public investigations,[18] the main limi-

[10] Recital n. 75, GDPR.

[11] See Article 29 Data Protection Working Party, 'Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679', adopted on 4 April 2017 as last revised and adopted on 4 October 2017 <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236> accessed 13 April 2018.

[12] See EDPS - Ethics Advisory Group, 'Towards a digital ethics' (2018) <https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf> accessed 4 March 2018.

[13] See e.g. CNIL, 'Privacy Impact Assessment (PIA). Knowledge Bases' (2018) <https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-3-en-knowledgebases-2018-02-19_diffusable_en_pdf_valide_jli.pdf> accessed 28 February 2018; CNIL, 'Privacy Impact Assessment (PIA). Methodology' (2018) <https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology-2018-02-19_diffusable_en_pdf_valide_jli.pdf> accessed 28 February 2018; CNIL, 'Privacy Impact Assessment (PIA). Templates' (2018) <https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-2-en-templates.pdf> accessed 28 February 2018; Information Commissioner's Office, 'Data Protection Impact Assessments draft guidance for consultation' (2018) and Information Commissioner's Office, 'Data Protection Impact Assessments draft template for consultation' (2018) <https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/data-protection-impact-assessments-dpias-guidance/> accessed 30 April 2018; Agencia Española de Protección de Datos, 'Guía Práctica de Análisis de riesgos en los tratamientos de datos personales sujetos al RGPD' (2018) <https://www.agpd.es/portalwebAGPD/canaldocumentacion/publicaciones/common/Guias/2018/AnalisisDeRiesgosRGPD.pdf> accesed 4 March 2018; Agencia Española de Protección de Datos, 'Guía práctica para las evaluaciones de impacto en la protección de los datos sujetas al RGPD' (2018) <https://www.agpd.es/portalwebAGPD/canaldocumentacion/publicaciones/common/Guias/2018/Guia_EvaluacionesImpacto.pdf> accesed 4 March 2018; Autoridad Catalana de Protección de Datos, 'Guía sobre la evaluación de impacto relativa a la protección de datos en el RGPD (2.0)' (January 2018) <http://apdcat.gencat.cat/web/.content/03-documentacio/Reglament_general_de_proteccio_de_dades/documents/GUIA-EVALUACION-DE-IMPACTO-CAST-2.0.pdf> accessed 28 February 2018.

[14] See, *ex multis*, Politico Staff, 'Full text: Mark Zuckerberg's Wednesday testimony to Congress on Cambridge Analytica' *Politico* (9 April 2018) <https://politi.co/2GNxFLx> accessed 9 May 2018; Llàcer, M. R., Casado, M., & Buisan, L. (eds), 'Document on bioethics and Big Data: exploitation and commercialisation of user data in public health care' (Barcelona, 2015); ProPubluca, series 'Machine Bias Investigating Algorithmic Injustice' <https://www.propublica.org/series/machine-bias> accessed 30 April 2018.

[15] See, in this sense, the increasing propensity of the big data-intensive and high-tech companies to set up their own ethics committees or advisory boards. See, e.g., Natasha Lomas, 'DeepMind now has an AI ethics research unit. We have a few questions for it…' *TechCrunch* (4 October 2017) <http://social.techcrunch.com/2017/10/04/deepmind-now-has-an-ai-ethics-research-unit-we-have-a-few-questions-for-it/> accessed; Axon AI Ethics Board <https://it.axon.com/info/ai-ethics> accessed 9 May 2018; DNA Web Team, 'Google drafting ethical guidelines to guide use of tech after employees protest defence project' *DNA India* (15 April 2018) <http://www.dnaindia.com/technology/report-google-drafting-ethical-guidelines-to-guide-use-of-tech-after-employees-protest-defence-project-2605149> accessed 7 May 2018. See also United Nations, 2011. Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. United Nations Human Rights Council (UN Doc. HR/PUB/11/04). See also Jordan Novet, 'Facebook Forms Ethics Team to Prevent Bias in AI Software', 3 May 2018. <https://www.cnbc.com/2018/05/03/facebook-ethics-team-prevents-bias-in-ai-software.html> accessed 9 May 2018; Microsoft, 'FATE: Fairness, Accountability, Transparency, and Ethics in AI'. Microsoft Research (blog) <https://www.microsoft.com/en-us/research/group/fate/> accessed 11 May 2018.

[16] See, e.g., Ester Fritsch, Irina Shklovski and Rachel Douglas-Jones, 'Calling for a revolution: An analysis of IoT manifestos' (2018) Proceedings of the 2018 ACM Conference on Human Factors in Computing (Montreal, Canada, 21-26 April 2018) <http://delivery.acm.org/10.1145/3180000/3173876/paper302.pdf?ip=80.180.146.48&id=3173876&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&__acm__=1525873755_622581693e4344f67627f0aec1be630b> accessed 3 May 2018.

[17] See above fn 15. See also the Asilomar AI Principles <https://futureoflife.org/ai-principles/> accessed 27 March 2018.

[18] See, e.g., Villani Cédric and others, 'Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne' (2018) <http://www.ladocumentationfrancaise.fr/rapports-publics/184000159/index.shtml> accessed 14 April 2018 ; House of Lords - Select Committee on Artificial Intelligence, 'AI in the UK: ready, willing and able?' (2018) <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf> accessed 16 April 2018. See also the ongoing initiative of the European Union Agency for Fundamental Rights (FRA) to map the impact of big data on fundamental rights <http://fra.europa.eu/en/event/2018/mapping-impact-big-data-fundamental-rights> accessed 5 April 2018, the ongoing studies on data processing and AI launched by the Council of Europe (MSI-AUT Committee and Consultative

tations of these initiatives concern the variety of values, approaches and models adopted.[19]

Against this background, this article tries to sketch out a uniform model, which provides a general common ground for value assessment in data processing[20] and, at the same time, offers a sufficient level of flexibility to give voice to different viewpoints. This not only provides a more systematic risk assessment scheme, described here in its main elements, but also outlines a more coherent theoretical framework for the proposed model.

Regarding the safeguarded interests that should be considered in assessing the potential negative impacts of data use, studies in the field of collective data protection[21] have pointed out the importance of the social and ethical implications of data processing in the context of data-intensive applications.[22]

Predictive policing software, credit scoring models and many other algorithmic decision-support systems highlight how data analysis strategies are centred on groups and society at large. The potential negative outcomes of data use are, therefore, no longer restricted to the more widely recognised privacy-related risks (e.g. illegitimate use of personal information, data security), but also include other potential prejudices (e.g. discrimination) that can be better addressed by placing data processing in the broader context of human rights.[23]

This article proposes a model, which is a variation of the Human Rights Impact Assessment.[24] The characteristic and particular features of this model can be seen from a comparison with other existing assessment strategies, such as the Privacy Impact Assessment (PIA), the Social Impact Assessment (SIA) and the Ethical Impact Assessment (EtIA).[25] On the one hand, the limitations affecting the existing PIA models and the Data Protection Impact Assessment (DPIA) described in the GDPR provide the key reason to embrace a broader standpoint, moving towards a Human Rights Impact Assessment (HRIA).[26] On the other hand, the granularity and the coverage of the SIA/EtIA models make them unsuitable as candidates for a general assessment of the consequences of a given data use.

The HRIA is not a new approach *per se*.[27] It has its roots in the environment impact assessment models and develop-

Committee of the Convention for the protection of individuals with regard to automatic processing of personal data, and the European Commission's Call for a High-Level Expert Group on Artificial Intelligence <https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence> accessed 12 April 2018.

[19] See European Commission - European Group on, Ethics in Science and, & New Technologies, 'Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems' (2018) 11 <https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf> accessed 4 April 2018 ("Current efforts represent a patchwork of disparate initiatives").

[20] See, in this sense, European Commission - European Group on, Ethics in Science and, & New Technologies (n 19) 11–12 ("There is a clear need for a collective, wide-ranging and inclusive process that would pave the way towards a common, internationally recognised ethical framework for the design, production, use and governance of AI, robots and 'autonomous' systems […] This statement calls for the launch of such a process and proposes a set of fundamental ethical principles and democratic prerequisites that could also guide reflection on binding law").

[21] See Mantelero, 'Personal data for decisional purposes' (n 7); Linnet Taylor, Luciano Floridi, Bart van der Sloot (eds), Group Privacy: New Challenges of Data Technologies (Springer International Publishing, 2017); Anton H. Vedder, 'Privatization, Information Technology and Privacy: Reconsidering the Social Responsibilities of Private Organizations' in Geoff Moore (ed) *Business Ethics: Principles and Practice* (Business Education Publishers 1997) 215–226; David Wright and Michael Friedewald, 'Integrating privacy and ethical impact assessments' (2013) 40(6) Science and Public Policy 755–766; David Wight and Emilio Mordini, 'Privacy and Ethical Impact Assessment' in David Wright and Paul De Hert (eds) *Privacy Impact Assessment* (Springer Netherlands 2012) 397–418; Charles Raab and David Wright, 'Surveillance: Extending the Limits of Privacy Impact Assessment' in David Wright and Paul De Hert (eds) *Privacy Impact Assessment* 363–383.

[22] See also Bernd Carsten Stahl and David Wright, 'Proactive Engagement with Ethics and Privacy in AI and Big Data - Implementing responsible research and innovation in AI-related projects' (2018) <https://www.dora.dmu.ac.uk/xmlui/handle/2086/15328> accessed 26 April 2018.

[23] Article 2, UN Universal Declaration of Human Rights (1948); art. 14, Council of Europe's Convention for the Protection of Human Rights and Fundamental Freedoms; art. 21 EU Charter of Fundamental Rights of the European Union. See also The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, 'Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1. IEEE, 2016' 16 <http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html> accessed 21 February 2018; Giovanni Sartor, 'Human Rights and Information Technologies' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds) *The Oxford Handbook of Law, Regulation, and Technology* (Oxford University Press 2017) 424–450.

[24] See below Section 2.

[25] Although other authors, e.g. Wright and Mordini (n 21), use the acronym EIA for Ethical Impact Assessment, the different acronym EtIA is used here to avoid any confusion with the Environmental Impact Assessment, which is usually identified with the acronym EIA.

[26] The notion of human rights adopted here refers to the rights recognised by the international human rights declarations. See also European Parliament, Resolution of 23 October 2008 on the impact of aviation security measures and body scanners on human rights, privacy, personal dignity and data protection <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2008-0521+0+DOC+XML+V0//EN> accessed 12 December 2017, where European Parliament asked the Commission to carry out an impact assessment relating to fundamental rights.

[27] See, *ex multis*, The Danish Institute for Human Rights, *Human rights impact assessment guidance and toolbox* (The Danish Institute for Human Rights, 2016) <https://www.humanrights.dk/business/tools/human-rights-impact-assessment-guidance-and-toolbox> accessed 20 December 2017; Paul De Hert, 'A Human Rights Perspective on Privacy and Data Protection Impact Assessments' in Wright and De Hert (n 21), 33-76; James Harrison, and Mary-Ann Stephenson, *Human Rights Impact Assessment: Review of Practice and Guidance for Future Assessments*. (Scottish Human Rights Commission, 2010) <http://fian-ch.org/content/uploads/HRIA-Review-of-Practice-and-Guidance-for-Future-Assessments.pdf > accessed 29 November 2017; Simon Mark Walker, The Future of Human Rights Impact Assessments of Trade Agreements. (Utrecht: G.J. Wiarda Institute for Legal Research 2009) <https://dspace.library.uu.nl/bitstream/handle/1874/36620/walker.pdf?sequence=2> accessed 26 April 2018.

ment studies,[28] but it has not yet been applied in the context of data processing.[29] Moreover, the HRIA can be enhanced by considering ethical and societal issues, which are playing an ever more central role today, given the enormous changes to society brought by technology and datafication.

Ethical and societal values necessarily influence the balance between the different interests in the HRIA model and attention to these values makes it possible to adopt a broader assessment covering questions that are not always properly addressed by human rights jurisprudence, data protection regulations or other laws safeguarding individual rights and freedoms. Moreover, the importance of ethical and social perspectives allows data controllers to better address the broader perspective of collective data protection, which may be partially limited by the individual dimension of fundamental rights and freedoms.

For these reasons, the Human Rights Impact Assessment in data protection should evolve into a more complete Human Rights, Ethical and Social Impact Assessment (HRESIA). Furthermore, the attention paid to the collective dimension of data protection suggests a design based on a participatory process, open to the contribution of the different stakeholders and characterised by a certain degree of transparency.

Finally, it should be pointed out how the HRESIA model differs from other broader approaches oriented more closely towards a Responsible Research Innovation assessment. The latter takes into account a variety of different societal issues, which do not necessarily concern fundamental rights and freedoms[30] (e.g. interoperability, openness).[31]

Having defined the general framework, Section 2 describes the boundaries of the proposed model and its main focus, while Section 3 discusses the main components of the HRESIA, which include expert committees that concur in defining the rights, freedoms and values that play a role in the assessment of a given data-driven application.

2. Outline of the HRESIA model

The use of algorithms in the context of modern data processing techniques[32] as well as data-intensive technological trends[33] have led to the adoption of a broader viewpoint in bringing into focus the issues concerning data processing. This has forced groups of experts[34] and scholars[35] to go beyond the traditional sphere of data protection and consider the impact of data use on fundamental rights and collective social and ethical values. This article sets out to embed these various suggestions in an assessment model with an architecture made up of two main elements: a self-assessment tool (questionnaire) and an *ad hoc* expert committee.

The assessment tool is used to define the framework – in terms of values – that data-intensive systems should comply with, while the expert committee contextualises this framework in a given data-intensive application. In this way, the general values can be operationalised by means of a tailored application consistent with the data processing context.

This approach does not involve any new procedures. Several assessment models are based on a set of benchmark values or principles and an assessment entity that applies these values/principles in a concrete case. Here the main challenge is represented by the complexity and variety of values that can be adopted as a benchmark.

Against this background, there are three preliminary decisions to be made regarding the benchmark. The first concerns the dimension to adopt when defining the values used in the model (common/universal values or local and context-specific values). The second concerns the type of assessment to be adopted (a value-oriented assessment or a risks/benefits assessment). The last one regards the values to be considered (legal or societal and ethical values).

These are not necessarily binary decisions. For example, universal and local approaches may be combined. The way these issues are addressed affects the core elements of the proposed model. To provide an overall idea of the model, the key decisions concerning its architecture are presented in this section, while the following sections focus on the rationale behind these decisions.

Although the proposed model combines the human rights assessment approach with attention to the societal and ethical consequences of data use,[36] this is not a broad social impact assessment, but remains focused on human rights. In this sense, ethical and social values are seen through the lens

[28] See Walker (n 27), 3–4; Tarek F. Massarani, Margo Tatgenhorst Drakos, and Joanna Pajkowska, 'Extracting Corporate Responsibility: Towards a Human Rights Impact Assessment' (2007) 40(1) Cornell International Law Journal 135, 143–149. See also Rabel J. Burdge and Frank Vanclay, 'Social Impact Assessment: A Contribution to the State of the Art Series' (1996) 14(1) *Impact Assessment* 59, 62–64.

[29] A suggestion in this sense was provided by the Council of Europe-Committee of experts on internet intermediaries (MSI-NET), 'Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications' (2018), 45 <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5> accessed 8 April 2018 ("Human rights impact assessments should be conducted before making use of algorithmic decision-making in all areas of public administration").

[30] Regarding this sort of hendiadys ("fundamental rights and freedoms"), see also Paul De Hert and Serge Gutwirth, 'Rawls' political conception of rights and liberties. An unliberal but pragmatic approach to the problems of harmonisation and globalisation' in Van Hoecke, Mark, *Epistemology and methodology of comparative law in the light of European Integration*. (Hart Publishing, 2004) 319-320 ("legal scholars in Europe have devoted much energy in transforming or translating liberty questions into questions of 'human rights'. One of the advantages of this 'rights approach' is purely strategic: it facilitates the bringing of cases before the European Court of Human Rights, a Court that is considered to have higher legal status […] There are however more reasons to think in terms of rights. It is rightly observed that the concept of human rights in legal practice is closely linked to the concept of subjective rights. Lawyers do like the idea of subjective rights. They think these offer better protection than 'liberty' or 'liberties' ").

[31] Regarding this approach in the context of data processing, see also H2020 Virt-EU project <https://virteuproject.eu/> accessed 19 December 2017.

[32] See Council of Europe-Committee of experts on internet intermediaries (MSI-NET) (n 29).

[33] See EDPS - Ethics Advisory Group (n 12).

[34] See above fn. 32 and 33.

[35] See above fn. 21.

[36] See below Section 2.

of human rights and are used to go beyond the limitations that the legal theory or practical implementation of such rights may imply in effectively addressing the current issues concerning the societal impacts of data use.

Moreover, ethical and social values are key to the interpretation of human rights in coherence with the regional context, in many cases representing the unspoken aspect of the legal reasoning behind the decisions of the Data Protection Authorities (DPAs) and courts.[37] In this sense, the suggested model is a Human Rights, Ethical and Social Impact Assessment (HRESIA).

The model proposed here is intended to provide a self-assessment tool, which data controllers can use to identify values and give them a clearer perception when designing their products/services. However, general background values and their contextual application may be not enough to address the societal changes when designing data-intensive systems. Although balanced with respect to the context, the definition of such rights and values may remain theoretical and need to be further tailored to the specific data processing application.

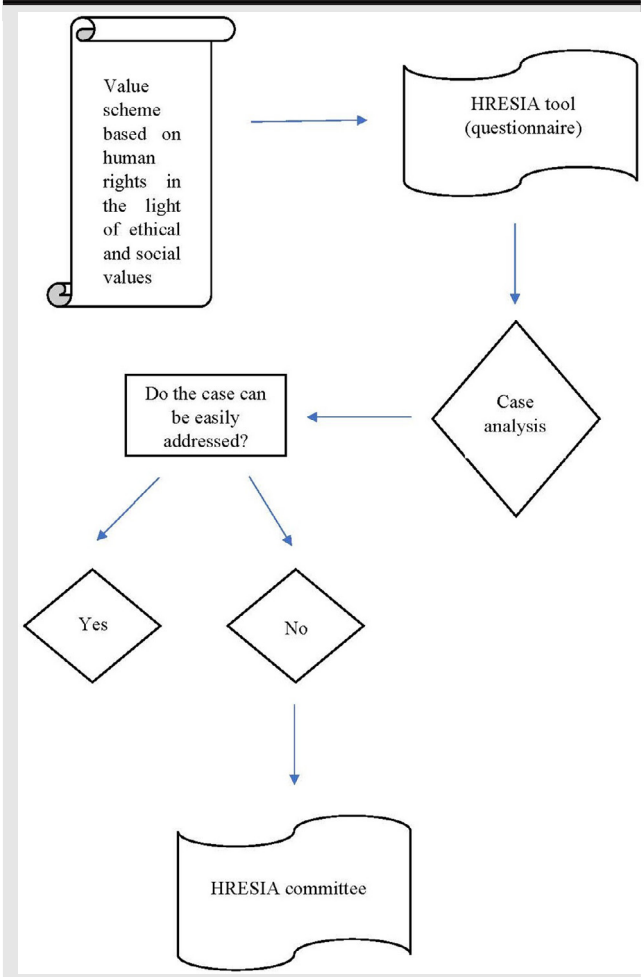
To achieve a balance in specific cases, individuals with the right skills are needed to apply this set of rights and values in the given situation. There are cases though in which bridging the gap between the theory of rights and values and their concrete application is complicated by the nature of data use and the complexity of the associated risks. In such cases, the assessment should be carried out by an ad hoc panel of experts, just as ethics committees[38] apply general principles and guidelines to a specific case.

This second element (the HRESIA committee) of the model makes it easier to provide specific answers to the issues raised by the given application (Table 1). The *ad hoc* committee also supports the development of an assessment characterised by transparency, participation and circularity.

Given the social issues that underpin this model, an essential requirement of the HRESIA is transparency. In this sense, the assessment is not only designed to mitigate the societal consequences, but also to give data subjects a better understanding of the data processing and, therefore, greater self-determination. Transparency is thus the basis for a participatory approach, as can be seen in other fields where impact assessments concern the societal consequences of technology (e.g. environment impact assessments). But transparency does not entail full disclosure of the assessment procedure and must be balanced against the safeguarding of other interests recognised by law (e.g. industrial secrets).

Finally, along the lines of risk management models, the assessment process should be characterised by a circular approach from the earliest stages (Table 2). This is also consistent with the circular product development models that focus on flexibility and interaction with users to address customers' needs,[39] which in this case are also legal and societal requirements (Table 3).

Table 1 – HRESIA model.



2.1. *Comparison with other assessment models*

The focus on data processing and the legal and societal impacts of data use require us to compare HRESIA with existing impact assessment models, both the specific data protection models (Privacy Impact Assessment-PIA and Data Protection Impact Assessment–DPIA) and those more interested in the societal (Social Impact Assessment–SIA) and ethical (Ethical Impact Assessment - EtIA[40]) consequences.

[37] See below Section 2.1.1.

[38] See Javier Arias Díaz and others, 'Ethics assessment and guidance in different types of organisations Research Ethics Committees' SATORI project <http://satoriproject.eu/media/3.a-Research-ethics-committees.pdf> accessed 23 April 2018. See also World Health Organization, *Research ethics committees basic concepts for capacity-building* (World Health Organization, 2009).

[39] See e.g., with regard to software development, the Manifesto for Agile Software Development <http://agilemanifesto.org/> accessed 5 February 2018. See also Seda Gürses and Joris Van Hoboken, 'Privacy after the Agile Turn' in Jules Polonetsky, Omer Tene, and Evan Selinger (eds) *Cambridge Handbook of Consumer Privacy* (Cambridge University Press, 2017) <https://osf.io/preprints/socarxiv/9gy73/ or https://osf.io/ufdvb/> accesed 28 March 2018.

[40] See SATORI project. 'Ethics assessment for research and innovation — Part 2: Ethical impact assessment framework' 6 <http://satoriproject.eu/media/CWA-SATORI_part-2_WD4-20170510W.pdf> accessed 24 April 2018, which defines ethical impact as the

Table 2 – HRESIA and product/service development.

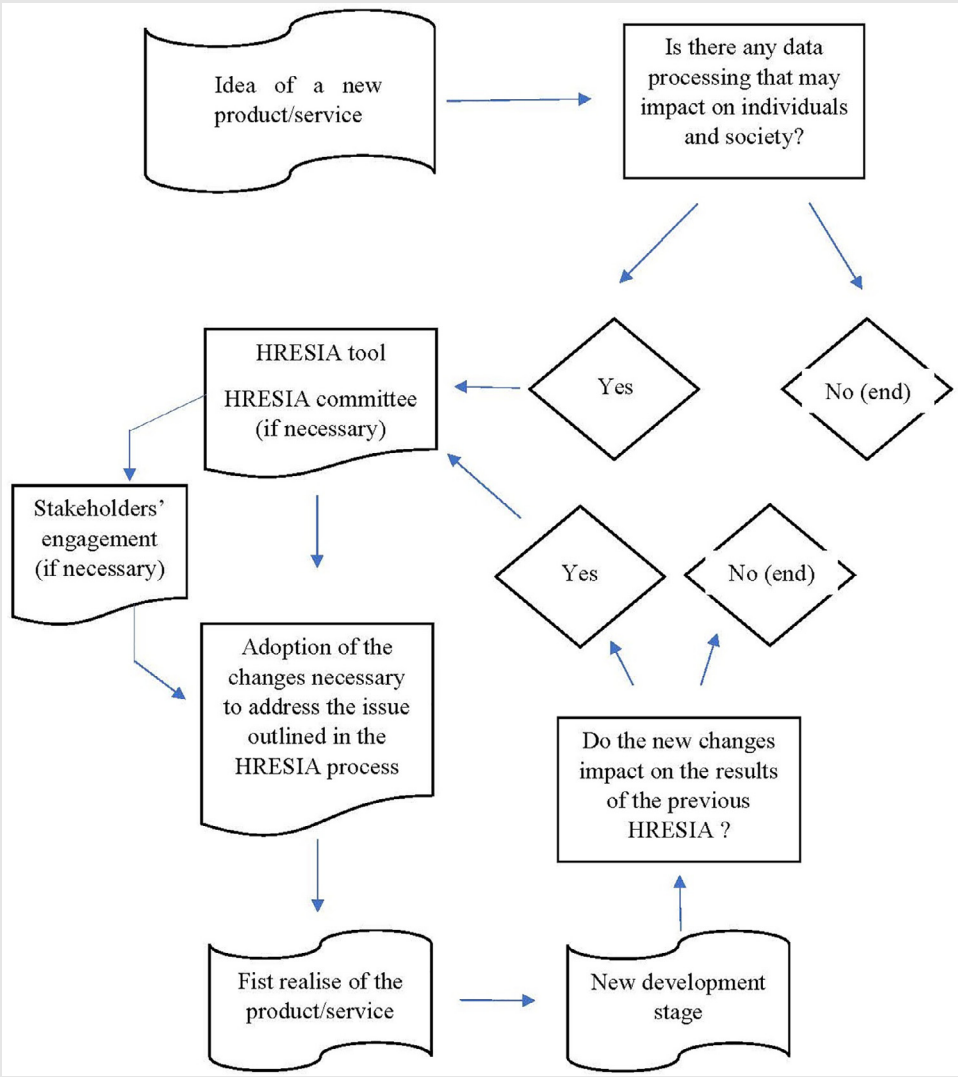
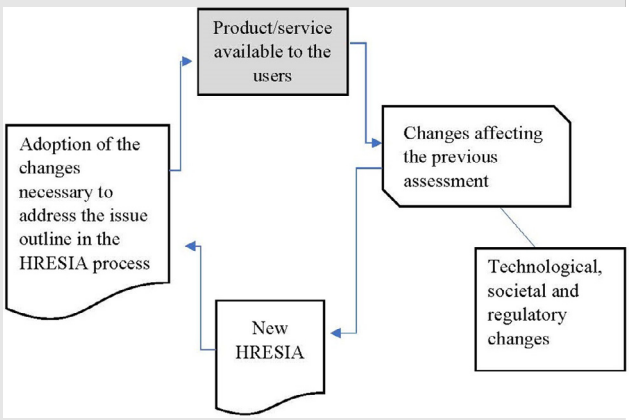


Table 3 – HRESIA and product/service life-circle.



This comparison takes a progressive approach, from impact assessments mainly focused on data (PIA and DPIA) to those more centred on societal and ethical issues (SIA and EtIA). The relationship between these different models can be thought of as a series of concentric rings,[41] where HRESIA is intermediate between the other two.

2.1.1. *HRESIA in the context of data processing impact assessments*

The focus on the risks arising from data processing has been an essential element of data protection regulation from the outset, though over the years this risk has evolved in

"impact that concerns or affects human rights and responsibilities, benefits and harms, justice and fairness, well-being and the social good". See also the Privacy, Ethical and Social Impact Assessment (PESIA) proposed in the context of the H2020 Virt-EU project <https://virteuproject.eu> accessed 27 April 2018.

[41] See Raab and Wright (n 21) 376–382.

a variety of ways.[42] The original concern about government surveillance[43] has been joined by new concerns regarding the economic exploitation of personal information (risk of unfair or unauthorised uses of personal information[44]) and, nowadays, by the increasing number of decision-making processes based on information (risk of discrimination, large-scale social surveillance, and bias in predictive analyses[45]).

From a theoretical perspective, this focus on the potential adverse effects of data use has not been an explicit element of data protection laws. Many of their provisions adopt a procedural approach that leaves in the shadows the safeguarded interests, which are encapsulated in the broad and general notion of data protection.

Moreover, compared to other personality rights, such as right to image or name, data protection has a proteiform nature, since data may consist in name, numbers, behavioural information, genetic data or many other forms of information. The progressive datafication of our world makes it difficult to find something that is not or cannot be transformed into data. The consequent broad notion of data protection covers different fields and has partially absorbed some elements traditionally protected by other personality rights.[46]

Against this background, the idea of control over information was used to aggregate the different forms of data protection and to find a common core.[47] The procedural approach is consistent with this idea, since it secures all the stages of data processing, from data collection to communication of data to third parties. Nevertheless, control over information describes the nature of the power, which the law grants to the data subject, not its theoretical foundations.

In this regard, part of the legal doctrine has pointed out the role of human dignity as a foundational ground of data protection in Europe.[48] However, interplay with the non-discrimination principle,[49] the role of data protection in the public sphere and in digital citizenship[50] suggest that a broader range of values underpin data protection.

Although more recently data protection regulations[51] and practices[52] have adopted a more explicit risk-based approach to address the varying challenges of data use, they still focus on the procedural aspects. Data management procedures represent, therefore, a form of risk management based on the regulation of the different stages of data processing (collection, analysis and communication) and the definition of the powers and tasks of the various subjects involved in this process.

This procedural approach and the focus of risk assessment on data management have led data protection authorities to propose assessment models (PIA) primarily centred on data quality and data security, leaving aside the nature of safeguarded interests. Instead, these interests are taken into account by DPAs and courts in their decisions, but – since data protection laws provide limited explicit references to the safeguarded values, rights and freedoms – the analysis of the relevant interest is frequently curtailed or not adequately elaborated.[53]

[42] See Lee A. Bygrave, *Data Protection Law. Approaching Its Rationale, Logic and Limits* (Kluwer Law International 2002), 107–112; Viktor Mayer-Schönberger, 'Generational development of data protection in Europe?' in Philip E. Agre and Marc Rotenberg (eds), Technology and privacy: The new landscape (MIT Press 1997) 221–225; Colin J. Bennett, *Regulating Privacy: Data Protection and Public Policy in Europe and the United States* (Cornell University Press 1992) 29–33, 47. See also Arthur R. Miller, *The Assault on Privacy Computers, Data Banks, Dossiers* (University of Michigan Press 1971) 54–67; Myron Brenton, *The Privacy Invaders* (Coward-McCann 1964); Vance Packard, *The Naked Society* (David McKay 1964); Secretary's Advisory Committee on Automated Personal Data Systems, 'Records, Computers and the Rights of Citizens' (1973) <http://epic.org/privacy/hew1973report/> accessed 27 September 2016.

[43] See Alan F. Westin, Privacy and Freedom (Atheneum 1970).

[44] See also Alessandro Acquisti, Laura Brandimarte and George Loewenstein, 'Privacy and human behavior in the age of information' (2015) 347(6221) Science 509–514; Laura Brandimarte, Alessandro Acquisti, and George Loewenstein, 'Misplaced Confidences: Privacy and the Control Paradox' (2010), Ninth Annual Workshop on the Economics of Information Security <http://www.heinz.cmu.edu/~acquisti/papers/acquisti-SPPS.pdf> accessed 27 February 2017; Joseph Turow and other, 'The Federal Trade Commission and Consumer Privacy in the Coming Decade' (2007) 3 ISJLP 723–749 <http://scholarship.law.berkeley.edu/facpubs/935> accessed 27 February 2017; Daniel J. Solove, 'Introduction: Privacy Self-management and The Consent Dilemma' (2013) 126 Harv. L. Rev. 1880, 1883–1888.

[45] See, *inter alia*, Andrew D. Selbst, 'Disparate Impact in Big Data Policing' (2018) 52(1) Georgia Law Review 109-195; Mireille Hildebrandt, *Smart Technologies and the End(s) of Law : Novel Entanglements of Law and Technology* (Edward Elgar Publishing, 2016) 191–195; Solon Barocas and Andrew D. Selbsr, 'Big Data's Disparate Impact' (2016) 104 (3) California Law Review 671–732; Mantelero, 'Personal data for decisional purposes' (n 7).

[46] See also bart van der Sloot, 'Privacy as Personality Right: Why the ECtHR's Focus on Ulterior Interests Might Prove Indispensable in the Age of "Big Data"' (2015) 31(80) Utrecht Journal of International and European Law 25–50 ("the right to privacy has been used by the Court to provide protection to a number of matters which fall primarily under the realm of other rights and freedoms contained in the Convention").

[47] See also Daniel J. Solove, Understanding Privacy (Harvard University Press, 2008) 12–38; Westin (n 43) 330–399.

[48] See James Q. Whitman, 'The Two Western Cultures of Privacy: Dignity versus Liberty' (2004) 113 The Yale Law Journal 1151–1221.

[49] See, in this sense, the notion of special categories of data in art. 6 of the Convention 108 adopted by the Council of Europe and in art. 9 of the GDPR. The White House, 'Administration Discussion Draft: Consumer Privacy Bill of Rights Act 2015' <https://obamawhitehouse.archives.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf> accessed 25 June 2017. See also The White House, 'A Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy' (2012), Appendix A: The Consumer Privacy Bill of Rights <https://obamawhitehouse.archives.gov/sites/default/files/privacy-final.pdf> accessed 4 December 2017.

[50] See Stefano Rodotà, 'Privacy, Freedom, and Dignity: Conclusive Remarks at the 26th International Conference on Privacy and Personal Data Protection' (2004) <http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/export/1049293> accessed 16 December 2017.

[51] See articles 24 and 35, GDPR.

[52] See Wright and De Hert (n 21).

[53] See, e.g., the following decisions: Garante per la protezione dei dati personali (Italian DPA), 1 February 2018, doc. web n. 8159221; Garante per la protezione dei dati personali, 8 September 2016, n. 350, doc. web 5497522; Garante per la protezione dei dati personali, 4 June 2015, n. 345, doc. web n. 4211000; Garante per la protezione

Data protection authorities and courts prefer using arguments grounded on the set of criteria provided by data protection regulations.[54] The legitimate nature of the purposes, lawfulness and fairness of processing, transparency, purpose limitation, data minimisation, accuracy, storage limitation, data integrity and confidentiality are general principles frequently used by data protection authorities in their argumentations.[55] However, these principles are only an indirect expression of the safeguarded interests. Most of them are general clauses that may be interpreted more or less broadly and require an implicit consideration of the interests underpinning data use.

Moreover, the indefinite nature of these clauses has frequently led to the adoption of the criterion of proportionality,[56] which is a sort of synthesis of the evaluation of the different competing interests[57] by courts or the DPAs. In fact, this balancing of interests and the reasoning that resulted in a specific border between them is often implicit in the notion of proportionality and not discussed in the decisions taken by the DPAs or discussed in an axiomatic manner.[58]

Against this scenario, it is difficult for data controllers to understand and outline the set of values that they should take into account in developing their data-intensive devices and services, since these values and their mutual interaction remain unclear and undeclared. Nor is this difficulty solved by the adoption of PIAs, since these assessment models merely point out the need to consider aspects other than data quality and data security, without specifically elaborating them and providing effective tools to identify and operationalise broader social values.

In the same way, the recent EU DPIA – according to the first models proposed by DPAs – does not offer a better answer. Despite specific references in the GDPR to the safeguarding of rights and freedoms in general as well as to societal issues,[59] the new assessment models do not seem to increase the focus on societal consequences that is present in the existing PIAs.[60]

In this light, the main goal of the HRESIA model is to fill this gap, providing an assessment model focused on the rights and freedoms[61] that may be affected by data processing. With regard to the EU context, this is in line with the declared intent of the GDPR and may provide a valuable tool on carrying out the risk assessment outlined in the Regulation, which focuses on the "risk to the rights and freedoms of natural persons".[62]

2.1.2. *The HRESIA and the collective dimension of data protection*

Shifting the focus from the traditional sphere of data quality and security to fundamental rights and freedoms, the HRESIA model help data controllers to address the collective dimension of data processing. In this sense, the issues concerning data-intensive applications and their use in decision-making processes concern a variety of interests related to several fundamental rights and freedoms. Not only does the risk of discrimination represent one of the biggest challenges of these applications, but other rights and freedoms also assume relevance, such as the right to the integrity of the person, to education, to be equal before the law, and the freedom of movement, of thought, of expression, of assembly and freedom in the workplace.[63]

Against this scenario, the last question that the proposed model must address from a theoretical standpoint concerns

dei dati personali, 8 May 2013, n. 230, doc. web n. 2433401; Agencia Española de Protección de Datos (Spanish DPA), Expediente n. 01769/2017; Agencia Española de Protección de Datos, Expediente n. 01760/2017; Agencia Española de Protección de Datos, Resolución R/01208/2014; Agencia Española de Protección de Datos, (Gabinet Juridico) Informe 0392/2011; Agencia Española de Protección de Datos, (Gabinet Juridico) Informe 368/2006; Commission de la protection de la vie privée (Belgian DPA), 15 December 2010, raccomandation n. 05/2010; Commission Nationale de l'Informatique et des Libertés (French DPA), 17 July 2014, deliberation n. 2014-307; Commission Nationale de l'Informatique et des Libertés, 21 June 1994, deliberation n. 94-056.

[54] Regarding the focus of DPAs' decisions on national data protection regulations and their provisions, see also the results of the empirical analysis carried out by Maria Grazia Porcedda, 'Use of the Charter of Fundamental Rights by National Data Protection Authorities and the EDPS' (Centre for Judicial Cooperation(CJC) Robert Schuman Centre for Advanced Studies, European University Institute, 2017) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3157786> accessed 24 April 2018.

[55] See above fn 53.

[56] See De Hert (n 27), 46, who defines the application of the principle of proportionality as a "political" test. With regard to the jurisprudence of the European Court of Human Rights, this author also points out how "The golden trick for Strasbourg is to see almost every privacy relevant element as one that has to do with the required legal basis".

[57] See also Sébastien van Drooghenbroeck, *La proportionnalité dans le droit de la Convention européenne des droits de l'homme: prendre l'idée simple au sérieux* (Publications Fac St Louis, 2001) 302.

[58] See e.g. Court of Justice of the European Union, 13 May 2014, Case C-131/12, Google Spain SL, Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González, para 81 ("In the light of the potential seriousness of that interference, *it is clear that* it cannot be justified by merely the economic interest which the operator of such an engine has in that processing", emphasis added).

[59] See Recital n. 75.

[60] For a proposal of integration of PIA and EtIA, see Wright and Friedewald (n 21) 760–762. However, these authors do not adopt a broader viewpoint focused on human rights assessment.

[61] Despite this difference, HRESIA and PIA/DPIA take a common approach in terms of architecture, since both are rights-based assessments. From this standpoint, the HRESIA interpretation of the general clauses and principles in the data protection regulations with respect to broader human rights becomes the core element of the rights-based approach. See also The Danish Institute for Human Rights (n 27) 76 ("Human rights impacts cannot be subject to 'offsetting' in the same way that, for example, environmental impacts can be. For example, a carbon offset is a reduction in emissions of carbon dioxide made in order to compensate for or to offset an emission made elsewhere. With human rights impacts on the other hand, due to the fact that human rights are indivisible and interrelated, it is not considered appropriate to offset one human rights impact with a 'positive contribution' elsewhere").

[62] See Article 35, GDPR on risk assessment. The same reference to the rights and freedoms is also present in several other provisions of the GDPR.

[63] See also Council of Europe-Committee of experts on internet intermediaries (MSI-NET) (n 29) and EDPS – Ethics Advisory Group (n 12). See also van der Sloot, B. (2015). Privacy as Personality Right: Why the ECtHR's Focus on Ulterior Interests Might Prove Indispensable in the Age of "Big Data". *Utrecht Journal of International and European Law*, *31*(80), 25–50.

the compatibility of the collective dimension of data protection[64] and the way human rights are framed by legal scholars. To answer to this question, it is necessary to highlight how the notion of collective data protection tried to go beyond the individual dimension of data protection and its focus on data quality and security, suggesting a broader range of safeguarded interests and affecting individuals as a group.

An impact assessment focused on the broader category of human rights, which also takes into account the ethical and societal issues concerning data use, can provide an answer to this need. This broader perspective and the varied range of human rights makes it possible to consider the impacts of data use more fully, not only limited to data protection. Moreover, several principles, rights and freedoms in the charters of human rights directly or indirectly address groups or collective issues.

However, in the context of human rights[65] as well as data protection, legal doctrine and the regulatory framework are focused primarily on the individual dimension. Furthermore, in some cases, the theoretical human rights background provides a limited notion of these rights and freedoms, which is inadequate to handle the new challenges of technology.[66]

In this sense, for example, the approach to classification adopted by modern algorithms does not merely focus on individuals and on the categories traditionally used to support unjust or prejudicial treatment of different groups of people.[67] On the contrary, algorithms create groups or clusters of people based on different and more varied characteristics (e.g. customer habits, lifestyle, online and offline behaviour, network of personal relationships etc.). For this reason, the wide application of predictive technologies based on these new categories and their use in decision-making processes suggests a broader notion[68] of discrimination.[69]

Additionally, the nature of the groups created by data-intensive applications poses challenging issue from the procedural viewpoint, which concern the potential remedies to the need for collective representation in the context of algorithmic-created groups.[70] Indeed, people belonging to groups that are the traditional targets of discriminatory practices are aware of their membership of these groups and they

[64] See above fn. 21.

[65] On the limits of an approach focused on induvial rather than on the collective dimension, see Walker (n 27), 21 ("Combatting discrimination is not simply a matter of prohibiting acts of discrimination or discriminatory legislation, but also entails an obligation on the State to take action to reverse the underlying biases in society that have led to discrimination and, where appropriate, take temporary special measures in favour of people living in disadvantaged situations so as to promote substantive equality"). See also Eric J. Mitnick, *Rights, Groups, and Self-Invention : Group-Differentiated Rights in Liberal Theory* (Routledge, 2018); Robert P. George, 'Individual rights, collective interests, public law, and American politics' (1989) 8 Law and Philosophy 245–261.

[66] For example, based on previous experiences, discrimination is primarily considered within the traditional categories (sex, religion, etc.), see e.g. Recital 71 of the GDPR on automated decision-making, which refers to "discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation". However, groups shaped by analytics and AI differ from the traditional notion of groups in the sociological sense of the term considered by the legislation: they have a variable geometry and individuals can shift from one group to another.

[67] These categories, used in discriminatory practice, are to a large extent the special categories mentioned in the data protection regulations.

[68] See also Article 14, European Convention on Human Rights, which contains an open-ended list of potential fields in which discriminatory practices can be adopted.

[69] This notion must encompass both the prejudicial treatment of groups of people – regardless of whether they belong to special categories –, and the consequences of unintentional bias in the design, data collection and decision-making stages of Big Data applications. Indeed, these consequences may negatively impact on individuals and society, even though they do not concern forms of discrimination based on racial or ethnic origin, political opinions, religious or philosophical beliefs or other elements that traditionally characterise minorities or vulnerable groups. For example, Kate Crawford has described the case of the City of Boston and its StreetBump smartphone app to passively detect potholes. The application had a signal problem, due to the bias generated by the low penetration of smartphones among lower income and older residents. While the Boston administration took this bias into account and solved the problem, less enlightened public officials might underestimate such considerations and make potentially discriminatory decisions See Kate Crawford, 'The Hidden Biases in Big Data' (2013) Harv. Bus. Rev. April 1, 2013, <https://hbr.org/2013/04/the-hidden-biases-in-big-data> accessed 29 January 2018; Jonas Lerman, 'Big Data and Its Exclusions' (2013) 66 Stan. L. Rev. Online 55. Another example is the Progressive case, in which an insurance company obliged drivers to install a small monitoring device in their cars in order to receive the company's best rates. The system considered as a negative factor driving late at night but did not take into account the potential bias against low-income individuals, who are more likely to work night shifts, compared with late-night party-goers, "forcing them [low-income individuals] to carry more of the cost of intoxicated and other irresponsible driving that happens disproportionately at night", see David Robinson, Harlan Yu and Aaron Rieke, 'Civil Rights, Big Data, and Our Algorithmic Future. A September 2014 report on social justice and technology' (2014), 18–19 <http://bigdata.fairness.io/wp-content/uploads/2014/09/Civil_Rights_Big_Data_and_Our_Algorithmic-Future_2014-09-12.pdf> accessed 10 March 2018. Finally, commercial practices may lead to price discrimination or the adoption of differential terms and conditions depending on the assignment of consumers to a specific cluster. Thus, consumers classified as "financially challenged" belong to a cluster "[i]n the prime working years of their lives […] including many single parents, struggl[ing] with some of the lowest incomes and little accumulation of wealth". This implies the following predictive viewpoint, based on big data analytics and regarding all consumers in the cluster: "[n]ot particularly loyal to any one financial institution, [and] they feel uncomfortable borrowing money and believe they are better off having what they want today as they never know what tomorrow will bring", see Federal Trade Commission. 2014. Data Brokers: A Call for Transparency and Accountability. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf (accessed February 27, 2018), 20. It is not hard to imagine the potential discriminatory consequences of similar classifications with regard to individuals and groups.

[70] See also Mantelero, 'Regulating Big Data (n 7).

know or may know the other members of the group. On the contrary, in the groups generated by algorithms, people do not know the other members of the group and, in many cases, are not aware of the consequences of their belonging to a group. Data subjects are not aware of the identity of the other members of the group, have no relationship with them and have a limited perception of their collective issues.

Hard law remedies in this field may be not easy to achieve in the short run and the existing or potential procedural rules often vary from one legal context to another.[71] In this scenario, a voluntary assessment procedure may represent a valid alternative to address these challenges. For these reasons, a model based on a participatory approach and in which human rights are seen through the lens of ethical and social values may provide broader safeguards both in terms of the interests taken into account and the categories of individuals engaged in the process.

Providing a framework for a societal impact assessment of data-intensive applications is in line with the ongoing debate on Responsible Research Innovation[72] and the demands of the data industry and product developers for practical tools to help them address the social issues of data use. Tools that can be more flexible open to new emerging values, easily reshaped and applicable in different legal and cultural contexts when built into self-assessment models.

2.1.3. *Human rights impact assessment in data processing*

The Human Rights Impact Assessment (HRIA) adopted in business[73] is a third-party assessment based on data collection and interviews with management, stakeholders and experts, which may take several months to carry out. This assessment is not focused on a specific process, but the bulk of the activities[74] carried out by a company in one or more countries.[75] On the other hand, the PIA and DPIA models concern a given data processing operation or, at least, may address a set of similar operations that present similar risks.[76]

The different scale of HRIA and PIA/DPIA does not rule out adoption of the HRIA approach – in terms of values and participatory model – in data protection and for the assessment of single processing operations. On the other hand, the focus on a specific process tends to scale down HRIA complexity.[77] Here, the data intensive and (third party) expert-based model adopted by HRIA is replaced by a self-assessment tool, which is grounded on the same principles and rights as the HRIA but is primarily aimed at data controllers. These may perform this assessment autonomously or with the support of an *ad hoc* committee,[78] which may be a permanent body supporting different assessment operations and their reviews.

Taking this approach, the participative method used in HRIA is also part of the HRESIA model in the sense that the self-assessment tool (questionnaire), or the *ad hoc* committee, help the data controller identify potential stakeholders and involve them in a participatory process. The proposed model therefore scales down the HRIA framework. The broader reach of the HRIA means that stakeholders' engagement cannot be seen as a mere opportunity, while – in the case of HRESIA – a single data processing operation may have a limited impact and this kind of third-party engagement may be superfluous.

Moreover, the wide array of business operations scrutinised in the HRIA is more likely to have an impact on a variety of human rights, whereas single data processing activities affect a more limited range of rights. In this sense, – as confirmed by ongoing studies in this field[79] – it is possible to point out the major role played by the principle of non-discrimination[80] (Article 2, UN Universal Declaration of Human Rights) and the right to privacy and private life (Article 12, UDHR). However, freedom of movement[81] (Article 13, UDHR),

[71] See, e.g., the case of redress procedures to safeguard consumers' rights.

[72] See Jack Stilgoe, Richard Owen, and Phil Macnaghten. 'Developing a Framework for Responsible Innovation' (2013) 42(9) Research Policy 1568–1580.

[73] See e.g. United Nations, *Guiding Principles on Business and Human Rights* (United Nations, 2011) <http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf> accessed 27 April 2018.

[74] See e.g. LKL International Consulting Inc., 'Human Rights Impact Assessment of the Bisha Mine in Eritrea 2015 Audit' (2015) <https://nevsuncsr.com/wp-content/uploads/2015/07/Bisha-HRIA-Audit-2015.pdf> accessed 27 April 2018; Tulika Bansal and Yann Wyss, *Talking the Human Rights Walk: Nestlé's Experience Assessing Human Rights Impacts in its Business Activities* (Danish Institute for Human Rights and Nestlé, 2013) <http://www.nestle.com/asset-library/documents/library/documents/corporate_social_responsibility/nestle-hria-white-paper.pdf> accessed accessed 27 April 2018; On Common Ground Consultants Inc., 'Human Rights Assessment of Goldcorp's Marlin Mine' (2010) <http://csr.goldcorp.com/2011/docs/2010_human_full_en.pdf> accessed 27 April 2018.

[75] See e.g. the short description of the HRIAs by different companies in The Danish Institute for Human Rights. (n 27) 12–15.

[76] See Article35.1, GDPR. See also Article 29 Data Protection Working Party (n 11) 7.

[77] See The Danish Institute for Human Rights (n 27) 39–124.

[78] See below Section 3.1.

[79] See EDPS - Ethics Advisory Group (n 12) and Council of Europe-Committee of experts on internet intermediaries (MSI-NET) (n 29). See also The White House, Executive Office of the President, 'Big Data: Seizing Opportunities, Preserving Values' (2014) <https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf> accessed 11 November 2017.

[80] See, e.g., Garante per la protezione dei dati personali (Italian DPA), 11 January 2007, doc. web n. 1381620; Commission de la protection de la vie privée, raccomandation (Belgian DPA), 18 March 2009, n. 01/2009. See also Information Commissioner's Office, 'The employment practices code' (2011) and 'The employment Practices. Code Supplementary Guidance' (2005) <https://ico.org.uk/for-organisations/guide-to-data-protection/employment/> accessed 16 April 2018.

[81] See, e.g., Commission Nationale de l'Informatique et des Libertés, 21 June 1994, deliberation n. 94-056; Commission Nationale de l'Informatique ed des Libertés, 23 November 2013, deliberation n. 2013-366; Commission Nationale de l'Informatique ed des Libertés, 16 March 2006, deliberation n. 2006-066 ; Garante per la protezione dei dati personali, 8 September 2016, doc. web n. 5497522; Garante per la protezione dei dati personali, 18 May 2016, doc. web n. 5217175; Garante per la protezione dei dati personali, 7 March 2013, doc. web n. 2471134; Garante per la protezione dei dati personali, 24 May 2017, doc. web n. 6495708; Garante per la protezione dei dati personali, 15 June 2017, doc. web n. 6697925; Garante per la protezione dei dati personali, 8 January 2015, doc web n. 3723437); Commission de la protection de la vie privée, raccomandation n. 03/2013, 24 April 2013; Commission de la protec-

freedom of thought (Article 18 UDHR), as well as freedom of expression[82] (Article 19 UDHR) and the right to education[83] (Article 26 UDHR) are also relevant in assessing the impact of data use in different contexts.

Finally, it should be pointed out that focusing the risk assessment on human rights allows for a universal model, which is unaffected, in its core values, by the variation in approaches to data protection in different geographical areas. At the same time, as described below, this universal approach should not underestimate the local dimension of the social issues[84] and the varying nuances in the safeguards to fundamental rights and freedoms in different contexts, including the balancing of competing interests.

2.1.4. *From HRIA to HRESIA*

Shifting the focus from data protection alone to human rights represents an important step in addressing the complexity of today's data-intensive and AI applications. However, the human rights-based approach does have its limitations, due to its historical origin and theoretical framework.

As mentioned above,[85] the conceptualization of these rights is based on past experience and threats to human dignity and freedom, as well as individual equality. For example, the principle of non-discrimination mainly focused on the conditions that are traditionally crucial in discriminatory practices (e.g. race, colour, sex, language, religion, political opinions), while the current algorithmic discrimination is based on blurrier, less clear-cut categories. This may make a HRIA based on this traditional notion of human rights less effective.

Moreover, human rights are largely safeguarded as individual rights, while Big Data and AI are often no longer primarily interested in the individual dimension and focus on groups and the collective level.[86] For this reason, it is necessary to address the societal consequences of data-intensive applications, such as predictive policing or healthcare analytics.

Additionally, data-intensive application may not necessarily be against the law and may pass a human rights assessment. However, this does not rule out that they may raise ethical and societal concerns, for example in terms of unforeseen bias or social acceptability (e.g. invasive or massive use of biometric data based on data subjects' consent), which cannot be left unaddressed.

These limitations concerning the safeguarding of the collective dimension of data protection lead us to consider societal and ethical issues in the HRIA. Moreover, ethical and social issues are not disconnected from legal assessment in this field. Data protection laws adopt general principles (e.g. fairness or proportionality) and general clauses (e.g. necessity, legitimacy[87]) which are used to introduce non-legal social values into the legal framework. Similarly, legal scholars have highlighted how the application of human rights is necessarily affected by social and political influences that are not explicitly formalised in court decisions.[88]

tion de la vie privée, raccomandation n. 05/2010 del 15 December 2010; Commission de la protection de la vie privée, raccomandation n. 01/2010 del 17 March 2010 ; Agencia Española de Protección de Datos, Resolución R/01208/2014; Agencia Española de Protección de Datos expediente n. E/02778/2010; Agencia Española de Protección de Datos, Expediente n. E/02689/2012. See also Information Commissioner's Office, 'In the picture: A data protection code of practice for surveillance cameras and personal information' (2017) <https://ico.org.uk/media/1542/cctv-code-of-practice.pdf> accessed 13 April 2018.

[82] See, e.g., Garante per la protezione dei dati personali, 8 maggio 2013, n. 230, doc. web n. 2433401; Commission de la protection de la vie privée, 12 April 2006, avis, n. 8/2006; Agencia Española de Protección de Datos, Gabinete Juridico, Informe 0464/2013; Agencia Española de Protección de Datos, Gabinete Juridico, Informe 0292/2010.

[83] See, e.g., Garante per la protezione dei dati personali, 8 May 2013, doc. web n. 2433401; Commission de la protection de la vie privée, 12 April 2006, avis, n. 8/2006.

[84] On this tension between universalist and context-dependent approach in social impact assessment, see Antonio Aledo-Tur and Andrés J. Domínguez-Gómez, 'Social Impact Assessment (SIA) from a Multidimensional Paradigmatic Perspective: Challenges and Opportunities' (2017) 195 Journal of Environmental Management 56–61; Deanna Kemp and Frank Vanclay, 'Human rights and impact assessment: clarifying the connections in practice' (2013) 31(2) Impact Assessment and Project Appraisal 86, 92. See also, with regard to Big Data and AI, Council of Europe (n 9), Section IV, para 1.2 ("Personal data processing should not be in conflict with the ethical values commonly accepted in the relevant community or communities and should not prejudice societal interests, values and norms, including the protection of human rights"); The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (n 23) 24.

[85] See above para 2.1.3.

[86] See above fn. 21.

[87] See Bygrave (n 42) 61–63 and 339 on processing data for legitimate purpose ("solid grounds exist for arguing that the notion of 'legitimate' denotes a criterion of social acceptability, such that personal data should only be processed for purposes that do not run counter to predominant social mores […] The bulk of data protection instruments comprehend legitimacy *prima facie* in terms of procedural norms hinging on a criterion of lawfulness […] Very few expressly operate with a broader criterion of social justification. Nevertheless, the discretionary powers given by some national laws to national data protection authorities have enabled the latter to apply a relatively wide-ranging test of social justification"). See also New South Wales Privacy Committee, 'Guidelines for the operations of personal data systems' (1977) <http://www.rogerclarke.com/DV/NSWPCGs.pdf> accessed 13 April 2018; Michael Kirby, 'Transborder Data Flows and the 'Basic Rules' of Data Privacy' (1981) 16 Stanford J. of Int. Law, 27–66.

[88] See De Hert (n 27); Nardell, Gordon, 'Levelling Up: Data Privacy and the European Court of Human Rights' in Serge Gutwirth, Yves Poullet and Paul De Hert (eds) *Data Protection in a Profiled World* (Springer, 2010) 43–52; Yutaka Arai-Takahashi and Yutaka Arai, *The Margin of Appreciation Doctrine and the Principle of Proportionality in the Jurisprudence of the ECHR* (Intersentia, 2002); van Drooghenbroeck. (n 57); Evans, Carolyn, and Simon Evans, 'Evaluating the Human Rights Performance of Legislatures' (2006) 6(3) Human Rights Law Review 545–570; Centre for European Policy Studies, 'Global Data Transfers: The Human Rights Implications' (2010) <https://www.ceps.eu/publications/global-data-transfers-human-rights-implications> accessed 13 November 2017; Steven Greer, *The margin of appreciation: interpretation and discretion under the European Convention on Human Rights* (Editions du Conseil de l'Europe, 2000); David Harris and others, *Law of the European Convention on Human Rights* (Oxford University Press, 2014); Paul De Hert, 'Balancing security and liberty within the European human rights framework. A critical reading of the Court's case law in the light of surveillance and criminal law en-

From this perspective, a HRESIA may be used to unveil the existing interplay between the legal and the societal dimensions,[89] making it explicit and mitigating the limitations of the HRIA approach. It is important to reveal this cross-fertilization between law and society, without allowing it to slip between the lines of the decisions of the courts, DPAs or other bodies. In this sense, providing a model that also considers the social and ethical dimensions helps to democratise assessment procedures, removing them from the exclusive hands of the courts, mediated by legal formalities.

Indeed, although the courts, DPAs and legal scholars are aware of the influence of societal issues on their reasoning, it is frequently not explicit in the decisions they adopt in data protection. Product developers are therefore unable to grasp the real sense of the existing provisions. Stressing the societal values that should be taken into account in the human rights assessment helps developers carry out self- assessments of the potential and complex negative consequences, from the early stages of product design.

Some may argue that one potential shortcoming of the proposed approach concerns the fact that, in the end, it introduces a paternalistic approach to data protection. In this sense, a HRESIA model necessarily encourages data controllers to exclude certain processing operations due to their ethical or social implications, even if some individual data subjects may take a different view and consider them in line with their own values. The model may therefore be seen as a limitation on self-determination, indirectly affecting and reducing the range of available data use.

The main pillar of this argument concerns the data subject's self-determination, but this notion is largely undermined by today's Big Data and AI-driven data processing.[90] The lack of knowledge and awareness in making decisions with regard to data processing, on the one hand, and the frequent lack of effective freedom of choice (due to social, economic and technical lock-ins), on the other, argue for a slightly paternalistic approach as a way to compensate these limitations on individual self-determination.[91]

Moreover, HRESIA is not a standard but a self-assessment tool. It aims to provide data controllers with a better awareness of the human rights, ethical and social implications of data processing that they should address. Data controllers remain free to decide whether and how to address them and to put in place participatory models to give voice to data subjects.

Finally, the publicity surrounding the HRESIA (in line with the HRIA) may help to reinforce data subjects' self-determination, as it makes explicit the implications of a certain data processing operation and fosters users' informed choice. Publicity increases not only the data subject's awareness, but also the data controller's accountability and is consistent with a human rights approach.[92]

There are cases in which full disclosure of the assessment results may be limited by the legitimate interests of the data controller, such as confidentiality of information, security and competition. For example, the Guidelines on Big Data adopted by the Council of Europe in 2017[93] – following the suggestions of legal scholars[94] – specify that the results of the assessment proposed in the guidelines "should be made publicly available, without prejudice to secrecy safeguarded by law. In the presence of such secrecy, controllers provide any confidential information in a separate annex to the assessment report. This annex shall not be public but may be accessed by the supervisory authorities".[95]

The HRESIA and the SIA (Social Impact Assessment) have a similar focus on societal issues and the collective dimension.[96] They also share an interest in public participation, individual and group empowerment through the assessment process, non-discrimination, equal participation in the assessment, focus on a range of different issues, accountability and a circular architecture. Since the model proposed sets out to embed social and ethical issues in the HRIA, it is worth pointing out the differences between the HRESIA and the EtIA/SIA models.

The main differences concern their rationale, the extent of the assessment and the way that the different interests are balanced in the assessment. The HRESIA aims to provide a universal tool, which, at the same time, takes into account the local dimension of the safeguarded interests. In this sense, it is based on a common architecture grounded on intentional instruments with normative strength (charters of fundamental rights). The core of the architecture is represented by human

forcement strategies after 9/11' (2005) 1(1) Utrecht Law Review 68–96.

[89] HRIA has its roots in the Social Impact Assessment (SIA) models. See Walker (n 27), 5. Nevertheless, due to the existing interplay between human rights and social and ethical values, it is hard to define this relationship as derivation, as human rights notions necessarily affected the values adopted in SIA models. For example, the International Association for Impact Assessment Principles refers to Article 1 of the UN Declaration on the Right to Development by which every human being and all peoples are entitled to participate in, contribute to, and enjoy economic, social, cultural and political development.

[90] See Mantelero 'The future of consumer data protection in the E.U.' (n 7).

[91] See also Bygrave (n 42) 86 ("Under many European data protection regimes, paternalistic forms of control have traditionally predominated over participatory forms, though implementation of the EC Directive changes this weighting somewhat in favour of the latter").

[92] Access to information is both a human right per se and a key process principle of HRIA.

[93] See above fn. 9.

[94] This is a critical aspect, because of the need to balance the transparency of data processing with security and firms' competitiveness. It is possible to provide business-sensitive information in a separate annex to the impact assessment report, which is not publicly available, or publish a short version of the report without the sensitive content. See Alessandro Mantelero, 'Competitive value of data protection: the impact of data protection regulation on online behaviour' (2013) 3(4) Int'l Data Privacy L. 234; Neil M. Richards and Jonathan H. King, 'Three Paradoxes of Big Data' (2013) 66 Stan. L. Rev. Online 41, 43; Wright (n 8) 222.

[95] Council of Europe (n 9), Section IV, para 3.3. See also Selbst (n 45) 190.

[96] See Frank Vanclay and others, *Social Impact Assessment: Guidance for assessing and managing the social impacts of projects* (International Association for Impact Assessment, 2015) <http://www.iaia.org/uploads/pdf/SIA_Guidance_Document_IAIA.pdf> accessed 26 April 2018; Walker (n 27), 39–42.

rights, which also play a role in the SIA models but are not pivotal as they take a wider approach.[97]

In fact, the greater extension of the SIA approach encompasses a wide range of issues,[98] broad theoretical categories and focuses on the specific context investigated.[99] The solutions proposed by the SIA models are therefore heterogeneous and differ in different contexts,[100] making it difficult to place them within a unique framework, which – on the contrary – is an essential requirement in the context of the global policies on data use.

By contrast, a model grounded on human rights[101] is more closely defined and universally applicable. Moreover, the scaled down human rights impact assessment proposed offers a common standard without requiring the significant effort of the SIA, which was designed for large-scale social phenomena. Finally, HRESIA is necessarily a rights-based assessment, in line with the approach adopted in data protection (PIA, DPIA), while both the SIA and the EtIA (Ethical Impact Assessment) are risks/benefits models.

Regarding the comparison between HRESIA and EtIA,[102] the considerations about SIA can be repeated in relation to EtIA.[103] Moreover, in the EtIA model, in the forms proposed in the context of data protection, there is a clearer link with the principles already recognised in law,[104] given the relationship between ethics and law discussed above. However, the potential risk of a mere ethical assessment does create some overlap between ethical guidance and legal provisions.

Finally, EtIA relies on a list of quite broad principles.[105] This is consistent with the assessment of specific processes (e.g. ethics committees involved in clinical trials) or, alternatively, entire branches of technology. But, data controllers may find it difficult to apply broad principles in practice faced with the enormous number of different data-intensive projects developed each year by companies.[106]

2.2. *Advantages of the HRESIA approach and the limits of PIA/DPIA models*

The proposed assessment model can achieve positive results in assessing the impact of data use for the various reasons mentioned above, here briefly summarised:

- The central role of human rights in HRESIA provides a universal set of values, which is the backbone of the model making it suited to various legal and social contexts.
- The HRESIA is necessarily a principle-based model, which makes it better at dealing with the rapid change of technological development, less easily addressed by detailed sets of provisions.
- The proposed model follows in the footsteps of the data protection assessments, as a rights-based assessment in line with the PIA and DPIA approaches. In this sense, the HRESIA can be classed an 'integrated' model since it integrates human rights into DPIA and social assessment. However, the HRESIA is not a multilayer assessment, but rather a human rights impact assessment in which the ethical and social dimensions serve to better understand

[97] See, *ex multis*, Thomas Dietz, 'Theory and method in social impact assessment' (1987) 57(1) Sociol. Inq. 54–69; Nicholas C. Taylor, C. Hobson Bryan and Colin G. Goodrich, *Social assessment: theory, process and techniques* (Centre for Resource Management, Lincoln College, 1990); Henk A. Becker, 'Social impact assessment' (2001) 128(2) Eur. J. Oper. Res. 311–321; Frank Vanclay, 'Conceptualising social impacts' (2002) 22(3) Environ. Impact. Assess. 183–211; Henk A. Becker and Frank Vanclay (eds) *The International Handbook of Social Impact Assessment. Conceptual and Methodological Advances* (Edward Elgar, 2003); James Harrison, 'Human rights measurement: Reflections on the current practice and future potential of human rights impact assessment' (2011) 3(2) J Hum Rights Prac. 162–187.

[98] See Burdge and Vanclay (n 28) 59 ("Social impacts include all social and cultural consequences to human populations of any public or private actions that alter the ways in which people live, work, play, relate to one another, organize to meet their needs, and generally cope as members of society"). See also Massarani, Drakos, and Pajkowska (n 28).

[99] In this sense, the ethical and social impact assessment is described as the outermost circle to which the PIA can be extended by Raab and Wright (n 21) 379–382.

[100] See also Jonas Svensson, Social impact assessment in Finland, Norway and Sweden: a descriptive and comparative study (Thesis, KTH Royal Institute of Technology 2011), 84 <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-86850> accessed 27 April 2018.

[101] See Kemp and Vanclay (n 84) 90–91 ("Human rights impact assessment (HRIA) differs from SIA in the sense that it proceeds from a clear starting point of the internationally recognised rights, whereas SIA proceeds following a scoping process whereby all stakeholders (including the affected communities) nominate key issues in conjunction with the expert opinion of the assessor in terms of what the key issues might be based on experience in similar cases elsewhere and a conceptual understanding").

[102] See also Ian Harris and others, 'Ethical Assessment of New Technologies: A Meta-methodology' (2011) 9(1) Journal of Information, Communication and Ethics in Society 49–64; Erin Kenneally, Michael Bailey, and Douglas Maughan. 'A Framework for Understanding and Applying Ethical Principles in Network and Security Research' in Radu Sion and others (eds) *Financial Cryptography and Data Security* (Springer, 2010).

[103] See, e.g., with regard to stakeholders' engagement Wright and Mordini (n 21) 397 ("One of the objectives of an ethical impact assessment is to engage stakeholders in order to identify, discuss and find ways of dealing with ethical issues arising from the development of new technologies, services, projects or whatever").

[104] See Wright and Mordini (n 21) 399 ("With specific regard to values, it draws on those stated in the EU Reform Treaty, signed by Heads of State and Government at the European Council in Lisbon on 13 December 2007, such as human dignity, freedom, democracy, human right protection, pluralism, non-discrimination, tolerance, justice, solidarity and gender equality"). See also Ingrid Callies and others, 'Outline of an Ethics Assessment Framework' (2017) 31 < http://satoriproject.eu/media/SATORI-FRAMEWORK-2017-05-03.pdf> accessed 27 April 2018. For broader analysis of ethical issue in risk assessment, see also Lotte Asveld and Sabine Roeser (eds) *The Ethics of Technological Risk* (Earthscan, 2009).

[105] See also SATORI project (n 40) 13–14 and 18–19 <http://satoriproject.eu/media/CWA-SATORI_part-2_WD4-20170510W.pdf> accessed 14 January 2018; Wright and Friedewald (n 21) 760-762.

[106] See Callies et al. (n 104). See also Jules Polonetsky, Omer Tene and Joseph Jerome, 'Beyond the Common Rule: Ethical Structures for Data Research in Non-Academic Settings' (2015) 13 Colorado Technology Law Journal 333–367; Consumer Privacy Bill of Rights, §103(c) (Administration Discussion Draft 2015) <https://www.whitehouse.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf> accessed 14 January 2018.

and operationalise human rights in a given context, beyond some theoretical limits of the human rights framework.[107] As an assessment tool, it fosters the adoption of a preventive approach to product/service development from the earliest stages, favouring a focus on safeguards to rights and values, and a responsible approach to technology development.

- By stressing ethical and social values, HRESIA makes explicit the non-legal values that inform the courts and DPAs in their reasoning when they apply general principles of data protection, interpret general clauses or balance conflicting interests.
- In considering ethical and social issues, this model makes it possible to give flexibility to the legal framework, going beyond its theoretical limits in dealing with of Big Data and AI applications. A human rights assessment that operates through the lens of ethical and social values can therefore better address the challenges of the developing digital society.

Given the above, it is worth pointing out the benefits of this model in respect of the different PIA/DPIA standards adopted in several countries. The PIA models mainly focus on the individual dimension of data protection[108] and ignore the ethical and social issues.[109] Moreover, in terms of safeguarded rights, the main focus of the PIA concerns data protection, leaving little room for other fundamental rights and freedoms. In this sense, the Data Protection Impact Assessment adopted by the EU legislator does not seem to significantly improve these forms of assessment, since the DPIA is a self-assessment combined with the potential control of the Supervisory Authorities.[110]

Moreover, the DPIA only partially addresses the main issues and challenges associated with data use. The EU legislator, both in the Directive 95/46/EC and in the GDPR, introduces provisions that are primarily focused on data security and data quality, without directly and broadly addressing the different social and ethical issues of data use or providing mechanisms to measure the various adverse effects on individuals and society.[111]

For these reasons, a self-assessment model such as the HRESIA may contribute to the evolution of the existing DPIA towards a more complete assessment model. From this perspective the HRESIA can be seen as putting into practice the EU legislator's intention to safeguard not only the right to the personal data protection, but also the "fundamental rights and freedoms of natural persons", as stated among the main aims of the Regulation[112] and the specific provisions on risk management.[113]

The HRESIA model also takes into account the social and collective dimension, which is still not adequately addressed by the data protection regulation and only partially explored by legal scholars.[114] Recital 75 of the GDPR describes the risk of a "significant economic or social disadvantage" because of data processing, but the GDPR and the DPIA models do not elaborated further on the societal consequences of data processing.

3. The HRESIA architecture and its components

This section describes the main elements of the HRESIA model. This article is intended to put forward a blueprint and as such does not analyse these elements in depth or discuss the different related issues, but provides a general overview of the model, which will be discussed further in a future publication drawing on this ongoing study.

To combine the universal approach of human rights, the local dimension of social and ethical values and the tailored application of these complementary frameworks to a given data processing operation the HRESIA model presents a three-layer architecture. The first general layer is represented by the common values,[115] i.e. human rights and related process principles,[116] whose relevance in the context of data protection should be examined in light of the jurisprudence of the DPAs and the courts.[117]

[107] See above Section 2.2..

[108] See Raab and Wright (n 21).

[109] See Wright and Mordini (n 21).

[110] The first stage (i.e. the DPIA) largely consists of an internal assessment, whose results are not publicly available. In this regard, the guidelines provided by the Article 29 Data Protection Working Party seem to be an attempt by the Supervisory Authorities to mitigate this shortcoming and encourage controllers to take an approach that is more oriented to the data subject's engagement and transparent assessment. However, the weaknesses of the GDPR legal framework in terms of the participatory assessment and transparency of the DPIA, as well as the evident scarcity of Supervisory Authority resources illustrate how the compromise reached in the GDPR is a missed opportunity to adopt a stronger risk management model. For these reasons, despite the potential fines for infringement of GDPR requirements (Article 83), there is a real risk that, in various countries, many controllers will prefer to underestimate the data processing risks and not seek prior consultation with the Supervisory Authorities for their processing operations.

[111] See Council of Europe (n 9), Section IV, para 2.3 ("Since the use of Big Data may affect not only individual privacy and data protection, but also the collective dimension of these rights, preventive policies and risk-assessment shall consider the legal, social and ethical impact of the use of Big Data, including with regard to the right to equal treatment and to non-discrimination").

[112] See Article 1.2, GDPR.

[113] See e.g. Articles 24.1,25.1,32.1, 33.1.53.1, GDPR.

[114] See above fn. 21.

[115] See also Reuben Binns, 'Algorithmic Accountability and Public Reason' (2017) *Philosophy & Technology*, 1–14 <https://link.springer.com/content/pdf/10.1007%2Fs13347-017-0263-5.pdf> accessed 12 April 2018 (on the role of public reason as a constraint in algorithmic accountability)

[116] The human rights-based approach includes a number of 'process' principles, namely: participation and inclusion, non-discrimination and equality, and transparency and accountability. See The Danish Institute for Human Rights (n 27) 35.

[117] Apart from the central role of privacy and data protection, a first analysis of the decisions concerning data processing reveals the crucial role played by the principles of non-discrimination, transparency and participation as well as the safeguarding of hu-

The second layer is represented by the effect of the social and ethical values on the interpretation of these human rights. These values represent the societal factors that influence the way the balance is achieved between the different human rights and freedoms, in different contexts and in different periods. Moreover, social and ethical values concur in defining the extension of rights and freedoms, making possible broader forms of protection when the regulatory framework does not provide adequate answers to emerging issues.[118]

Finally, the third layer concerns the assessment of the concrete case based on specific sets of rights, values and principles. To operationalise this theoretical framework in an assessment tool, the suggested model is composed of two different components: a HRESIA questionnaire and an *ad hoc* committee (hereinafter HRESIA Committee).

Since it is impossible to adopt a prescriptive approach when assessing the impact of data use, data controllers must consider the elements that are relevant in the specific case, both in terms of data processing and its potential impacts. The questionnaire therefore serves as a tool to support data controllers in identifying the relevant human rights issues for any given application along the lines of similar models adopted in the field of data protection (PIA and DPIA). The questionnaire embeds human rights principles and values and, depending on the context, may also place them in a framework of the local ethical and social values.

The HRESIA committee assists in this contextualization and moreover applies the HRESIA benchmark values to the given case, balancing interests that may be in conflict, assessing and mitigating the risks. Of course, where assessment is easy the committee may be not necessary and data controllers can assess the risks and mitigate them on their own using the questionnaire alone.

This blueprint does not intend to provide a list of questions to be adopted in the HRESIA model. It is part of an ongoing research study[119] and the questionnaire must be carefully drafted and tested further to validate the questions. However, from a methodological perspective, the questionnaire can be built based on the various existing HRIA, PIA and PDIA models, adapting them to the specific perspective of the HRESIA. It should cover a range of areas concerning not only the various human rights and freedoms relevant to data processing, but also the procedural aspects of the participatory approach and the disclosure of assessment results.

Regarding the potential role of the participatory approach, the results of the HRESIA may suggest the engagement of specific categories of individuals – giving voice to the different groups of persons potentially affected by the use of data – and other stakeholders[120] (e.g. NGOs, public bodies).[121] The same conclusion might follow from the advice of the HRESIA Committee,[122] which represents the second component of the model. Moreover, this participatory approach[123] can also be useful to get a better understanding of the different competing interests and ethical and social values.[124]

man dignity, physical integrity and identity, as well as freedom of choice, of expression, of education, and of movement. These results are part of an ongoing research programme on "Legal and regulatory issues of data processing and related social impacts", see below in the section Acknowledgments.

[118] In order to absorb social values in the legal framework, a role can be played by the open clause that authorises restriction to fundamental rights when "necessary in a democratic society", a limitation present in several articles of the ECHR and, including Article 8.2. See De Hert (n 27) 53-54. Anyway, also in this case a balance of interests in terms of proportionality is necessary, see European Court of Human Rights, *S. and Marper* v. *The United Kingdom*, Judgement of 4 December 2008, Applications nos. 30562/04 and 30566/04, § 125; see also Serge Gutwirth and Paul De Hert, 'Privacy, data protection and law enforcement. Opacity of the individual and transparency of power' in Erik Claes, Antony Duff and Serge Gutwirth (eds) *Privacy and the criminal law* (Intersentia, 2006), 91.

[119] See below Acknowledgement.

[120] Stakeholders, different from groups directly affected by data processing, play a more relevant role in those contexts where direct consultation may put groups at risk, due to the law safeguards provided by local jurisdictions to human rights. See also Kemp and Vanclay (n 84) 92 ("For situations where direct consultation may put groups at risk, it may be necessary to engage third parties, such as NGOs or other agencies or individuals who have worked closely with particular groups. Assessment teams must be vigilant about ensuring that individuals and groups are not put at risk by virtue of the human rights assessment itself").

[121] For a different approach to participation, more oriented towards participation of lay people in committees of experts – in the context of Technology Assessment, see Skorupinski and Ott (n 3) 117–120.

[122] See also the following section.

[123] The role of participatory approaches and stakeholders' engagement is specifically recognised in the context of fundamental rights. See The Danish Institute for Human Rights (n 27) 24; De Hert (n 27) 72 ("Further case law is required to clarify the scope of the duty to study the impact of certain technologies and initiatives, also outside the context of environmental health. Regardless of the terms used, one can safely adduce that the current human rights framework requires States to organise solid decision-making procedures that involve the persons affected by technologies").

[124] Participation of the different stakeholders (e.g. engagement of civil society and the business community in defining sectoral guidelines on values) can achieve a more effective result than mere transparency, although the latter has been emphasized in the recent debate on data processing. See The Danish Institute for Human Rights (n 27) 10 ("Engagement with rights-holders and other stakeholders are essential in HRIA […] Stakeholder engagement has therefore been situated as the core cross-cutting component"). See also Walker (n 27), 41 ("participation is not only an end – a right – in itself, it is also a means of empowering communities to influence the policies and projects that affect them, as well as building the capacity of decision-makers to take into account the rights of individuals and communities when formulating and implementing projects and policies"). A more limited level of engagement, focused on awareness, was suggested by Council of Europe-Committee of experts on internet intermediaries (MSI-NET) (n 29) 45 ("Public awareness and discourse are crucially important. All available means should be used to inform and engage the general public so that users are empowered to critically understand and deal with the logic and operation of algorithms. This can include but is not limited to information and media literacy campaigns. Institutions using algorithmic processes should be encouraged to provide easily accessible explanations with respect to the procedures followed by the algorithms and to how decisions are made. Industries that develop the analytical systems used in algorithmic decision-making and data collection processes have a particular responsibility to create awareness and understanding, including

Finally, stakeholder engagement is a development goal for the assessment,[125] since it reduces the risk of under-representing certain groups and may also flag up critical issues that have been underestimated or ignored by data controller.[126]

However, as has been pointed out in risk theory,[127] stakeholder engagement should not become a way for decision makers (data controllers, in this case) to avoid their responsibilities as leaders of the entire process. Decision makers remain committed to achieving the best results in terms of minimising the potential negative impacts of data processing on individuals and society.

3.1. *The role of expert committees in the HRESIA*

The increasing and granular availability of data about individuals provided by IoT devices and online services enable private corporations to collect large amounts of data and use this to extract further information about individuals and groups. This has led private companies to carry out social investigations, which can be classed as research activities, traditionally carried out by research bodies. This raises new issues since private firms do not have the same ethical[128] and scientific background as researchers.[129]

To address this lack of expertise, the literature has suggested the adoption of ethical boards, which may act at a national level, providing general guidelines, or at a company level, supporting data controllers with regard to specific data applications.[130] However, these proposals limit their focus to ethical issues on the one hand and on the other do not situate these ethical boards within a broader framework of rights and values.[131] Such shortcomings highlight the self-regulatory nature of these solutions lacking a strong general framework that could provide a common baseline for data processing.

On the other hand, the adoption of HRESIA committees would build on the human rights framework outlined above representing a sound and common set of values to guide the committees' decisions. HRESIA committees are not asked to define the general ethical principles but, based on the HRESIA questionnaire, contextualise these human rights and freedoms.

In defining the status of these committees, the first issue concerns their internal or external nature.[132] This question is closely linked to their independence, which is vital if they are to be above the competing interests of the data controllers whose activities are assessed.[133]

The main issue underpinning the internal/external nature of the HRESIA committees[134] therefore depends on the degree of independence they are allowed. In theory, they could be both internal and external, providing there are no conflicts of interests, which may occur more frequently with in-house committees, but cannot be excluded in the case of external committees. Best practice in the prevention of conflicts of interests should therefore be applied.

As regard their function, the committee plays an important role in scaling down the complexity of traditional human rights impact assessment models. In this sense, the committee experts (and the HRESIA questionnaire) replace empirical analysis, which is required in the HRIA to define the assessment baseline in terms of the relevant rights and their local dimension as framed by jurisprudence and local socio-ethical issues.

---

with respect to the possible biases that may be induced by the design and use of algorithms").

[125] See also United Nations Office of the High Commissioner for Human Rights, 'Frequently asked questions on a human rights-based approach to development cooperation' (New York and Geneva: United Nations, 2006).

[126] See Wright and Mordini (n 21) 402.

[127] See Elin Palm and Sven Ove Hansson, 'The case for ethical technology assessment (eTA)' (2006) 73(5) Technological Forecasting & Social Change 543, 550–551.

[128] See, e.g., Council of Europe, 'Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine' (1997); National Commission for the Protection of Humans Subjects of Biomedical and Behavioural Research, 'Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research' (1979) <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html> accessed 12 March 2018.

[129] See e.g. Stuart Schechter & Cristian Bravo-Lillo, 'Using Ethical-Response Surveys to Identify Sources of Disapproval and Concern with Facebook's Emotional Contagion Experiment and Other Controversial Studies' (2014) <http://research.microsoft.com/pubs/220718/CURRENT%20DRAFT%20-%20Ethical-Response%20Survey.pdf> accessed 12 March 2018; Adam D. I. Kramer, Jamie E. Guillory & Jeffrey T. Hancock, 'Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks' (2014) 24 Proc. Nat'l Acad. Sci. <http://www.pnas.org/content/111/24/8788.full.pdf> accessed 12 March 2018. See also Ryan Calo, 'Digital Market Manipulation' (2014) 82(4) Geo. Wash. L. Rev. 995, 1046.

[130] See Ryan Calo, 'Consumer Subject Review Boards: A Thought Experiment' (2013) 66 Stan. L. Rev. Online 97 (2013), http://www.stanfordlawreview.org/online/privacy-and-big-data/consumer-subject-review-boards; Polonetsky, Tene and Jerome (n 106). See also White House, 'Consumer Privacy Bill of Rights' §103(c) (Administration Discussion Draft 2015) <https://www.whitehouse.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf> accessed 12 March 2018; The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (n 23) 43-44, 47, 49, 53.

[131] See, e.g., The Information Accountability Foundation, 'Artificial Intelligence, Ethics and Enhanced Data Stewardship' (2017) <http://informationaccountability.org/wp-content/uploads/Artificial-Intelligence-Ethics-and-Enhanced-Data-Stewardship.pdf> accessed 12 March 2018.

[132] The internal and external nature of the committees also depends on the availability of in-house skills and the costs of setting up an internal committee. See in this sense Polonetsky, Tene and Jerome (n 106) 354.

[133] See also Polonetsky, Tene and Jerome (n 106), 341, who point out the consequences of this choice with regard to confidentiality of information ("On the other hand, advocates would not be satisfied with a process that is governed internally and opaque. The feasibility of CSRBs thus hinges on the development of a model that can ensure rapid response and business confidentiality while at the same time guaranteeing transparency and accountability").

[134] Regarding other potential consequences of the different internal/eternal nature of committees assessing data use, see also Polonetsky, Tene and Jerome (n 106), 353–356.

These tasks will obviously have an influence on the composition of the committees, since the people involved must be able to carry out this kind of analysis. Legal expertise, an ethical and sociological background, as well as domain-specific knowledge (of the data application) is required. Moreover, the composition and number of experts will also depend on the complexity of data use. To offset the costs, permanent committees may be set up by groups of enterprises or serving all SMEs in a given area.

The quality of these committees and their work might be enhanced by empowering the Data Protection Authorities (DPAs) to supervise them. These authorities would not have to authorise the creation of the committees or perform any prior assessment of their compositions or activities, but data subjects could ask the DPAs to scrutinize the committees when shortcomings in their abilities or decisions affect the data processing.

The HRESIA committees' main job is to consider the specific data use and place it in the local context, providing a tailored and more granular application of the rights and freedoms underpinning the HRESIA model. Committees may therefore decide that this contextual application of general principles and values requires the engagement of the groups of individuals potentially affected by data processing[135] or institutional stakeholders. In this sense, the HRESIA is not a mere desk analysis, but takes a participatory approach – as described earlier[136] – which may be enhanced by the work of the HRESIA committee.

To guarantee the transparency and the independence of the HRESIA committee and its deliberations, it should adopt procedures to regulate its activity, also with regard to stakeholder engagement. In addition, it should fully document its decisional process and the documentation should be recorded and archived for a specific period depending on the type of the data use.

Finally, given the self-assessment character of the HRESIA model, the committees' opinions are not mandatory for the data controllers but help them to better assess the impact of their data processing decisions. On the other hand, data controllers may cite the committee's conclusions in defending the adequacy of their data processing decisions. Thus, the HRESIA committee can play an indirect role in the data controller's accountability.

4. Conclusions

The increasing use of Big Data analytics and AI in decision-making processes highlights the importance of examining their potential impact on individuals and society at large. The consequences of data processing are no longer restricted to the well-known privacy-related issues, but encompass prejudices against groups of individuals and a broader array of fundamental rights. Moreover, the tension between the extensive use of Big Data and AI, on the one hand, and the increasing demand for ethically and socially responsible data use on the other, reveals the lack of a regulatory framework that can address the societal issues raised by these data-intensive technologies.

Against this background, neither the traditional data protection impact assessment models (PIA and DPIA) nor the broader social or ethical impact assessment procedures (SIA and EtIA) appear to provide an adequate answer to the challenges of our algorithmic society. While the former have a narrow focus – centred on data quality and data security – the latter cover a wide range of issues, broad theoretical categories and heterogeneous solutions.

A human rights-centred assessment therefore offers a better answer to the demand for a more comprehensive assessment, including not only data protection issue, but also the effects of data use on other fundamental rights and freedoms (e.g. freedom of movement, freedom of expression, of assembly and freedom in the workplace) and related principles (e.g. non-discrimination). Moreover, a human rights assessment is grounded on the charters of fundamental rights, which provide the common baseline for assessing data use required in the context of the global data processing policy.

The development of a self-assessment model based on human rights can contribute to the evolution of the existing DPIA models towards a more complete assessment model. The proposed Human Rights, Ethical and Social Impact Assessment (HRESIA) is more closely aligned with the true intention of the EU legislator to safeguard not only the right to personal data protection, but also the fundamental rights and freedoms of natural persons. Furthermore, attention to the ethical and social issues in the HRESIA model enables it to take into account the social and collective dimension of data use, which is still not adequately addressed by the data protection regulations and only partially studied by legal scholars.

Finally, ethical and social values, viewed through the lens of human rights, make it possible overcome the limitations of the traditional human rights impact assessment and help to interpret human rights in a manner consistent with the regional context. In this way, the HRESIA aims to provide a universal tool that can take the local dimension of the safeguarded interests into account.

To achieve these goals the HRESIA model includes a self-assessment questionnaire in line with the traditional impact assessment approach, and an *ad hoc* committee. These two components make it possible to scale the complexity of the human rights impact assessment down, defining the value framework (questionnaire) and placing it in the context of the local dimension, as well as providing a tailored and more granular application of the rights and freedoms underpinning the HRESIA model.

Based on this architecture, this assessment tool can raise awareness among data controllers with respect to the impact of their data processing choices on individuals and society. At the same time, a participatory and transparent assessment model like the HRESIA also gives data subjects an opportunity for more informed choices concerning the use of their data, and increases their awareness about the consequences of data processing.

Though this assessment may represent an additional burden for data controllers, its voluntary self-assessment nature

[135] On the nature of these groups and its potential influence on the difficulty to engage them in the assessment, see also Mantelero, 'Personal data for decisional purposes' (n 7).

[136] See the previous section.

may encourage its diffusion in spheres where the data subjects pay greater attention to the ethical and social implications of data use (healthcare, services/products for kids, etc.) or in the presence of socially oriented entities or developers' communities. Moreover, as has happened in other sectors, a greater attention to human rights and societal impacts may represent a competitive advantage for companies that deal with responsible consumers and partners.

Finally, the focus of policymakers, industry and communities on ethics and responsible use of data, on the one hand, and the lack of adequate tools to assess the impacts of data processing on the fundamental rights and freedoms protected by legislators (e.g. GDPR), on the other, make the HRESIA a possible solution in the direction of a broader consideration of the consequences of data processing.

Acknowledgement

This article presents the first results of an ongoing research programme on "Legal and regulatory issues of data processing and related social impacts" (Polytechnic University of Turin, 2017–2022, PI: Prof. Alessandro Mantelero). The results of this project, regarding the HRESIA model outlined here, are expected at the begin of 2019 and will be published in A. Mantelero (ed.), Addressing social and ethical issues in data processing, Springer (forthcoming 2019). I am grateful to Joe Cannataci for the comments he provided during the first presentation of my thoughts on this topic at the Expert workshop on the right to privacy in the digital age organised by the Office of the United Nations High Commissioner for Human Rights (Geneva, February 19–20, 2018). I am also grateful to Samantha Esposito for the analysis of DPAs' jurisprudence; her research has been partially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 732027 (Virt-EU project).