Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Legal framework for the use of big data and blockchain in public governance

Tombal, Thomas; Willem, Pauline; De Terwangne, Cecile

Published in:
The new digital era governance

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Citation for pulished version (HARVARD):
Tombal, T, Willem, P & De Terwangne, C 2022, Legal framework for the use of big data and blockchain in public governance. in E Tan & J Crompvoets (eds), *The new digital era governance: how new digital technologies are shaping public governance.* Wageningen Academic Publishers, Wageningen , pp. 111-140.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 02. May. 2026

Chapter 4

Legal framework for the use of big data and blockchain in public governance

T. Tombal[1,2], P. Willem[3*] and C. De Terwangne[3]

[1]Faculty of Law, University of Namur, Rempart de la Vierge, 5, 5000 Namur, Belgium; [2]Tilburg Institute for Law and Technology, Tilburg University, Prof. Cobbenhagenlaan 221, 5037DE, Tilburg, the Netherlands; [3]Centre de recherche Information, Droit et Société, Namur Digital Institute, University of Namur, Rempart de la Vierge, 5, 5000 Namur, Belgium; pauline.willem@unamur.be

Abstract

By using technologies, such as big data and blockchain, public administrations are able to process a large amount of citizens' personal data. When processing these personal data, the administration must comply with the personal data protection rules contained in the General Data Protection Regulation. As the use of such technologies could have dramatic impacts on the lives of their citizens, it is fundamental to understand the limits that this legal framework puts on their use. This chapter analyses the interactions between personal data protection and big data and blockchain technologies. It explains the legal framework within which such technologies can be leveraged by public administrations for the provision of their public services. Focussing on data collection and combination (big data), this chapter concludes that the processing of personal should be based on a law meeting certain requirements. Data subjects' rights must also be respected. Regarding data storage (blockchain), several key and concrete takeaways are formulated for public administrations. While most of the following analysis will be equally applicable to any public administration within the European Union, this chapter will focus on two case studies within the Belgian public administration, namely the public policies and decision-making linked to social security infringements and tax fraud at the Belgian federal level. Indeed, these could have a significant impact on citizens' finances in particular, and their lives in general.

Keywords: big data, blockchain, GDPR, public administration, fraud

Evrim Tan and Joep Crompvoets (ed.) **The new digital era governance**
DOI 10.3920/978-90-8686-930-5_4, © Wageningen Academic Publishers 2022

4.1 Introduction

Public administrations consistently use more and more data, including their citizens' personal data,[1] to deliver their public services. Yet, when processing[2] these personal data, they have to comply with the personal data protection rules. For the public administrations of Members States of the European Union,[3] these rules are contained in the General Data Protection Regulation (hereafter 'GDPR').[4] The adoption of the GDPR presents several challenges for the administration as it is directly applicable to them and they might thus have to revise their former way of processing personal data. Indeed, the principles of accountability (Article 29 Working Party, 2010)[5] and of data protection by design and by default,[6] which are at the core of the GDPR, were not explicit in the former Directive 95/46,[7] and the administrations had to adapt their practices in order to meet the new standards set by the GDPR.

These privacy and personal data protection rules are especially important with the advent of new technologies, such as big data, artificial intelligence and blockchain, which bring new capabilities to public administration to process greater amounts of data in public service processes with unique implications in data privacy, transparency, security and governance. As the use of such technologies could have strong impacts on the lives of their citizens, it is fundamental to understand the limits that this legal framework puts on their use.

[1] Personal data is defined as 'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person' (Art. 4.1 of the GDPR).

[2] Processing 'means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction' (Art. 4.2 of the GDPR).

[3] For examples of data protection rules in other countries, see for instance the UK Data Protection Act (2018), available at https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted; or the Australian Privacy Act (1988), available at https://www.legislation.gov.au/Details/C2021C00242. In the USA, there is no single Federal State legislation pertaining to data protection, but rather hundreds of (sector-specific) Federal and State legislations, some of which focus on a particular type of data (for more information and a brief overview, see https://iclg.com/practice-areas/data-protection-laws-and-regulations/usa).

[4] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46 (General Data Protection Regulation), OJ [2016] L 199/1. The GDPR was adopted on the 27 April 2016 and is applicable since the 25 May 2018.

[5] Art. 5.2 of the GDPR.

[6] Art. 25 of the GDPR.

[7] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ [1995] L 281/31. In Belgium, which is the country that will be used as a case study (see *infra*), the provisions of the Directive had been included in an existing law of 1992 (Loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel, *M.B.*, 18 mars 1993).

Accordingly, this chapter and the following one will delve into the interactions between personal data protection and big data, artificial intelligence and blockchain technologies, in order to draw the legal framework within which such technologies can be leveraged by public administrations for the provision of their public services. To do so, the legal challenges will be structured around the data-processing lifecycle for decision-making, focussing first on data collection and data combination (big data – Section 4.2), and then on data storage (blockchain – Section 4.3). The legal considerations pertaining to the use of artificial intelligence to facilitate data analysis will, on the other hand, be dealt with in Chapter 5. Naturally, all of these technologies can potentially apply at all stages of the data-processing lifecycle, but for readability purposes, it has been decided to relate each technology to the stage of the lifecycle where it has the most impact. However, this does not mean that it does not have an impact on the other stages. Indeed, these technologies are intertwined and influence each other throughout the data-processing lifecycle.

While most of the following analysis will be equally applicable to any public administration within the European Union, this chapter will focus on one Member State, namely Belgium, in order to illustrate in more detail some of the legal challenges that public administrations face in practice when they wish to rely on these new technologies. More specifically, particular attention will be devoted, in this chapter, to two case studies within the Belgian public administration, namely the public policies and decision-making linked to social security infringements and tax fraud at the Belgian federal level, as these could have a significant impact on citizens' finances in particular, and their lives in general.

4.2 Data collection and combination – personal data protection and big data

As pointed out by Zarsky, 'when striving to define the 'big data' concept, the professional literature refers to the four V's: the Volume of data collected, the Variety of sources, the Velocity with which the analysis of the data can unfold, and the Veracity of the data which could (arguably) be achieved through the analytical process' (Zarsky, 2017). Big data represents 'a fundamental change in the way data is *collected*, *stored*, and subsequently *used* – all a result of recent technological developments' (emphasis in the original text) (Zarsky, 2017). Big data relies on the fact that large volumes of various data are gathered in order to extract information and draw inferences that would have otherwise not been possible to extract/draw with smaller volumes of data (Mayer-Schönberger and Padova, 2016). The more data is gathered, the more information can potentially be extracted from combining this data, in order, for example, to draw up public policies or to take decisions. For public administration, implementing big data technologies could notably derive from a greater sharing of data between administrations, thus breaking the existing silos that exist between these administrations (Chantillon *et al.*, 2017). It could also derive from the collection of data, held by private sector firms, for public interest purposes (High-Level Expert Group on Business-to-Government Data Sharing, 2020; Richter, 2020).[8]

[8] See also the French 'Loi Lemaire': Loi n° 2016-1321 pour une République numérique, 7 octobre 2016 (especially Articles 17 to 22 pertaining to 'data of general interest' ('données d'intérêt general')).

With the advent of today's technology, data can be collected more easily, notably through sensors and the Internet of Things (IoT); it can be stored in larger quantities, as the cost of storage is constantly decreasing; and it can be used in a wide variety of contexts, notably for decision-making (Zarsky, 2017). However, such an increase in the power of data analytics can have an impact on the data subjects' privacy and personal data protection. As pointed out by Zarsky, there is a 'double-sided tension' between big data analytics and personal data protection (Zarsky, 2017). As this author outlines, 'on the one hand, these advanced forms of data analyses can compromise the individuals' privacy rights and the control citizens have over their personal data. Thus, the availability of these tools might require stricter enforcement of privacy laws to so limit privacy-related harms. On the other hand, (...) stricter data protection and privacy laws compromise the growth of the big data industry and the benefits to be derived from it' (Zarsky, 2017).

Therefore, big data analytics, at least in some of its forms, is in tension with personal data protection. Indeed, the potential of big data relies on the fact that 'data needs to be gathered at an unprecedented scale whenever possible, and reused for different purposes over and over again (...) This puts big data on a direct collision course with the core principles of existing data protection laws' (Mayer-Schönberger and Padova, 2016). This is specifically apparent when focussing on five of the GDPR's provisions, namely the prohibition to process special categories of data,[9] the purpose limitation principle,[10] the data minimisation principle,[11] the data subject's right to information,[12] and the data subject's right not to be subject to automated individual decision-making[13] (Zarsky, 2017). Before delving into each of these provisions, it must first be outlined that any processing of personal data by the public administration must be lawful and fair.[14]

4.2.1 Lawfulness and fairness of processing and big data

The GDPR stipulates that the processing of personal data will be lawful 'only if and to the extent that at least one of the following applies: (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes; (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; (c) processing is necessary for compliance with a legal obligation to which the controller is subject; (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (f) processing is necessary for the purposes of the legitimate interests

[9] Art. 9 of the GDPR.
[10] Art. 5.1.b) of the GDPR.
[11] Art. 5.1.c) of the GDPR.
[12] Arts. 12 to 14 of the GDPR.
[13] Art. 22 of the GDPR.
[14] Art. 5.1.a) of the GDPR.

pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child'.[15]

While a quick read of the GDPR might give the impression that the public authorities could thus potentially rely on six lawful bases of processing, in reality, two of those bases should be avoided by public authorities. On the one hand, the GDPR explicitly states that the 'legitimate interests' basis shall not apply to processing carried out by public authorities in the performance of their tasks.[16] In practice, this is not a major problem for public administrations as they will rely instead on the 'performance of a task carried out in the public interest' lawful basis.

On the other hand, public administrations should avoid relying on the data subject's consent as a lawful basis of processing. To understand why this is the case, the definition of consent should first be re-stated. According to the GDPR, the consent of the data subject 'means any *freely given*, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her'[17] (emphasis added). The consent of the data subject thus needs to be freely given. In this regard, the GDPR outlines that 'consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment'.[18] This will especially be the case 'where there is a clear imbalance between the data subject and the controller, in particular *where the controller is a public authority and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation*'[19] (emphasis added). This is confirmed by the European Data Protection Board (hereafter 'EDPB') in its revised guidelines on consent, though it outlines that it might be appropriate to use consent under certain circumstances, such as consenting to be included in a mailing list to receive information about the progress of road works (European Data Protection Board, 2020a).

Nevertheless, as suggested by the EDPB, other lawful bases are more appropriate to the activity of public authorities, namely either the processing necessary for compliance with a legal obligation or processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller (European Data Protection Board, 2020a).[20] In practice, the lawful basis will often be a law. For instance, at the Belgian federal level, personal data processing to fight tax fraud is organised by the law of 3 August 2012,[21] while personal data

[15] Art. 6.1 of the GDPR.
[16] Art. 5.1 of the GDPR.
[17] Art. 4.11 of the GDPR.
[18] Recital 42 of the GDPR.
[19] Recital 43 of the GDPR.
[20] Respectively arts. 6.1.c) and 6.1.e) of the GDPR.
[21] Loi du 3 août 2012 portant dispositions relatives aux traitements de données à caractère personnel réalisés par le Service public fédéral Finances dans le cadre de ses missions, *M.B.*, 24 août 2012.

processing to tackle social security fraud is notably addressed by the law of 15 January 1990[22] and the 'Program Law (I)' of 29 March 2012.[23] It should be outlined from the outset that some of the provisions of the law of 3 August 2012 and of the law of 15 January 1990, which predate the adoption of the GDPR, have been modified or inserted in September 2018, in order to adapt them to the entry into force of the GDPR.[24]

Such laws, however, need to meet certain standards. Namely, they should specify 'the general conditions governing the lawfulness of processing by the controller; the types of data which are subject to the processing; the data subjects concerned; the entities to, and the purposes for which, the personal data may be disclosed; the purpose limitation; storage periods; and processing operations and processing procedures, including measures to ensure lawful and fair processing (…) [it shall also] meet an objective of public interest and be proportionate to the legitimate aim pursued'.[25] Moreover, as a rule of thumb, any personal data processing must be fair,[26] which implies that the laws on which this processing is based must be sufficiently explicit and understandable for the citizens. They cannot be taken by surprise and must be informed about this processing. This point, which is fundamental in terms of transparency, will be addressed in Section 4.2.5.

4.2.2 The prohibition to process special categories of data and big data

The processing of 'special categories of data', listed in Article 9.1 of the GDPR,[27] is, in principle, prohibited.[28] 'Data concerning health' are part of that list, and are defined as 'personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status'.[29] These types of data are important to mention in the big data context, where more and more data will be collected by sensors, such as a connected watch calculating your number of daily steps. Indeed, according to the Article 29 Working Party (today the European Data Protection Board), such data could be considered as revealing health-related information, and should accordingly benefit from the reinforced protection for 'special categories of data' (Article 29 Working Party, 2015; Zarsky, 2017). This finding is important for public policies and decision-making linked to social security infringements, as they might rely on 'data concerning health'.

[22] Loi du 15 janvier 1990 relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale, *M.B.*, 22 février 1990.
[23] Loi-programme (I) du 29 mars 2012, *M.B.*, 6 avril 2012. See in particular Article 101 of this Law.
[24] Loi du 5 septembre 2018 instituant le comité de sécurité de l'information et modifiant diverses lois concernant la mise en œuvre du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE, *M.B.*, 10 septembre 2018, arts. 9-40 and 70-85.
[25] Art. 6.3 of the GDPR.
[26] Art. 5.1.a) of the GDPR.
[27] 'Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, (…) genetic data, biometric data, data concerning health or data concerning a natural person's sex life or sexual orientation' (Art. 9.1 of the GDPR).
[28] Art. 9 of the GDPR.
[29] Art. 4.15 of the GDPR.

Moreover, some forms of big data analytics might blur the lines between special categories of data and 'regular' personal data. Indeed, 'an analysis merely relying on and addressing 'regular' categories can quite quickly end up pertaining to 'special categories.' For instance, health data can be deduced from a variety of datasets, such as shopping databases, and therefore this category has quickly and sharply expanded' (Zarsky, 2017). Once again, this is an important factor to consider for public administration relying on big data analytics for decision-making linked to social security infringements and tax fraud, as, even if they believe that they are relying on 'regular' data, the crossing of large numbers of such regular data could give rise to special categories of data, which can only be processed under stricter conditions.

Finally, it must be outlined that specific additional safeguards also apply to data relating to criminal convictions and offences (Art. 10 of the GDPR),[30] which might have to be considered for tax and social security fraud, for instance if the elaboration of algorithmic models to identify 'suspicious' profiles includes the analysis of historical criminal convictions and offences data.

4.2.3 Purpose limitation principle and big data

The purpose limitation principle provides that personal data shall be 'collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes' (Article 29 Working Party, 2013).[31] This means that the purpose of processing must be defined prior to the collection of the data, and that data that has been collected for a specific purpose cannot be further processed for a purpose that does not fit within this initial purpose. This principle, which is a core principle of personal data protection enshrined in Article 8.2 of the Charter of Fundamental Rights of the European Union, can be at odds with some forms of big data analytics, where a large number of data are collected 'for the sake of it', without a clearly defined purpose. Indeed, Mayer-Schönberger and Padova argue that 'there is a strong economic incentive to keep the data for as long as possible, and much beyond the initial use of it, to reuse it repeatedly as well as to combine it with other data' (Mayer-Schönberger and Padova, 2016).

One might think that resorting to broad definitions of purposes of processing could solve this incompatibility. However, as pointed out by Zarsky, 'trying to circumvent this limitation by initially defining a very broad and vague purpose for future uses would most likely not resolve this matter, as the stated purposes must also be 'specific.' Furthermore, stating an unnecessarily broad purpose might even be considered as 'illegitimate' and thus lead to unacceptable processing' (Article 29 Working Party, 2013; Hahn, 2021; Zarsky, 2017). In the same vein, obtaining consent from the data subjects for very broadly defined purposes would likely not

[30] 'Processing of personal data relating to criminal convictions and offences or related security measures based on Article 6(1) shall be carried out only under the control of official authority or when the processing is authorised by Union or Member State law providing for appropriate safeguards for the rights and freedoms of data subjects. Any comprehensive register of criminal convictions shall be kept only under the control of official authority'.
[31] Art. 5.1.b) of the GDPR.

amount to specific and informed consent (European Data Protection Board, 2020a; Mayer-Schönberger and Padova, 2016).[32] Indeed, the core idea of the purpose limitation principle is that the data subject should be aware of what is done with their data, and should be able to exercise control over this processing.

However, it should be remembered here that, according to Recital 50 and Article 6.4 of the GDPR, if data is processed for a new purpose that is compatible with the initial purpose of processing, no separate lawful basis is required. In this regard, it is important to outline that the GDPR provides that further processing for archiving purposes[33] in the public interest, scientific or historical research[34] purposes or statistical purposes[35] shall, in accordance with Article 89.1, not be considered incompatible with the initial purposes.[36] In such cases, appropriate technical and organisational safeguards, such as pseudonymisation,[37] would, however, have to be set.[38] As pointed out by Mayer-Schönberger and Padova, these exceptions could notably be used for some big data applications for statistical purposes (Mayer-Schönberger and Padova, 2016). Nevertheless, Recital 162 of the GDPR makes it explicit that pursuing statistical purposes implies that the personal data cannot be used in support of individual measures or decisions regarding any particular natural person. Accordingly, using big data statistical analysis 'to influence decision-making directly affecting a particular individual would be outside the meaning of 'statistical purposes,' and also violate the restrictions on automated individual decision-making, including profiling' (Mayer-Schönberger and Padova, 2016).

If one applies the above considerations to the specific topic of this contribution, the purpose limitation principle does not prevent the use of big data analytics for social security infringements and tax fraud per se, but only of big data analytics which do not have a specified, explicit and legitimate purpose. Accordingly, big data analytics can be applied for purposes that have been clearly defined in advance. Moreover, there must be a lawful basis for this processing. As a matter

[32] Art. 4.11 of the GDPR.
[33] According to Recital 158 of the GDPR, this should apply to bodies that have a 'legal obligation to acquire, preserve, appraise, arrange, describe, communicate, promote, disseminate and provide access to records of enduring value for general public interest'. This Recital also adds that the GDPR does not apply to deceased people.
[34] According to Recital 159 of the GDPR, scientific research purposes 'should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research (...) [and] should also include studies conducted in the public interest in the area of public health'. According to Recital 160 of the GDPR, historical research purposes 'should also include historical research and research for genealogical purposes, bearing in mind that this Regulation should not apply to deceased persons'.
[35] According to Recital 162 of the GDPR, statistical purposes mean 'any operation of collection and the processing of personal data necessary for statistical surveys or for the production of statistical results. Those statistical results may further be used for different purposes, including a scientific research purpose. The statistical purpose implies that (...) the personal data are not used in support of measures or decisions regarding any particular natural person'.
[36] Art. 5.1.b) and Recital 50 of the GDPR.
[37] 'Pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person' (Art. 4.5 of the GDPR).
[38] Art. 89.1 of the GDPR.

of principle, this should be a law (Art. 6.1.c) of the GDPR) because these personal data processing interfere with citizens' privacy (Degrave, 2015), and this law should be very specific regarding the purposes of processing it allows.[39]

In terms of tax fraud, Article 3 of the Law of 3 August 2012[40] states that the Belgian Federal Public Service (hereafter 'FPS') Finances can collect and process personal data to execute its legal missions, and that the data cannot be used for other purposes.[41] Within the FPS Finances, the various administrations and/or services of the FPS can exchange personal data, provided that they have the authorisation from the President of the Executive Committee.[42] The President can ask an opinion from the Information Security Committee in this regard.[43] Regarding, more specifically, the use of big data to fight tax fraud, Article 5.1 provides that the FPS Finances may aggregate data, collected to execute its legal missions, in a 'data warehouse' enabling 'data mining' and 'data matching' operations, including profiling. This can only be done to carry out, in the context of its legal missions, targeted controls on the basis of 'risk indicators' and of analyses on data coming from different administrations and/or services of the FPS Finances.

According to Degrave and Lachapelle, these purposes of data processing might be defined too broadly, as they simply refer to the execution of the FPS Finances' 'legal missions' (Degrave and Lachapelle, 2014). This might thus be problematic in terms of the purpose limitation principle. However, this concern is somewhat alleviated by the fact that the data miners have to fill in a Data Access Management (DAM) fiche, which has to be validated by the President of the Executive Committee of the SPF Finances.[44] In this DAM fiche, they have to state the objectives and purposes of the data mining and explain how it fits the organisation's mission. This is, however, not ideal from a democratic perspective (as Parliament does not define the concrete purposes of processing) and from a legal perspective (according to Article 8 of the European Convention on Human Rights[45] and Article 22 of the Belgian Constitution, the key elements of personal data processing by public administration must be clearly defined by law).

Personal data resulting from processing operations in the data warehouse shall be kept for no longer than is necessary for the purposes for which they are processed, with a maximal retention period of one year after the prescription of all actions falling within the competence of the controller.[46] In practice, a relevance check is performed every three months by the head of the data miners to see if the data are used in conformity with the DAM fiche and if the project is still relevant and advancing. If it is not, the data access ceases and the data must be deleted.

[39] Art. 6.3 of the GDPR.
[40] Loi du 3 août 2012 portant dispositions relatives aux traitements de données à caractère personnel réalisés par le Service public fédéral Finances dans le cadre de ses missions, *M.B.*, 24 août 2012.
[41] Art. 3, al.1 and 2 of the Law of 3 August 2012.
[42] Art. 4, al.1 of the Law of 3 August 2012.
[43] Art. 4, al.4 of the Law of 3 August 2012.
[44] Art. 4, al.1 of the Law of 3 August 2012.
[45] European Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 4 November 1950.
[46] Art. 5.1, al.3 of the Law of 3 August 2012.

In terms of social security fraud, the use of data-matching techniques, aimed at identifying incompatibilities in terms of social allocations, must be subject to a data transfer protocol, as provided in Article 20.1 of the Law of 30 July 2018,[47] unless provided otherwise in specific laws (e.g. in Article 15 of the Law of 15 January 1990 which requires, in some cases, a prior deliberation of the Information Security Committee). The protocol, which must notably contain the purposes of processing, must be submitted to the Data Protection Officers of the social security institutions (hereafter 'SSIs') involved in the sharing.[48] However, they are not subject to a prior validation by the Data Protection Authority, which would bring more certainty in terms of the legitimacy of the purpose of processing (Degrave, 2020a). Once this purpose is achieved, the data must be deleted.

SSIs also use data mining techniques. According to Article 5*bis* of the Law of 15 January 1990, they may aggregate and process data in a data warehouse, enabling them to carry out data-mining operations to prevent, establish, prosecute, and punish offences against social legislations that fall within their respective powers. This data warehouse is known as OASIS and has existed since 2005. According to Degrave, the purposes of processing in OASIS that are authorised by the law are not clearly defined, which could be problematic in terms of the purpose limitation principle (Belgian Privacy Commission, 2018; Degrave, 2020b). However, this concern is somewhat alleviated, although not optimally either from a democratic and legal perspective (see above), in the hypotheses contained in Articles 5*bis*, al.7 and 15 of the Law of 15 January 1990, as the authorisation to process data from the data warehouse must be subject to a prior deliberation by the Information Security Committee (ISC), which will evaluate the purposes of processing.[49] It must nevertheless be underlined here that the ISC should, in theory, be independent of the administrations (including the SSIs) to which it grants authorisations to process the data, which implies that its members should not also exercise mandates within these administrations.[50] This is currently not the case, which creates independence issues as some members of the CSI are both players and referees, and this has led to the launch of an infringement procedure by the European Commission against Belgium (Anonymous, 2021a,b; European Commission, 2021; Laloux, 2021). This situation will need to be remedied as soon as possible.

Moreover, if such a deliberation is not imposed, the data controllers taking part in the fraud detection processing will nevertheless have to conclude a protocol in this regard, notably specifying the desired processing purposes, as this is the standard for any exchange of personal data between administrations, in light of the accountability principle of the GDPR (Art. 5.2, GDPR).[51] In any case, personal data resulting from processing operations in the data warehouse shall be kept for no

[47] Loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel, *M.B.*, 5 septembre 2018.
[48] Art. 20.2 of the Law of 30 July 2018.
[49] Art. 5*bis*, al. 1 of the Law of 15 January 1990.
[50] Art. 52 of the GDPR; Loi du 5 septembre 2018 instituant le comité de sécurité de l'information et modifiant diverses lois concernant la mise en œuvre du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE, *M.B.*, 10 septembre 2018, arts. 3 and 5.
[51] Article 20.1 of the Law of 30 July 2018.

longer than is necessary for the purposes for which they are processed, with a maximal retention period not exceeding one year after the prescription of all actions falling within the competence of the data controller.[52]

4.2.4 Data minimisation principle and big data

The data minimisation principle is another fundamental principle of the GDPR. It provides that only the adequate, relevant and necessary data for the fulfilment of the specific purpose of processing shall be processed.[53] This implies that, in combination with the purpose limitation principle, the categories and amount of data that can be processed should be limited to what is necessary to meet this purpose. Once again, this core principle of the GDPR is in stark contrast with some forms of big data analytics, which aim at collecting as much data as possible, in order to extract information and draw inferences that would have otherwise not been possible to extract/draw with smaller volumes of data. Yet, there is a core tension between, on the one hand, collecting as much data as possible, without knowing which of it will actually be useful, and on the other hand collecting only the data that is necessary for a pre-defined specific purpose.

Moreover, the data minimisation principle can also be linked to the principle of storage limitation (Art. 5.1.e) of the GDPR), as the data can only be stored for the period during which it is useful for the specific purpose of processing, and has to be erased as soon as it is no longer necessary (Zarsky, 2017). Once again, this clashes with some forms of big data analytics that aim at retaining as much data as possible, for as long as possible, in order to discover potential new uses for the data (Zarsky, 2017).

Similar to the purpose limitation principle, the data minimisation principle can be more flexible for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, provided that they are subject to appropriate technical and organisational safeguards.[54] Here, pseudonymisation is explicitly mentioned as a potential safeguard to ensure the respect of the data minimisation principle. Mayer-Schönberger and Padova argue that, since most of big data analytics are statistical in nature, this exception could allow data controllers to retain personal data for longer than is necessary for the original purpose of processing (Mayer-Schönberger and Padova, 2016). However, according to Zarsky, this might not be a viable option for some big data applications, as 'removing identifiers to achieve pseudonymity can potentially undermine the quality of the results derived, as the data would be purposefully altered and the aggregation of different datasets would be rendered difficult' (Zarsky, 2017). Alternatively, anonymisation[55] does not seem to be a viable option either, as truly effective anonymisation is

[52] Art. 5*bis*, al. 4 of the Law of 15 January 1990.
[53] Art. 5.1.c) of the GDPR.
[54] Art. 89.1 of the GDPR.
[55] The ISO 29100 standard defines anonymisation as the 'process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party' (ISO 29100:2011, point 2.2, available at https://www.iso.org/obp/ui/#iso:std:iso-iec:29100:ed-1:v1:en).

difficult to achieve (Franceschi *et al.*, 2018; Graef *et al.*, 2018; Mayer-Schönberger and Padova, 2016; Wendehorst, 2017). This is especially true in light of the constant development of big data analytics, which increases the risk of re-identification of the data subjects. This failure to effectively anonymise personal data has been demonstrated several times in the literature (Rocher *et al.*, 2019; Sweeney, 2017), leading to the conclusion that what is often presented as anonymisation techniques are, in fact, merely pseudonymisation techniques.

Much like the purpose limitation principle, the data minimisation principle does not per se preclude the use of big data analytics by public administrations for decision-making linked to social security infringements and tax fraud. Rather, this principle frames the scope and scale of data that can be used. This therefore reinforces the need to clearly define in advance the specific purpose of the processing used to combat social security infringements and tax fraud, on the one hand, and the data that will be necessary to do so, on the other. The key is to be proportionate in the types of data collected and used. Even if public administrations could potentially have access to troves of data, a balance must be found not only with citizens' privacy and data protection, but also with commercial and professional secrecy requirements.

In terms of tax fraud, personal data exchange between the various administrations and/or services of the FPS Finances must be authorised by the President of the Executive Committee.[56] The President decides which types of personal data can be exchanged, on a systematic or ad hoc basis and for specific purposes, after having verified their adequacy, relevance and non-excessiveness.[57] Regarding data-mining operations in the data warehouse, Article 5 of the Law of 3 August 2012 provides that the FPS Finance can use 'data collected to execute its legal missions'. These are notably data collected from people's and undertakings' tax declarations, from the newspapers, from their own experience, from whistleblowers and from outputs of investigations.

According to Degrave and Lachapelle, the fact that Article 5 provides that the FPS Finance can use, via the data warehouse, any 'data collected in order to execute its legal missions' might be problematic from a data minimisation perspective, as this is a broad definition that does not define exactly which types of data can be used (Degrave and Lachapelle, 2014). However, this concern is somewhat alleviated by the fact that, as outlined above, a DAM fiche must be completed and submitted to the President of the Executive Committee. This constrains the data that data miners can access for a specific project. This is a pragmatic solution, as it would be very difficult for the legislator to predefine all the types of data that could be processed in this regard. Moreover, the technical access to the data warehouse is built in such a way that the agents of FPS Finances can only access the electronic records, data or applications that are adequate, relevant and non-excessive in light of the execution of the tasks that fall within their legal missions,[58] and this can

[56] Art. 4, al.1 of the Law of 3 August 2012.
[57] Art. 4, al.2 of the Law of 3 August 2012.
[58] Art. 10.1 of the Law of 3 August 2012.

be checked through access logs, which allows third party auditing (e.g. by the Data Protection Authority) of the process. This also materialises the data protection by design and by default principle.[59]

Regarding data-matching operations to fight social security infringements, they must be subject to a data transfer protocol or to a prior deliberation of the Information Security Committee (Section 4.2.3). In this regard, the SSIs must identify the data that are necessary and adequate for the data-matching purpose they pursue. Regarding the data-mining operations conducted in the OASIS database, Art 5*bis* of the Law of 15 January 1990 provides that 'all the necessary data for the purposes of applying the labour law and social security legislation' can be used. According to Degrave, this definition may be too broad as it does not allow citizens to know exactly which types of data are (or can be) processed (Degrave, 2020b). However, this concern is somewhat alleviated by the fact that access to data from the data warehouse must also be subject to a data transfer protocol or to a prior deliberation of the Information Security Committee, in which the necessary and proportionate nature of the accessed data will be controlled (Section 4.2.3).

Moreover, the data minimisation principle is enshrined in the fact that the data warehouse contains solely pseudonymised data and that it can only be accessed by a limited number of data miners/investigators. Importantly, the people who pseudonymise the data to be uploaded in the data warehouse are not the same as those who use the data warehouse in order to spot fraudulent patterns. In practice, the SSIs must draw up a list of the categories of persons with access to the personal data in the data warehouse, with a description of their role in relation to the data processing in question, and this list shall be kept at the disposal of the Data Protection Authority.[60] It is only once these data miners/investigators have identified a potential fraudulent case that the data at hand is de-pseudonymised, following a risk analysis, and extracted from the data warehouse, in order to start an investigation assessing whether there is indeed fraud.

4.2.5 The data subjects' right to information and big data

The right to information

Data has to be processed fairly and in a transparent manner.[61] This implies that citizens cannot be taken by surprise and that the public administration shall take appropriate measures to provide any information referred to in Articles 13 and 14 of the GDPR to the data subjects in a concise, transparent, intelligible and easily accessible form, using clear and plain language (Article 29 Working Party, 2018b).[62] If the public administration relies on big data to feed AI applications used to fight social security infringements and tax fraud, this means that the data subjects, whose data are collected, should notably receive information about:

[59] Art. 25 of the GDPR.
[60] Art. 5*bis*, al. 5 of the Law of 15 January 1990.
[61] Art. 5.1.a) of the GDPR.
[62] Art. 12.1 of the GDPR.

- The identity and contact details of the data controller and the contact details of its data protection officer;[63]
- The specific purposes of the processing for which the personal data are intended as well as the legal basis for the processing;[64]
- In cases where the data is acquired from a third party and thus not collected directly from the data subjects by the data controller, the categories of personal data that are processed and the source from which the personal data originate (notably whether it came from publicly accessible sources);[65]
- The recipients or categories of recipients of the personal data;[66]
- The period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period;[67]
- The existence of the data subjects' rights and of the right to withdraw consent at any time if the processing is based on consent;[68] and
- The right to lodge a complaint with a supervisory authority.[69]

Moreover, in the specific case where the public administration intends to make use of this data for individual decision-making based 'solely' on automated processes, for instance to fight social security infringements and tax fraud, it will have to inform the data subject about the existence of such processes.[70]

Turning back to the examples of tax and social security fraud, citizens are generally informed about the existence of data-matching and data-mining operations through the laws mentioned in Section 4.2.3. Yet, according to some authors, these laws do not provide sufficiently clear information to the citizens, notably in terms of the concrete processing that will be conducted and in terms of the types of data that will be used (Degrave 2020b; Degrave and Lachapelle, 2014). To some extent, this lack of transparency is reduced by the fact that these concrete data processing will be subject to a DAM fiche, to an authorisation from the President of the Executive Committee of the FPS Finance, to a prior deliberation of the Information Security Committee or to the conclusion of a data transfer protocol, which will provide more specific information (Section 4.2.3). However, citizens do not have access to the DAM fiches or to the authorisations of the President of the Executive Committee. Moreover, while the deliberations of the Information Security Committee are published on the website of the CBSS,[71] it is hard to obtain information about a specific processing operation, as the search tool is quite basic. In a similar vein, while the

[63] Arts. 13.1.a), 13.1.b), 14.1.a) and 14.1.b) of the GDPR.
[64] Arts. 13.1.c) and 14.1.c) of the GDPR.
[65] Arts. 14.1.d) and 14.2.f) of the GDPR.
[66] Arts. 13.1.e) and 14.1.e) of the GDPR.
[67] Arts. 13.2.a) and 14.2.a) of the GDPR.
[68] Arts. 13.2.b), 13.2.c), 14.2.c) and 14.2.d) of the GDPR.
[69] Arts. 13.2.d) and 14.2.e) of the GDPR.
[70] Arts. 13.1.f) and 14.2.g) of the GDPR. For more details on this obligation, see Chapter 5, Section 5.2.2.
[71] https://www.ksz-bcss.fgov.be/fr/deliberations-csi-list?term_node_tid_depth=51.

data transfer protocols have to be published on the websites of the relevant data controllers,[72] the result is that they are published on a wide variety of websites, whose quality can vary greatly, making it almost impossible for citizens to have a good overview of the types of processing that are done with their data (Degrave, 2020b). For transparency purposes, it would be preferable to centralise the publication of all of these protocols and deliberations in a single source, such as the Data Protection Authority's website, where it should be possible to search through them on the basis of several criteria, such as the types of purposes or of data concerned (Degrave, 2020a). A good example of this is the city of Amsterdam's 'Algorithm register'.[73]

Restrictions to the right to information: the 'SyRI' case example

However, it must be pointed out that, according to Article 23.1 of the GDPR, the right to information, like any other data subject's right, can be restricted by a Member State law when such a restriction respects the essence of the fundamental rights and freedoms and is a necessary and proportionate measure in a democratic society to safeguard an important objective of general public interest (De Raedt, 2017; European Data Protection Board, 2020b; Scarcella, 2019). These objectives are listed in Article 23.1, and, in the context of big data collection to feed AI applications used to fight social security infringements and tax fraud, it could be resorted to 'an important economic or financial interest of the (...) Member State, including monetary, budgetary and taxation matters, public health and social security'.[74]

This provision is in line with Article 8.2 of the European Convention on Human Rights[75] and with Article 52.1 of the Charter of Fundamental Rights of the European Union.[76] Therefore, the interpretation of these provisions by the Court of Justice of the European Union and the European Court of Human Rights, respectively, is perfectly transposable to the interpretation of Article 23 of the GDPR. Accordingly, the restriction of data subjects' right must be provided by law, must respect the essence of the restricted fundamental rights and freedoms, must be necessary and proportionate in a democratic society, and must safeguard an important objective of general public interest (Article 29 Working Party, 2016; Tombal, 2018). For instance, in the context of the fight against tax fraud, the right to information, as well as the other data subject rights, can be delayed, limited or excluded, with regard to the processing of personal data for which the FPS Finances is the data controller, to guarantee public interest objectives in the budgetary, monetary

[72] Art. 20.3 of the Law of 30 July 2018.
[73] https://algoritmeregister.amsterdam.nl/en/ai-register/
[74] Art. 23.1.e) of the GDPR.
[75] 'There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others'.
[76] Charter of Fundamental Rights of the European Union, OJ [2012] C 326/391. Art. 52.1: 'Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others.

and fiscal field.[77] The goal is to prevent a citizen suspected of committing tax fraud from using this information in order to prejudice the investigation and to escape a sentence (Degrave and Lachapelle, 2014).[78]

In this regard, it is highly relevant to present the recent decision (5 February 2020) of the Rechtbank Den Haag,[79] on the compatibility, with Article 8 of the European Convention on Human Rights, of the Dutch government's 'Systeem Risico Indicatie ('SyRI' – system risk indication), which is a legal instrument used to detect various forms of fraud, including social benefits, allowances, and tax fraud. SyRI is a technical infrastructure in which data can be linked and analysed in a secure environment, in order to generate a risk report, which means that a legal or natural person is deemed worthy of investigating with regard to possible fraud, unlawful use and non-compliance with legislation.[80] The instrument is applied at the request of government bodies or other bodies with a public function, who decide to collaborate and exchange data.

The technique underlying the application of SyRI was first used between 2003 and 2013 without a legal basis.[81] On 1 January 2014, the SUWI Act[82] was amended in order to enshrine in law the application of SyRI, and the conditions for the application of SyRI are detailed in the SUWI Decree.[83,84] More precisely, the legal basis for the processing of the necessary information in SyRI for the purpose of carrying out risk analyses is Section 65 of the SUWI Act,[85] and Article 5a.1, §3 of the SUWI Decree lists the categories of data that qualify for processing in SyRI, such as work data, tax data, social assistance benefit data, pension data, etc.[86]

The legality of the SyRI instrument has been challenged, in the Court of The Hague, by a coalition of civil society interest groups and two natural persons.[87] In substance, they claimed that the SyRI instrument breached Article 8 of the European Convention on Human Rights because it constituted an infringement of people's privacy, and that the SyRI legislation did not provide

[77] Arts. 11.1, al.1; 11/1.1, al.1; 11/2.1, al.1; and 11/3.1, al.1 of the Law of 3 August 2012.
[78] Projet de loi portant dispositions relatives aux traitements de données à caractère personnel réalisés par le Service public fédéral Finances dans le cadre de ses missions, 6 juillet 2012, *Doc. parl.*, Chambre, sess. ord., 2011-2012, no 53-2343/001, p. 11.
[79] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, ECLI:NL:RBDHA:2020:865 (ECLI:NL:RBDHA:2020:1878 for the English version).
[80] *Ibid.*, points 3.1 and 3.2.
[81] *Ibid.*, points 3.5 to 3.10.
[82] Wet van 9 oktober 2013 tot wijziging van de Wet structuur uitvoeringsorganisatie werk en inkomen en enige andere wetten in verband met fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekende zijnde gegevens, *Stb.*, 2013, p. 405.
[83] Besluit van 1 september 2014 tot wijziging van het Besluit SUWI in verband met regels voor fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekend zijnde gegevens met inzet van SyRI, *Stb.*, 2014, p. 320.
[84] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, points 4.1 to 4.30.
[85] *Ibid.*, point 4.8.
[86] *Ibid.*, point 4.17.
[87] *Ibid.*, point 2.1.

sufficient safeguards.[88] The Court thus assessed whether the SyRI legislation complies with Article 8.2 of the European Convention on Human Rights. This particular provision requires striking a fair balance between the interests of the community as a whole, which the legislation serves, and the right to privacy of the individuals affected by the legislation.[89]

In this regard, the Court outlined that the State has a special responsibility when applying new technologies, and that it must strike the right balance between the benefits that technologies such as SyRI bring, and the violation of the right to a private life that they might entail.[90] In the case at hand, the Court concluded that, in its current form, the SyRI legislation failed to comply with Article 8.2 of the European Convention on Human Rights, because the legislation does not strike the 'fair balance' required under the ECHR between the social interest the legislation serves and the violation of private life to which the legislation gives rise.[91] Indeed, taking into account the GDPR principles of transparency, purpose limitation and data minimisation, the Court ruled that the application of the SyRI legislation was especially deemed to be unlawful because it was insufficiently transparent and verifiable.[92]

Without entering into too much detail here, we will simply outline that the following circumstances were deemed relevant by the Court in reaching this conclusion: (1) the principle of transparency is insufficiently respected in the SyRI legislation because it does not provide information on which objective factual data can justifiably lead to the conclusion that there is an increased risk, but only provides a few examples of indicators that can indicate an increased risk and a potential hit;[93] (2) the SyRI legislation does not provide information on the functioning of the risk model, for instance the type of algorithms used in the model, nor does it provide information on the risk analysis method as applied by the inspection services;[94] (3) a data subject whose data were processed in SyRI, but which did not result in a risk report, will not be informed about this processing, and therefore cannot verify that her data was processed on correct grounds;[95] (4) because it is impossible to verify how the risk report has been generated and which steps it comprises, a data subject will not be able to defend herself against the fact that a risk report has been submitted about her;[96] and (5) because large amounts of data qualify for processing in SyRI and because, in a concrete SyRI project, the test of necessity is carried out by the separate participants in the project, with no independent assessment, the SyRI legislation therefore contains insufficient safeguards and breaches the principles of purpose limitation and data minimisation.[97]

[88] *Ibid.*, point 6.1.
[89] *Ibid.*, executive summary.
[90] *Ibid.*, point 6.6. See also ECtHR, *S. and Marper v. the United Kingdom*, 4 December 2008, req. n^{os} 30562/04 and 30566/04, §112.
[91] Rechtbank Den Haag, 5 februari 2020, Zaak n° C-09-550982-HA ZA 18-388, point 6.7.
[92] *Ibidem.*
[93] *Ibid.*, point 6.87.
[94] *Ibid.*, point 6.89.
[95] *Ibid.*, point 6.90.
[96] *Ibidem.*
[97] *Ibid.*, point 6.106.

4.2.6 The data subjects' right not to be subject to automated individual decision-making and big data

According to Article 22.1 of the GDPR, the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning her or similarly significantly affects her (Article 29 Working Party, 2018a). We will further address this right in our chapter dedicated to AI,[98] but we wish to point out here that, according to Rouvroy, achieving Article 22's aims in a big data world is 'both unrealistic and deeply paradoxical (...) especially when [big data analytics] involve self-learning algorithms' (Rouvroy, 2016). Indeed, as pointed out by Zarsky, there are several tensions between Article 22 of the GDPR and big data analytics:

> First, prohibiting automated analysis obviously undermines many of the big data practices (...). Second, even if one of the many exceptions to the prohibition on automation (...) is met, the specific disclosures (...) which call for enabling a human response to the machines' decisions are still required. To meet these disclosure obligations, big data processes must be conducted in a manner that would assure they are interpretable – i.e. they can be explained to the inquiring individual. Constantly meeting an 'interpretability' requirement might call upon those designing the automated processes to compromise some of the system's precision to enable the delivery of this form of detailed disclosure. Third, allowing human interjection would further encumber the automated process and slow down the innovative technologies they bring about (Zarsky, 2017).

4.3 Data storage – personal data protection and blockchain

Blockchain technology is a distributed ledger technology (DLT) 'that makes it possible to avoid the use of a trusted third party for transactions, and which is notably at the basis of Bitcoin' (Villani *et al.*, 2018). As pointed out by Finck, 'a distributed ledger can be described as a shared and synchronised digital database that is maintained by a consensus algorithm and stored on multiple nodes (computers)' (Finck, 2017). From a technical point of view, blockchains 'only designate the variants of DLT that record data in packages ('blocks') that are hashed ('chained') to another' (Finck, 2017). As Finck outlines, 'data is usually grouped into blocks that, upon reaching a certain size, are chained to the existing ledger through a hashing process. Through this process, data is chronologically ordered in a manner that makes it difficult to tamper with information without altering subsequent blocks' (Finck, 2017). Concretely, digital ledger technologies, such as blockchains 'rely on a two-step verification process with asymmetric encryption. Every user has a public key (...) that is shared with others to enable transactions. In addition, each user holds a private key (...) that must never be shared with others. Both keys have a mathematical relationship by virtue of which the private key can decrypt data that is encrypted through the public key. Public keys thus hide the identity of the individual unless they are linked to additional identifiers.

[98] See Chapter 5, Section 5.2.5.

> The nodes are the computers on which the ledger is stored. (...) In public and permission-less blockchains, anyone can entertain a node by downloading and running the relevant software. Some (but not all) nodes also function as 'miners', which aggregate transactions into candidate blocks and hash a new block to the chain on the basis of a predetermined consensus protocol (such as proof-of-work or proof-of-stake)' (Finck, 2017).

Similarly to big data and AI, blockchain, as a new technology, creates challenges in terms of compatibility with the GDPR. This is notable because the GDPR was built for a world where data is collected, stored and processed centrally, whereas blockchain technology decentralises each of these processes (Finck, 2017; Lyons *et al.*, 2018). Accordingly, some types of blockchains (notably those that are public and permissionless) may be in tension with the GDPR as the decentralised method of data storage and protection of blockchains cannot be easily reconciled with data protection mechanisms developed for centralised data silos (Finck, 2017). However, 'GDPR compliance is not about the technology, it is about how the technology is used' (Lyons *et al.*, 2018). Therefore, a creative legal interpretation of the GDPR and the implementation of technical solutions could reconcile some of these tensions (Finck, 2017; Lyons *et al.*, 2018).

Before diving into these potential tensions between blockchain technology and personal data protection, a difference must be made between two types of blockchains, namely public and permissionless blockchains on the one hand, and private and permissioned blockchains on the other. Public and permissionless blockchains are open-source and open-access, which means that anyone can download or design software to run nodes (Finck, 2017). Private and permissioned blockchains run on a private network (an intranet or a VPN) and a person must have been granted permission by an administrator in order to maintain a node (Finck, 2017). To take an example, the European Blockchain Services Infrastructure (EBSI) will be a public but permissioned network, as only pre-validated participating nodes will have writing, storing, processing and transmitting of (personal and non-personal) data on the ledger (European Commission, 2020). For those types of blockchains, it will be easier to comply with the GDPR (Lyons *et al.*, 2018). Public and permissionless blockchains, on the other hand, are the ones that create the most issues from a personal data protection perspective, because of their extremely distributed nature (Lyons *et al.*, 2018). Therefore, those are the ones we will focus on in the remainder of this section.

4.3.1 Two preliminary questions

Two preliminary questions must first be outlined, namely whether data stored on the blockchain are personal data and, if this is the case, who should be responsible for the application of the GDPR.

Two sets of data could potentially be considered as personal data, namely the data stored in the blocks themselves (we will refer to these as the 'block data'), and the public keys that are used to link a specific operation to an individual (Finck, 2017). Regarding the block data, personal data can either be stored in plain text, in an encrypted form or hashed (Finck, 2017). For the plain text data, it is obvious that this remains personal data. For the encrypted and hashed data,

the question is whether this constitutes pseudonymised data,[99] to which the GDPR applies, or anonymised data,[100] to which the GDPR does not apply. The difference between the two concepts is that it is impossible to link back to the data subject when their data has been anonymised, while this is possible with pseudonymised data. Yet, when data is encrypted, the data subject can still be reidentified through the use of the right encryption keys, so it must be considered as pseudonymised data subject to the GDPR (Finck, 2017; Lyons *et al.*, 2018). The same goes for hashed data (Article 29 Working Party, 2014; Lyons *et al.*, 2018), though some authors argue that this is debatable, and that it will, in fact, depend on the circumstances of the case and on the hashing technique that has been used (Lyons *et al.*, 2018). In any case, truly effective anonymisation is difficult to achieve at present (Franceschi *et al.*, 2018; Graef *et al.*, 2018; Wendehorst, 2017).

That being said, advanced cryptographic (such as zero-knowledge proofs) and data aggregation techniques could potentially lead to robust anonymisation in the future (Lyons *et al.*, 2018). For blockchain technology, this could notably be the case if 'personal data could be stored off-chain and merely linked to the blockchain through a hash pointer. In such a scenario, personal data is recorded in a referenced encrypted and modifiable database and not on the blockchain. Under this formula, no personal data is stored on-chain' (Finck, 2017). Storing personal data off-chain could indeed be an interesting avenue to explore, especially in light of the tensions that will be presented below, as a number of them could be solved through storing the personal data off-chain. However, this is an attempt to design GDPR compliant blockchains that limits the role of the chain, as it merely holds proof that the data, which is held on a private storage, is valid (Finck, 2017).

Regarding the public keys, as they are used for the pseudonymous identification of an individual's operations on the blockchain, they shall also be considered as personal data, as this individual can be reidentified by matching additional information to the key (Finck, 2017). Contrary to the block data, the public keys could not be stored off-chain, as they are necessary to the chain's functioning (Finck, 2017).

Since the block data and the public keys can be considered as personal data, this raises the question of who should be the data controller, i.e. who is accountable for the application of the GDPR. In the GDPR, the data controller is defined as 'the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data'.[101] For private and permissioned blockchains, it might not be too complicated to identify a central entity that can qualify as the controller, for example the

[99] Pseudonymisation means 'the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person' (Art. 4.5 of the GDPR).

[100] The ISO 29100 standard defines anonymisation as the 'process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party' (ISO 29100:2011, point 2.2, available at: https://www.iso.org/obp/ui/#iso:std:iso-iec:29100:ed-1:v1:en).

[101] Art. 4.7 of the GDPR.

administrator that grants permissions to maintain a node (Finck, 2017; Lyons *et al.*, 2018). Public and permissionless blockchains, on the other hand, are a decentralised peer-to-peer network of nodes, and consequently, this leads to uncertainties as to who should be considered as the data controller(s) (Lyons *et al.*, 2018). According to Finck, either none of the nodes can be defined as *the* data controller, as none of them determines the purposes and means of the processing, or, more likely, *every* node qualifies as a data controller, which means that each of them would be accountable for the application of the GDPR and would have to be able to answer to a data subject's request regarding one of its rights (Finck, 2017). The same conclusion is reached by the European Commission in its EBSI assessment, where it indicates that all the nodes should be qualified as 'joint data controllers for the transactional data that they to verify, store, and put on/off chain' (European Commission, 2020). Moreover, these decentralised models might make it difficult to draw clear boundaries between data controllers and mere data processors (European Commission, 2020). This clearly shows that the GDPR was built for a world where data is collected, stored and processed centrally, which clashes with the functioning of decentralised blockchains. This is especially problematic because 'nodes: (1) only see the encrypted or hashed version of the data; and (2) are unable to make any changes thereto. Nodes are thus decentralised entities that cannot respond to the tasks the GDPR requires of centralised agents' (Finck, 2017).

Another issue derives from the GDPR's territorial scope, as decentralised blockchains rely on nodes that are spread out across the world (Finck, 2017). Yet, the GDPR's scope of application is relatively wide, as it applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not;[102] and to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to the offering of goods or services to these data subjects, or to the monitoring of their behaviour.[103] In light of this wide territorial scope of application, the GDPR obligations may be binding for a lot of blockchain-based applications throughout the world, even if they only have an indirect link with the EU (Finck, 2017).

4.3.2 Tensions between blockchain technology and personal data protection

We will now turn to the potential tensions between blockchain technology and personal data protection. In this regard, we should clarify from the outset that our analysis starts from the assumption that the data controller uses the blockchain for a specified, explicit and legitimate purpose,[104] relies on a lawful basis of processing to do so (Lyons *et al.*, 2018),[105] and that the data subject has been informed about this.[106] Accordingly, we start from the assumption that the blockchain is simply a technical means to achieving that purpose.

[102] Art. 3.1 of the GDPR.

[103] Art. 3.2 of the GDPR.

[104] Art. 5.1.b) of the GDPR. See Section 4.2.3. Purpose limitation principle and big data.

[105] Art. 6.1 of the GDPR. See Section 4.2.1. Lawfulness and fairness of processing and big data.

[106] Arts. 12 to 14 of the GDPR. See Section 4.2.5. The data subjects' right to information and big data.

Data minimisation and storage limitation principle

According to the data minimisation principle, only the adequate, relevant and necessary data for the fulfilment of the specific purpose of processing shall be processed.[107] Yet, integral copies of the chain are stored on each node, which multiplies, rather than minimises, the personal data that is used (Finck, 2017). Additionality, in light of the storage limitation principle, data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.[108] This principle is inherently at odds with blockchain technology, as personal data that has been added to the chain will perpetually remain part of it, as the chain is an append-only database that continuously grows and expands (Finck, 2017; Lyons *et al.*, 2018). Because data will never be deleted from the blockchain, the chain might contain data that are no longer adequate, relevant and necessary for the purpose of processing for which they were originally processed, and might contain data that is stored for longer than necessary, thus contradicting these two key principles of the GDPR. It should however be outlined that, in its EBSI assessment, the European Commission indicated that 'it can be argued that the existence of public keys on blockchains, combined with necessary privacy enhancing mechanisms (PEMs), will fulfil the data minimisation requirements of the GDPR' (European Commission, 2020). The efficiency of these PEMs however remains to be demonstrated (Lyons *et al.*, 2018).

One way to circumvent this tension would be to store the block data off-chain, as, in this case, the chain simply contains a link, via a hash pointer, towards an off-chain database, where the personal data can more easily be minimised or deleted, in full compliance with the GDPR (Finck, 2017). Indeed, if the data is deleted off-chain, the link will remain on the chain, but it will point to nothing. However, whether this will actually work in practice remains to be seen.

Data accuracy principle and right to rectification

According to the data accuracy principle, the data controller must ensure that the personal data that it processes is accurate and that it is kept up to date.[109] This principle is complemented by the right to rectification, which grants the data subject the right to obtain, without undue delay, the rectification of inaccurate personal data concerning them, and the right to have incomplete personal data completed.[110] Once again, these principles are at odds with blockchain technology, in light of the latter's immutability (Finck, 2017; Lyons *et al.*, 2018). Regarding the right to have incomplete data completed, the GDPR does outline that this could be done by 'providing a supplementary statement' (Finck, 2017). Adding a new block with supplementary information might thus potentially satisfy this requirement, though the former block containing the incomplete

[107] Art. 5.1.c) of the GDPR. See Section 4.2.4. Data minimisation principle and big data.
[108] Art. 5.1.e) of the GDPR.
[109] Art. 5.1.d) of the GDPR.
[110] Art. 16 of the GDPR.

information will remain. If the request is to rectify data rather than to complete it, then adding a new block might not be sufficient (Finck, 2017). Alternatively, the block data could be stored off-chain, in order to facilitate its rectification and update.

Right of access

The data subjects' right of access provides that the data subject has the right to obtain, from the controller, the confirmation as to whether or not it processes personal data concerning him or her.[111] If it is the case, the controller will have to provide access to the data, as well as to the information listed in points (a) to (h) of Article 15.1 of the GDPR.[112] Moreover, the right of access provides that the data subject has a right to obtain a copy of the personal data that is processed by the controller.[113] In the context of blockchain technology, this right might be extremely difficult to implement for the various nodes of the chain, as they will not know exactly which data is stored on the chain because they will likely only have access to encrypted or hashed data, and will thus not be able to tell a specific data subject whether their data is being processed in the chain or to provide them with a copy of her data (Finck, 2017; Lyons *et al.*, 2018). Nevertheless, if the block data is stored off-chain, the controller can more easily identify if data concerning the data subject is processed, and provide them with a copy of the said data.

Right to erasure

Article 17 of the GDPR stipulates that the data subject shall have the right to obtain from the controller the erasure of personal data concerning them, without undue delay, if: '(a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed; (b) the data subject withdraws consent on which the processing is based (...) and where there is no other legal ground for the processing; (c) the data subject objects to the processing (...); (d) the personal data have been unlawfully processed; (e) the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject; (f) the personal data have been collected in relation to the offer of information society services [to a child]'.[114] Once again, this right is at odds with blockchain technology, in light of the latter's immutability (Finck, 2017; Lyons *et al.*, 2018). By definition, it is almost impossible to delete the data from the chain. Nevertheless, as pointed out by Finck, 'the precise meaning of 'erasure' is not defined in the GDPR, opening the door to other interpretations than absolute deletion' (Finck, 2017). For instance, the French CNIL (CNIL, 2018), indicates that the combination of encryption techniques and key destruction could potentially be considered as an erasure (Lyons *et al.*, 2018). Moreover, the German legislator accepts that, if the specific technical mean of storage makes it impossible to delete

[111] Art. 15 of the GDPR.
[112] For example: the purposes of the processing, the categories of personal data concerned, the recipients to whom the data will be or has been disclosed, the storage period, etc.
[113] Art. 15.3 of the GDPR.
[114] Art. 17.1 of the GDPR.

the data, erasure can be achieved through other means, such as limiting the processing that are tolerated,[115] e.g. the data can only remain stored but cannot be used, and this example might open the door to an interpretation of 'erasure' that accounts for the blockchain's immutability (Finck, 2017). However, we do not find this alternative compelling, as the GDPR provides for a distinct right to the restriction of the processing, which precisely aims at limiting the processing that can be performed on the data to its sole storage.[116] Accordingly, if the GDPR was meant to make erasure possible through restriction rather than through deletion, even if only in specific cases, two distinct rights would probably not have been created. That being said, data erasure could be facilitated by storing the block data off-chain.

Data protection by design and by default

According to the data protection by design principle, the controller will have to implement appropriate technical and organisational measures, both at the time of the determination of the means for processing and at the time of the processing itself, which are designed to implement data protection principles in an effective manner and to integrate the necessary safeguards into the processing.[117] For instance, pseudonymisation could be used in order to ensure the respect of the data minimisation principle.[118] Here, the encryption and hashing of the block data before including it in the chain might actually be considered as a form of privacy by design, and the objectives of the GDPR and of blockchain technology are thus aligned in this regard (Finck, 2017; Lyons *et al.*, 2018). This will especially be the case if the data is stored off-chain.

According to the data protection by default principle, the controller shall implement appropriate technical and organisational measures in order to ensure that, by default, only personal data which are necessary for each specific purpose of the processing are processed.[119] This echoes the data minimisation principle that has been presented above. By default, the amount of personal data collected, the extent of its processing, the period of its storage and accessibility should be reduced to the minimum necessary for the purpose of processing.[120] In terms of accessibility, this means that, by default, the personal data should only be made accessible to a limited number of people.[121] Yet, as mentioned above, in decentralised blockchains, integral copies of the chain are stored on each node, which multiplies, rather than minimises, the number of people that have access to the data (Finck, 2017). Additionality, because data will never be deleted from the blockchain, the chain might contain data that are no longer adequate, relevant and necessary for the purpose of processing for which they were originally processed (Finck, 2017).

[115] Article 35 of the Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680.
[116] Art. 18 of the GDPR.
[117] Art. 25.1 of the GDPR.
[118] Art. 25.1 of the GDPR.
[119] Art. 25.2 of the GDPR.
[120] Art. 25.2 of the GDPR.
[121] Art. 25.2 of the GDPR.

As outlined above, a way to address this issue would be to store the block data off-chain, where the personal data can more easily be minimised and where the access can be limited to specific people. Another solution would be to apply PETs, such as zero knowledge proofs (European Commission, 2020). The efficiency in practice of these PETs is however uncertain at the moment (Lyons *et al.*, 2018).

Transborder data flows

Finally, since the nodes of a chain are located all around the world, and each of them contains a full copy of the chain, the rules pertaining to the transfer of personal data to 'third countries', namely countries outside of the EU and the European Economic Space (EES), must be kept in mind (Finck, 2017; Lyons *et al.*, 2018). As a matter of principle, data can only be transferred to a third country if the European Commission has decided that it ensures an adequate level of protection.[122] As this list is quite limited, some of the nodes will very likely be located in third countries that are not deemed to grant adequate protection. In such cases, Article 46.1 of the GDPR provides that a controller may transfer personal data to a third country if it has provided appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available. This could notably be done through binding corporate rules or a code of conduct.[123] In theory, the nodes could be requested to adhere to one of these mechanisms in order to join the blockchain, but in light of the other tensions that we have highlighted, such mechanisms may, in fact, not be deemed to grant appropriate safeguards (Finck, 2017). A final solution would be to rely on the data subject's explicit consent for such a transfer,[124] but this consent may be very hard to get in practice (Finck, 2017).

4.3.3 Use of blockchain, by the public administration, to fight social security and tax fraud

In light of the above, it stems that the use, by public administration, of blockchain technology to fight social security infringements and tax fraud raises numerous challenges and potential tensions. The starting point for the public administration should be to determine whether they actually need blockchain technology, or whether other technical solutions could be used, as it should not be assumed that using blockchain will automatically be more secure and cheaper (Lyons *et al.*, 2018).

If a public administration wishes to rely on blockchain technology, it is recommended that they use a private and permissioned blockchain with a clearly identified controller, who will grant authorisations to run nodes of the blockchain, in order to limit the number of people that can

[122] Art. 45.1 of the GDPR. For a list of these countries, see https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en.
[123] Respectively arts. 46.2.b) and e) of the GDPR.
[124] Art. 49.1.a) of the GDPR.

get access to the block data (Lyons *et al.*, 2018). In this regard, authorisations should be solely granted to bodies, whether public or private, such as SMALS,[125] that have been authorised by a law to process such data in order to fight social security infringements and tax fraud (Section 4.2.1).

Moreover, the blockchain should, ideally, be designed in such a way that the block data, which qualifies as personal data,[126] should be encrypted, aggregated and stored off-chain (Lyons *et al.*, 2018). In this regard, 'it could be argued that blockchain networks should be used to store immutable proofs that certain data exists, rather than to store the data itself' (Lyons *et al.*, 2018). As we have seen above, this would facilitate the compliance with the data minimisation, storage limitation, data accuracy, and data protection by design and by default principles, and would facilitate the exercise of the data subjects' rights of rectification, access and erasure, all of which may not be possible to do if the block data was directly stored on the chain. However, it is uncertain whether this will actually be workable in practice.

Finally, it should be outlined that, depending on the circumstances of the case, blockchain technology might enhance transparency about the processing done by the data controllers, which are stored on the chain. Indeed, blocks in the chain contain not only 'block data' pertaining to the different processing that occurred (e.g. which data has been processed), but also headers that outline the identity of the data source and data recipient, as well as a timestamp (Finck, 2017). Accordingly, depending on how the blockchain is built (e.g. whether the data subject can have access to the block data and whether the data controller can link the block data to a specific data subject), the data subject could check the chain in order to obtain information about whether data concerning them has been processed, but also about the entities between which their data has been transferred, and at what moment this occurred. Thus, if information about the processing of citizens' data in order to fight social security infringement and tax fraud were to be stored on a blockchain, the citizen could check, depending on how the blockchain is built, which of their data has been processed in this regard, among which public administration the data has circulated and at what moment. Naturally, in light of the sensitive nature of the data at hand, and in light of the data protection by default principle, natural or legal persons other than the data subject and the relevant public administration should not be able to access the chain (private and permissioned blockchain).

An interesting example to mention in this regard is Estonia's use of blockchain technology to secure its residents' health records (Einaste, 2018).[127] Indeed, Estonia has set up a blockchain that stores all the log files that record the data-processing activities performed on its residents' health records (which are themselves not stored on the blockchain) (Einaste, 2018). Each access or each change to a patient's electronic records is recorded and timestamped, and this can be checked by the patients (Einaste, 2018).

[125] SMALS is a Belgian non-profit organisation whose objective is to support the Belgian Federal public administration regarding information management and related issues for integrated IT service provision (https://www.smals.be/fr). It is one of the Belgian nodes of the European Blockchain Services Infrastructure (EBSI) (see https://ebsi4be.eu/).
[126] See Section 4.3.1.
[127] See https://e-estonia.com/blockchain-healthcare-estonian-experience/.

4.4 Conclusions

As outlined throughout this chapter, public administrations are increasingly relying on new technologies, such as big data and blockchain, which increase their capability to process greater amounts of data, in order to provide public services and to support their decision-making. While this can lead to significant benefits, notably in terms of efficiency, for these administrations, it must not be overlooked that the use of such technologies could also have significant impacts on the lives of their citizens. Therefore, administrations need to comply with the legal framework that aims at limiting the uses they can make of such technologies, in order to circumscribe these impacts. In this regard, the aim of this chapter was precisely to outline how this could be done in practice.

In terms of key takeaways pertaining to the use of big data by public administrations, it should be remembered that the lawful basis underlying big data processing should ideally be a law.[128] Importantly, this law needs to meet certain standards in order to be sufficiently clear and predictable. Indeed, as a rule of thumb, any personal data processing must be fair, which implies that the laws on which this processing is based must be sufficiently explicit and understandable for citizens. Furthermore, the purpose of processing must be defined prior to the collection of the data, and data that has been collected for a specific purpose cannot be further processed for a purpose that does not fit within this initial purpose.[129] This requirement can be at odds with big data analytics if a large number of data are collected 'for the sake of it', without a clearly defined purpose. Accordingly, if public administrations wish to rely on big data analytics, they will need to clearly define the purposes of processing in advance. Linked to this is the data minimisation requirement, according to which only the adequate, relevant and necessary data for the fulfilment of the specific purpose of processing shall be processed.[130] This can also be at odds with big data analytics that rely on the processing of data coming from the largest possible number of sources, independently of their relevance. Therefore, public administrations willing to rely on big data analytics will also need to clarify, in advance, which data they will use and why these data are necessary to achieve the desired purpose. Finally, public administration relying on big data analytics to process their citizen's personal data must respect the latter's data subject rights, such as the right to information and the right not to be subject to automated individual decision-making.[131]

Regarding the use of blockchain, several key takeaways can also be formulated for public administration.[132] First, it is advised to use a private and permissioned blockchain with a clearly identified controller, who will grant authorisations to run nodes of the blockchain, in order to limit the number of people that can get access to the block data. Moreover, such authorisations should only be granted to bodies that have been authorised by a law to process such data. Second,

[128] See Section 4.2.1.
[129] See Section 4.2.3.
[130] See Section 4.2.4.
[131] See Sections 4.2.5 and 4.2.6.
[132] See Section 4.3.3.

the blockchain should, ideally, be designed in such a way that the personal data from the blocks are encrypted, aggregated and stored off-chain. Third, the blockchain should be designed in a way that enhances transparency about the processing done by the data controllers on the data stored on the chain. In this regard, the data subject should be able to check the chain in order to obtain information about whether data concerning them has been processed, but also about the entities between which their data has been transferred, and at what moment this occurred.

Finally, it must be outlined that while the analysis pertaining to big data has focussed on data collection and combination, and the analysis pertaining to blockchain has focussed on data storage, this does not mean that these technologies do not have an impact on the other stages of the data processing lifecycle. Indeed, these technologies are intertwined and influence each other. For instance, a blockchain based system can enable the sharing of personal data between public administration in order to complement other sources of data, which would allow these administrations to conduct big data analytics on a greater scale and scope of data. Furthermore, big data and blockchain technologies can contribute to the development of artificial intelligence techniques, by increasing the amount of available data on which the AI algorithms can be trained. Naturally, the use of artificial intelligence technologies by public administration must also comply with the rules of personal data protection, as will be analysed in Chapter 5.

References

Anonymous, 2021a. European Commission gives Belgium two months to restore independence to ODA. Available at: https://www.archyde.com/european-commission-gives-belgium-two-months-to-restore-independence-to-oda/.

Anonymous, 2021b. Mathieu Michel invites the federal parliament to assume its responsibilities in the APD file: 'A highly problematic situation'. Available at: https://www.archyde.com/mathieu-michel-invites-the-federal-parliament-to-assume-its-responsibilities-in-the-apd-file-a-highly-problematic-situation/.

Article 29 Working Party, 2010. Opinion 3/2010 on the principle of accountability. WP 173. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp173_en.pdf.

Article 29 Working Party, 2013. Opinion 03/2013 on purpose limitation. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

Article 29 Working Party, 2014. Opinion 05/2014 on Anonymisation Techniques. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

Article 29 Working Party, 2015. Annex – Health data an apps and devices (annex to a letter to the European Commission on the scope of health data in relation to lifestyle and well-being apps). Available at: https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf.

Article 29 Working Party, 2016. Working Document 01/2016 on the justification of interferences with the fundamental rights to privacy and data protection through surveillance measures when transferring personal data (European Essential Guarantees). Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2016/wp237_en.pdf.

Article 29 Working Party, 2018a. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Available at: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

Article 29 Working Party, 2018b. Guidelines on transparency under Regulation 2016/679. Available at: http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227.

Belgian Privacy Commission, 2018. Opinion n° 34/2018. Available at: https://www.autoriteprotectiondonnees.be/publications/avis-n-34-2018.pdf.

Chantillon, M., R. Kruk, A. Simonofski, T. Tombal, J. Crompvoets, C. de Terwangne, N. Habra, M. Snoeck, and B. Vanderose. 2017. FLEXPUB Public e-Service Strategy – Work package 2 – Baseline Measurement. KU Leuven Public Governance Institute, Leuven, Belgium.

CNIL, 2018. Blockchain et RGPD: quelles solutions pour un usage responsable en présence de données personnelles? Available at: https://www.cnil.fr/fr/blockchain-et-rgpd-quelles-solutions-pour-un-usage-responsable-en-presence-de-donnees-personnelles.

De Raedt, S. 2017. The impact of the GDPR for tax authorities. *Revue du droit des technologies de l'information* 66-67: 129-143.

Degrave, E. 2015. La protection de la vie privée dans l'e-gouvernement. In C. De Terwangne, and E. Degrave (eds.) *Vie privée et données à caractère personnel*. 963-1003. Politeia, Brussels, Belgium.

Degrave, E. 2020a. Le R.G.P.D., les lois belges et le secteur public: Les traitements de données dans l'administration en réseaux et l'Autorité de protection des données. *Anthémis* 195: 281-317.

Degrave, E. 2020b. The use of secret algorithms to combat social fraud in Belgium. *European Review of Digital Administration & Law* 1: 167-177.

Degrave, E. and A. Lachapelle. 2014. Le droit d'accès du contribuable à ses données à caractère personnel et la lutte contre la fraude fiscal. *Revue Générale du Contentieux Fiscal* 28: 322-335.

Einaste, T. 2018. Blockchain and healthcare: the Estonian experience. Available at: https://nortal.com/blog/blockchain-healthcare-estonia/.

European Commission, 2020. EBSI GDPR Assessment: Report on data protection within the EBSI version 1.0 infrastructure. Available at: https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITALEBSI/Legal+Assessment+Reports.

European Commission, 2021. June infringements package: key decisions. Available at https://ec.europa.eu/commission/presscorner/detail/en/inf_21_2743.

European Data Protection Board, 2020a. Guidelines 05/2020 on consent under Regulation 2016/679 (V.1.1). Available at: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf.

European Data Protection Board, 2020b. Guidelines 10/2020 on restrictions under Article 23 GDPR, Version 1.0. Available at: https://edpb.europa.eu/sites/default/files/consultation/edpb_guidelines_202010_article23_en.pdf.

Finck, M. 2017. Blockchains and data protection in the European Union. Max Planck Institute for Innovation & Competition Research Paper No. 18-01. https://doi.org/10.2139/ssrn.3080322

Franceschi, A., R. Schulze, M. Graziadei, O. Pollicino, F. Riente, S. Sica, and P. Sirena. 2018. Legal challenges of the changing role of personal and non-personal data in the data economy. In: De Franceschi, A. and Schulze, R. (eds.) *Digital revolution: data protection, smart products, blockchain technology and bitcoins challenges for law in practice*. 19-41. C.H. Beck Verlag, München, Germany. https://doi.org/10.17104/9783406759048-19

Graef, I., R. Gellert, and M. Husovec. 2018. Towards a holistic regulatory approach for the european data economy: why the illusive notion of non-personal data is counterproductive to data innovation. TILEC Discussion Paper No. 2018-028. https://doi.org/10.2139/ssrn.3256189

Hahn, I. 2021. Purpose limitation in the time of data power: is there a way forward? *European Data Protection Law Review* 7: 31-44.

High-Level Expert Group on Business-to-Government Data Sharing, 2020. Towards a European strategy on business-to-government data sharing for the public interest – Final Report. Available at: https://ec.europa.eu/digital-single-market/en/news/experts-say-privately-held-data-available-european-union-should-be-used-better-and-more.

Laloux, P. 2021. Vie privée: la Commission lance une procédure d'infraction au RGPD contre la Belgique. Available at https://plus.lesoir.be/376968/article/2021-06-08/vie-privee-la-commission-lance-une-procedure-dinfraction-au-rgpd-contre-la.

Lyons, T., L. Courcelas, and K. Timsit, 2018. Blockchain and the GDPR: report prepared by the European Union Blockchain Observatory and Forum. Available at: https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf.

Mayer-Schönberger, V., and Y. Padova, 2016. Regime change? Enabling big data through Europe's new Data Protection Regulation. *Columbia Science & Technology Law Review* XVII: 315-335.

Richter, H. 2020. The law and policy of government access to private sector data ('B2G data sharing'). Max Planck Institute for Innovation and Competition Research Paper No. 20-06. https://doi.org/10.2139/ssrn.3594109

Rocher, L., J. Hendrickx, and Y.-A. de Montjoye. 2019. Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications* 10: 3069. https://doi.org/10.1038/s41467-019-10933-3

Rouvroy, A. 2016. 'Of data and men': fundamental rights and freedoms in a world of big data. Bureau of the Consultative committee of the Convention for the protection of individuals with regard to automatic processing of personal data. Available at: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020.

Scarcella, L. 2019. Tax compliance and privacy rights in profiling and automated decision making. *Internet Policy Review* 8: 1-19.

Sweeney, L. 1997. Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine & Ethics* 25: 98-110.

Tombal, T. 2018. Les droits de la personne concernée dans le RGPD. In C. de Terwangne, and K. Rosier (eds.) *Le Règlement Général sur la Protection des Données (RGPD/GDPR): analyse approfondie.* 407-557. Larcier, Brussels, Belgium.

Villani, C., M. Schoenauer, Y. Bonnet, C. Berthet, A-C. Cornut, F. Levin, and B. Rondepierre. 2018. Donner un sens à l'intelligence artificielle: pour une stratégie nationale et européenne. Available at: https://hal.inria.fr/hal-01967551/document.

Wendehorst, C. 2017. Of elephants in the room and paper tigers: how to reconcile data protection and the data economy. In S. Lohsse, R. Schulze, and D. Staudenmayer (eds.) *Trading data in the digital economy: legal concepts and tools.* 327-355. Hart Publishing, Baden Baden, Germany.

Zarsky, T., 2017. Incompatible: the GDPR in the age of big data. *Seton Hall Law Review* 47: 995-1020.