Empowerment and Big Personal Data: from Portability to Personal Agency

Nicolas Anciaux, Célia Zolynski, Sébastien Chaudat, Riad Ladjel

Empowerment and Big Personal Data: from Portability to Personal Agency

N. Anciaux
Director of Research at INRIA, PETRUS team

Célia Zolynski
Professor at Sorbonne Law School, IRJS

S. Chaudat
PhD student at University of Versailles, DANTE lab.

R. Ladjel
Post doc student at INRIA and University of Versailles, PETRUS team, DAVID Lab.

Abstract

The place of individuals and the control of their data have emerged as central issues in the European data protection regulation. The "empowerment" of the individual has notably resulted in the recognition of a new prerogative for the individual: the right to the portability of personal data. The corollary of this new right is the design and deployment of technical platforms, commonly known as Personal Cloud, Personal Server or PIMS, allowing the individual to consolidate all his or her data in a single system managed under his or her control. On the strength of these technical and legal innovations, several questions arise: what forms of empowerment are targeted in practice? What are the appropriate conditions to guarantee the objective pursued? At the crossroads of these questions, one dimension appears to be insufficiently exploited: that of "agentivity". This article transposes this notion from the social sciences to the management of personal data, and opens up a new reading of the empowerment measures of Big Data functionalities on personal data.

Keywords

Personal Data, Portability Right, GDPR, Data Governance Act, PIMS, PDMS, Personal Agency, Mutual Trust

Introduction

While the world is being turned upside down by Artificial intelligence and the use of personal data, the place of individuals and control over their data have become central issues in the new European Data Protection Regulation that came into force in May 2018. The European Union’s intention with this regulation was to

empower individuals[1], which notably involved recognising a new prerogative: the right to personal data portability. Portability gives individuals the ability to extricate themselves from a captive ecosystem, and provides them with enhanced control over their personal data. According to the Article 29 Working Party, it should "re-balance the relationship between data subjects and data controllers"[2], and represents a new medium in the development of innovative and virtuous European economics around personal data. The corollary to this new right is the conception and implementation of technical platforms to "empower individuals by improving their right to self-determination regarding their personal data"[3], commonly known as PIMS[4]. These provide individuals with personal data management systems[5] to collate all their data in a single system – to be managed under their control. This gives rise to commercial structures such as Digi.me and Cozy Cloud, as well as governmental initiatives like Mydata.org[6] and Self-Data[7], supported by personal data protection agencies.

However, most analysts agree that the objectives of empowerment are only partially achieved today, with many barriers still to overcome. A recent CERRE Report[8] underlines that the way data portability rights can be exercised remains "minimal and far from ideal", due for instance to delays in processing data portability requests and a lack of standard models for retrieved data. The implementation of data portability still requires new mechanisms to "allow user's trust and controls on the procedures of right of data sharing"[9]. Needless to say, the obstacles are not merely technological, but also of a legal and economic nature. Recent publications suggest that portability rights need to be clarified to enable their most ambitious promises[10], to better adapt to the business model of the collaborative economy[11] and to quantify the expected gain for citizens' privacy (e.g., when using PIMS) whereas all one's personal data would be delegated to a provider anyway[12]. The European Commission is also conducting discussions along these lines: as part of its Data Strategy[13] and the recent Data governance Act[14], it supports the creation of personal data spaces, which implies strengthening the right to portability enshrined in the General Data Protection Regulation.

Considering the actual state of the regulation, one dimension underpinning the notion of empowerment appears insufficiently explored – that of "personal agency". The concept of personal agency is a product of

---

[1] Communication from the Commission to the European Parliament and the Council, *Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition – two years of application of the General Data Protection Regulation*, COM(2020) 254 final, p. 2

[2] Guidelines of 13 April 2017 of Article 29 Working Party on the right to data portability, WP242 rev.01, https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233

[3] The motto of the MyData movement, which unifies PIMS editors and organiations, see https://mydata.org/about/

[4] S. Abiteboul, B. André, D. Kaplan, "Managing your digital life with a Personal information management system", *Communications of the ACM* 2015, 58 (5), pp. 32-35

[5] Anciaux, N., Bonnet, P., Bouganim, L., Nguyen, B., Pucheral, P., Popa, I. S., & Scerri, G. (2019). Personal data management systems: The security and functionality standpoint. *Information Systems*, *80*, 13-35

[6] See https://mydata.org/about/

[7] See http://mesinfos.fing.org/english/

[8] Centre for regulation in Europe (CERRE). June 2020. Krämer, J., Senellart, P., de Streel, A. (2020). Making data portability more effective for the digital economy: economic implications and regulatory challenges. https://cerre.eu/sites/cerre/files/cerre_making_data_portability_more_effective_for_the_digital_economy_june2020.pdf

[9] Martinelli, S. (2019). Sharing data and privacy in the platform economy: the right to data portability and "porting rights". In Regulating New Technologies in Uncertain Times (pp. 133-152). TMC Asser Press, The Hague.

[10] See in particular the so-called "Fusing scenario", where data portability fosters the creation of platforms of interoperable services, in: De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L., & Sanchez, I. (2018). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review*, *34*(2), 193-203

[11] Drechsler, L. (2018, June). Practical Challenges to the Right to Data Portability in the Collaborative Economy. In *Collaborative Economy: Challenges and Opportunities, Proceedings of the 14th International Conference on Internet, Law & Politics. Universitat Oberta de Catalunya, Barcelona* (pp. 21-22)

[12] Urquhart, L., Sailaja, N., & McAuley, D. (2018). Realising the right to data portability for the domestic Internet of things. *Personal and Ubiquitous Computing*, *22*(2), 317-332

[13] *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, "A European strategy for data"*, COM(2020) 66 Final, 19 February 2020, p. 10

[14] Proposal for a Regulation on European data Governance (Data Governance Act), 25 November 2020, COM(2020) 767 final

social sciences and forms the basis of individual empowerment. It was coined by Martha Nussbaum and Amartya Sen[15], then further developed by Ruth Alsop[16] and Deepa Narayan[17]. Personal agency characterises individual empowerment through two key components: the individual's "ability to envisage options and make a choice"[18] and "the capacity to transform choices into desired actions"[19]. According to these authors, the concept of personal agency can be used to characterise various vectors of empowerment related to human development, poverty reduction and women's status improvement.

We argue in this paper that this concept can be transposed to personal data management, to offer a new reading of empowerment measures in this context. Hence, our point is to extend the statements made by Tim Berners-Lee, winner of the Turing Award, criticising the current situation of the Web and the monopolies it engenders in personal data management. Now working on a new PIMS solution, he uses the concept of *personal agency* as the key to empowering individuals[20] ("you will have far more personal agency over data").[21]

Based on its initial meaning, personal agency – transposed to the context of personal data management by individuals – could be said to rest on two pillars.

(1) The first aim is to enable all individuals to "**make decisions**" about their own data. On one hand, this requires a range of different options to be open to the individual, as promoted by portability which allows migration from one service to another. On the other hand, individuals should not only be able to give their consent and hence to access necessary information (e.g. in the general terms of use for services that process their data), but also to understand it (e.g. by adequately designing information to ensure educated consent). In other words, they should be capable of measuring the effects – and the risks – entailed by their decisions around the use of their data, especially by considering each party's responsibilities. Such are the required conditions to ensure informed decision-making.

(2) Each individual should be in a position to "**become an agent**" of the way their decisions are implemented, *i.e*. to be able to orchestrate how their data is processed and ensure that this complies with their decisions. Varying scales of control and safeguards can be argued for, bearing in mind that delegation of control is not always sufficient to secure personal agency[22]. In essence, the digital ecosystem where individuals with personal agency evolve would enable them to confidentially compute and deduce certain appropriate information from subsets of their personal data and to transmit this information to third parties of their choice, with means to proove to said third parties that the information was indeed produced by the individuals, using given computation codes applied to given personal data subsets. The underlying IT architecture and the levels of protection offered to individuals and third-party services with whom they interact should therefore be carefully considered to uphold personal agency.

This definition being established, our goal is to further investigate the meaning of personal agency and identify key conditions to empower individuals regarding Big Data features. Therefore, personal agency should be broken down according to the particular types of decision and use at stake. These may be divided into two

---

[15] Nobel-Prize winning economist A. Sen defines personal agency as a dimension of his capability approach ("Well-being, Personal Agency and Freedom: The Dewey Lectures 1984", *The Journal of Philosophy* 1985, vol. 82, p. 206)

[16] R. Alsop *et al.*, "Measuring Empowerment in Practice: Structuring Analysis and Framing Indicators", 2006

[17] D. Narayan *et al.*, "Measuring Empowerment: Cross-disciplinary Perspectives", 2005, p. 6: "*personal agency is defined by the capacity of actors to take purposeful action, a function of both individual and collective assets and capabilities* […]".
https://openknowledge.worldbank.org/bitstream/handle/10986/7441/344100PAPER0Me101Official0use0only1.pdf?sequence=1&isAllowed=y

[18] R. Alsop *et al.*, esp. p. 6

[19] R. Alsop *et al.*, esp. p. 3

[20] Open letter by Tim Berners-Lee, 23 Oct. 2018. See: https://inrupt.com/blog/one-small-step-for-the-web

[21] Other writers propounding control by individuals over their personal data also recommend exploring this avenue: C. Lazaro and D. Le Métayer, "Control over personal data: true remedy or fairy tale?", SCRIPTed, 12:1, 2015, INRIA Research Report vol. 8681. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2689223

[22] The notion of personal agency introduced by A. Sen, Ruth Alsop or Deepa Narayan includes the effective power of a person (or group) and direct control of the procedure through the means at its disposal (see: Alkire, S, 2008, Concepts and measures of agency). Other works (see: Bandura, 1989, 2000; Crocker & Robeyns, 2009) introduce the related concept of "proxy agency" to consider delegation to another individual or a third party system. For the sake of simplicity, we focus here on personal agency exercised directly by the individual, using the available legal and technical tools.

main groups: exclusively individual uses related to a single person's data (often referred to as *Personal Big Data*[23]) and collective uses by a community of individuals (called *Big Personal Data*[24]).

The first part of the paper is thus dedicated to personal agency in the Personal Big Data context. We first review data portability and its current implementation through PIMS. Then, we underline the strongest form of empowerment to be currently suggested. Finally, we introduce a new "*bilateral trust*" condition, which should be met for the individual to interact with personal agency with third parties.

In the second part of the article, we investigate the case of empowerment in a collective context, where Big Data functionalities would be provided to a community of citizens. Firstly, we describe the collective uses of personal data in the field of AI and review the current dominant scenarios, in order to conclude that they disregard personal agency. Secondly, we review alternative suggestions to provide for personal agency. Finally, we introduce a second necessary condition of trust driven by personal agency –*mutual trust*– and outline a preliminary proposal for a legal and technical construction on this basis. The last section summarises our main findings and concludes the paper.

I. Empowerment with personal agency for Personal Big Data

This part follows a two-phase development. In the first stage, we shall briefly review the history of data portability and its implementation using PIMS. Then, we identify two different levels of empowerment currently set out –weak and strong empowerment–, to examine the strongest form of empowerment as implemented by PIMS. In the second stage, we propose a new condition based on personal agency, to enable empowerment in the context of Personal Big Data. This new condition allows individuals to transmit calculated results (i.e., personal information resulting from a calculation) to third parties of their choice, rather than their detailed personal data. Indeed, the entropy of information of an aggregated result is much less informative than that of a complete personal data history. As a result, personal agency enhances the ability of individuals to further regulate the dissemination of their personal data. We illustrate this by an example and discuss two important questions to make it applicable, centred on architectural choices and the individual's liability.

A. Asserting individual empowerment: an overview of the current context

Data portability, as a salient new right in the GDPR[25], opens new legal and technological opportunities. Emerging from joint projects such as Blue Button (medical data) and Green Button (electricity consumption data) in the United States, MiData (related to energy, financial, telecommunications and retail data) in Great Britain or MesInfos in France, piloted by FING (a non-profit French think tank) and supported by CNIL (the French personal data protection agency), this allows citizens to download all or some of their data in a structured, commonly used and machine-readable[26] digital format.[27] Those projects might pave the way for empowering individuals at different scales. We propose to highlight existing PIMS in order to ascertain to what extent they empower individuals, and confront them to our vision of a developed empowerment based on personal agency.

---

[23] This terminology was first introduced in the Personal Information Management domain, in the context of Lifelogging technology. The precise definition of "Personal Big Data" introduced in this context can be found in Section 3.2 of the paper : Gurrin, C., Smeaton, A. F., & Doherty, A. R. (2014). Lifelogging: Personal Big Data. Foundations and trends in information retrieval, 8(1), 1-125

[24] The term "Big Personal Data" refers to a Big Data processing using the personal data of a large number of individuals, as opposed to the term "Big Personal Data" (where data of only one individual is involved, see the previous note). For a detailed definition of "Big Personal Data", we refer the reader to Section 2 of the paper: McDonagh, M. P. Data Protection in the Age of Big Data: The Challenges Posed by Big Personal Data.

[25] Not only as an extension of the right of access.

[26] On limits of the terms used: S. ELFERING, *Unlocking the Right to Data Portability: An Analysis of the Interface with the Sui Generis Database Right*, MILPC Studies vol. 38, spec. p. 21 and f.

[27] Provision 68 encourages data controllers « *to develop interoperable formats* » enabling data portability, without creating an obligation to adopt or maintain systems that are technically compatible. The final version of the GDPR went back on Amendment 111 of the Parliament; European Parliament, legislative resolution of 12 March 2014, COM(2012)0011, 2012/0011(COD), art. 15(2a). For the extension of this right to standardised format, see: European Commission, op. cit., COM(2020) 254 final, p. 8, and P. Jyrcys, C. Donewald, J. Globocnik, M. Lampinen, "My Data, My Terms: A Proposal for Personal Data Use Licences", *Harvard Journal of Law & Technology*, vol. 33, Digest Spring 2020, p. 9. See, also: CNIL, *Le corps, nouvel objet connecté*, *Cahiers IP* 2014, vol. 2, p. 23 et seq.

1. A brief history of PIMS: technical and legal frameworks

Regarding the technical framework, one of the earliest systems allowing individuals to manage their personal data under their exclusive control was introduced in the United States in 2008 by Eben Moglen, Professor at Columbia University. Called "FreedomBox"[28] this system uses personal servers like plug-computer (a low-cost mini-PC such as RaspberryPI) and free software to help individuals elude State control and keep social exchanges private.

The concept of a personal data server then emerged in academia, with proposals from INRIA[29] and MIT[30], possibly benefiting individuals (cross-analysis of personal data hosted in different data silos, quantified self-tracking), third parties (sharing results of personal data computations) or society as a whole (collaboration between groups of individuals).

Commercial proposals appeared from 2012 onwards (Meeco, Cozy Cloud, etc.) and the terminology shifted towards the concept of "personal cloud". These solutions include online offerings ("cloud") and are exclusively aimed at individuals ("personal"), who are given a "digital home" ("Welcome to your new digital home"[31] is the Cozy Cloud motto) with advanced capacities for quantified self-analysis ("Meeco's manifesto reads: "[...] What if you and I had the same power?").[32]

FING uses the term PIMS, or "personal information management systems", as a technical solution that integrates and applies big data processing to an individual's data for self-tracking purposes.[33] FING has also introduced the concept of "self data", defined as the exploitation of personal data by individuals for their own purposes.[34]

Lastly, Tim Berners-Lee, founder of the web, published an open letter[35] in September 2018 criticising the current state of the web and the monopolies it engenders in personal data management. With his own roadmap including personal data management techniques, the Turing Award winner has in turn launched a personal data management solution.[36] Tim Berners-Lee invokes the principle of "personal empowerment through data" ("data should empower you") but also uses the concept of personal agency ("you will have far more personal agency over data"),[37] which he believes is fundamental to the success of the next era of the web.

As a follow-up to these projects, reforms have been made in European and French law enshrining the right to personal data portability for data subjects,[38] with the intent to turn this new prerogative into an empowerment tool to adjust the balance of power between major service suppliers and their users[39]. Individuals can retrieve their data free of charge in an open-access, machine-readable format and can thus move from one operator to another without losing their track record. They can also take control and manage their data and its use

[28] Initiated by E. Moglen and G. Bdale. Presented at FOSDEM 2013. Foundation website: https://freedomboxfoundation.org/.

[29] T. Allard, N. Anciaux, L. Bouganim, Y. Guo, L. Le Folgoc, B. Nguyen, P. Pucheral, I. Ray, I. Ray et S. Yin, "Secure Personal Data Servers: a Vision Paper", proceedings of the international conference on Very Large Data Bases (VLDB), vol. 3, p. 25-35, 2010

[30] Y.-A. de Montjoye, E. Shmueli, S. S. Wang, A. S. Pentland, "OpenPDS: Protecting the Privacy of Metadata through SafeAnswers", *PLOS ONE* 2014, vol. 9(7)

[31] See the Cozy Cloud website: https://support.cozy.io/article/280-etape-3-bienvenue-dans-votre-domicile-numerique.

[32] See the Meeco manifesto: https://www.meeco.me/manifesto

[33] "La gestion de votre 'vie numérique' avec un système de gestion des informations personnelles", by S. Abiteboul (INRIA and ENS Cachan), B. André (Cozy Cloud) and D. Kaplan (FING and MesInfos), on the Le Monde *Blog Binaire*, 2014. http://binaire.blog.lemonde.fr/files/2014/07/personalInfoSystem.short_.fr_.3.pdf.

[34] Wikipedia French: https://fr.wikipedia.org/wiki/Self_Data

[35] Open letter by Tim Berners-Lee, dated 23 Oct. 2018: https://inrupt.com/blog/one-small-step-for-the-web.

[36] MIT Solid project: https://solid.mit.edu/

[37] Open letter by Tim Berners-Lee, op. cit.

[38] Art. 20 GDPR and Art. 39 of the French Data Protection Act [LIL] (as amended by Act no. 2018-493 of 20 June 2018) – Art. L. 224-42-3 et seq. of the Consumer Code, established by Act no. 2016-1321 of 7 Oct. 2016 on the digital Republic, subsequently repealed by the Act of 20 June 2018

[39] On the scope of regulation through data as a counterpower given to final users on a market, see: Autorité de la concurrence, AMF, Arafer, Arcep, CNIL, CRE, CSA, *Nouvelles modalités de régulation. La régulation par la donnée*, 8 juill. 2019, esp. p. 3, for further references, see ref. 54

themselves[40]. Portability consequently has become an instrument for "privacy by using"[41], a tool to learn about privacy protection mechanisms, encouraging individuals to reclaim their informational autonomy[42], as an essential part of digital literacy. In fact, this new right enables individuals to take back some control over their personal data, in two different ways: by both "receiv[ing] the personal data concerning him or her"[43] and having "the personal data transmitted directly from one controller to another"[44]. Yet, the strict scope of application of this right counteracts the idea of empowered individuals[45], fully able to control their personal data, even more in the Big Data era. This right is subject to various conditions narrowing its field of application[46]. Indeed, it can only be exercised for data that (i) the individual "provided to a controller"[47] on the basis of consent or contractual performance[48] and (ii) only if the processing is carried out by automated means. Therefore, portability appears considerably limited, concerning its legal scope[49] and material scope[50].

As this right has not yet reached its full potential, academics and institutions call for an enhanced regulation, some through competition law[51], others through data protection law[52]. For instance, the Commission underlines the "absence of technical tools and standards that make the exercise of their rights simple and not only burdensome" and aknowledges that true empowerment should not be limited to mere portability[53] as "switching of service providers", but also aim at "enabling data reuse in digital ecosystems".[54] The Commission therefore highlights the need to create a supportive environment for the development of these solutions, requiring a more progressive interpretation of the article 20 provisions. Recital 68 of the GDPR, which focuses on the sole transmission to another data processor[55], might, according to the European Commission, require to enhance data

[40]See C. Zolynski and M. Leroy, "La portabilité des données personnelles et non personnelles, ou comment penser une stratégie européenne de la donnée", *Légicom* 2018, p. 105, esp. p. 108 – See also, C. Berthet and C. Zolynski, "L'empouvoirement des citoyens de la République numérique : regards sur une réforme en construction", *RLDI* 2018, p. 60, esp. no. 9 et seq.

[41]A. Rallet, F. Rochelandet, C. Zolynski, "De la Privacy by Design à la Privacy by Using: Regards croisés droit/économie", *Réseaux* 2015, vol. 189(1), p. 15-46

[42] Article 29 Data Protection Working Party, *Guidelines on the right to data portability*, Adopted on 13 December 2016, as last revised and adopted on 5 April 2017, WP 242 rev. 01, footnote 1, p. 4

[43] Art. 20(1) of the GDPR

[44] Art. 20(2) of the GDPR

[45] H. Ursic, "The Failure of Control Rights in the Big Data Era – Does a Holistic Approach Offer a Solution ?", in M. Bakhoum, B. Gallogo Conde, M.-O. Mackernordt & G. Surblyte (Eds.), *Personal Data in Competition, Consumer Protection and IP Law – Towards a Holistic Approach ?*, Berlin Heidelberg, Springer, 2017, Available at SSRN: https://ssrn.com/abstract=3134745

[46] See : J. Belo, P. Macedo Alves, "The right to data portability: an in-depth look", *Journal of Data Protection & Privacy* 2018, vol. 2, 1, pp. 53-61

[47] For a suggested interpretation, see: P. De Hert, V. Papakonstantinou, G. Malgieri, L.Beslay, I. Sanchez, « The right to data portability in the GDPR. Towards user-centric interoperability of digital services », *Computer Law & Security Review* (2018), pp. 193-203, spec. p. 199

[48] Art. 20(1)(a) of the GDPR

[49] O. Tambou, *Manuel de droit européen de la protection des données à caractère personnel*, Bruylant, 2020, spec. 203, n°255 and further

[50] I. Graef, M. Husovec, N. Purtova, "Data portability and Data Control: Lessons for an Emerging Concept in EU Law", *German Law Journal*, 2018, Vol. 19, No. 06, spec. p. 1370 and f., J. Drexl, *Data Access and Control in the Era of Connected Devices, Study on Behalf of the European Consumer Organisation BEUC*, BEUC Study, Brussels, 2018, n°43, 110

[51] O. Lynksey, *The Foundations of EU Data Protection Law*, Oxford University Press, 2015, 265, as quoted in H. Ursic, 2017, EDPS, *Preliminary Opinion of the European Data Protection Supervisor, Privacy and competittiveness in the age of big data : The interplay between data protection, competition law and consumer protection in the Digital* Economy, March 2014, §26, p. 15

[52] COM(2020) 66 final, p. 21: "*Explore enhancing the portability right for individuals under Article 20 of the GDPR (…) (possibly as part of the Data Act in 2021)*", Centre for regulation in Europe (CERRE), June 2020. Krämer, J., Senellart, P., & de Streel, A. (2020). *Making data portability more effective for the digital economy: economic implications and regulatory challenges*, spec. n°6.2.1, p. 78

[53] H. Ursic, *op. cit.*

[54] *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, "A European strategy for data"*, COM(2020)66 Final, p.10

[55] Art. 20§3 of the GDPR

portability to give individuals "more control over who can access and use machine-generated data" such as "real-time data access and making machine-readable formats compulsory".[56] Consequently, the EU Commission expresses its willingness to possibly include this extension as part of the 2021 Data Act.

2. Current status: weak empowerment and strong empowerment

In its actual state, empowerment gained from exercising the right to portability seems to range from "weak empowerment", where individuals merely switch from one commercial service to another, to "strong empowerment", where individuals migrate to a personal data management service and thus regain sovereignty over their data. Legislators plainly conceived this prerogative as competition-focused[57], in the manner of the phone number portability advocated to force telecommunications operators to open up the market by lowering the exit barriers[58]. The right to data portability is therefore an instrument for extricating oneself from a captive ecosystem: it allows individuals to migrate from one operator to another without losing their data and without the drudgery of retrieving data from different systems.[59] It imparts service users with more freedom of choice, and could stimulate competition through innovation. Empowerment limited to this choice can be referred to as "weak". Empowerment may be characterised as "strong" when data recovery gives individuals an active role in the data lifecycle. It was in this context that personal cloud solutions were developed as a corollary to these new portability rights. They form the technical lever to exercise the right to portability. An individual's personal cloud has a range of connectors (to their bank, their employer and any external service that possesses their personal data) which lets them retrieve their personal data automatically.[60] With these offerings, they can combine all their data in a single system and adjust access in favour of innovative services.

Considering "strong empowerment" being the most advanced and promising version of data portability, we shall clarify its meaning in terms of features, in the context of PIMS. Recent reviews of personal cloud solutions,

---

[56] COM(2020) 66 final, p. 20

[57] Commission Staff Working Paper, *Impact Assessment, Accompanying the document Regulation of the European Parliament and of the Council on the protection of individuals to the processing of personal data and on the free movement of such data (General Data Protection Regulation) and Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data* (*COM(2012) 10 final, COM(2012)) 11 final*, Jan., 2012, SEC(2012) 72 final, *e.g.* at p. 28 stating that "*Portability is a key factor for effective competition, as evidenced in other market sectors, e.g. number portability in the telecom sector*.", Commissioner Joaquin Almunia, Speech, *Competition and personal data protection*, 26th, November 2012, available at: https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_12_860, European Commission, Staff Working Document on the free flow of data and emerging issues of the European data economy Accompanying the document Communication of the Commission, *Building a European Data Economy* (COM(2017) 9 final), SWD(2017) 2 final, spec. p. 47, Article 29 Data Protection Working Party, WP 242 rev. 01, p. 4, De Hert P., Papakonstantinou V., Malgieri G., Beslay L., Sanchez I., « The right to data portability in the GDPR. Towards user-centric interoperability of digital services », *Computer Law & Security Review* (2018), pp. 193-203, I. Graef, J. Verschakelen, P. Valcke, "Putting the right to data portability into a competition law perspective", *Law: The Journal of the Higher School of Economics, Annual Review* 2013, p. 53-63, Y. Poullet, "Is the general data protection regulation the solution?", *Computer Law &Security Review* 34 (2018), pp. 773-778, spec. p. 777, For a broader view on data over competition aspects: Autorité de la Concurrence, Bunderskartellamt, *Competition Law and Data*, 10th May, 2016, p. 11 and f.

[58] For an overview of mobile number portability and its effects on competition, see: B. Usero Sánchez, G. Asimakopoulos, "Regulation and competition in the European mobile communications industry: an examination of the implementation of mobile number portability", *Telecommunications Policy* 36 (2012), pp. 187-196, on cross-border portability of online content services, see: European Commission, DG for Communications Networks, Content and Technology, *Annual Activity Report. 2019*, 31st March, 2020, Ares(2020)1859706, p. 5 and further

[59] C. Zolynski and M. Leroy, *op. cit.*

[60] PIMS plateforms propose a set of connectors to easily retrieve personal data for the individual from many sources. Existing connectors can be used in personal cloud applications, and additional ones can be implemented by developers at will. See for instance Cozy Cloud documentation: https://docs.cozy.io/en/tutorials/konnector/ and Digi.Me presentation: https://digi.me/sources/

whether conducted from a social sciences perspective,[61] a technical one[62] or in experimental form,[63] unanimously agree on the intended purposes. The key point here is to re-establish the individuals' control over the lifecycle of their personal data[64], from collection to destruction, while enhancing the use by individuals of their own data.

In terms of features[65] , the first and foremost promise is to automatically reconstitute full personal records, which were originally stored in different data silos[66] (banking, medical history, internet search history, geolocation, social exchanges, etc.). The second key promise made to individuals is the cross-analysis of personal information, allowing them to benefit from the interconnection of personal records from different sources. For instance, a medical examination and its corresponding prescription can be automatically retrieved from bank records of the reimbursement for medication. As a third promise, cross-exploitation of individuals' data also allows them to derive statistical information and complex computed results from their records, in a quantified self-tracking perspective. For example, comparing medical data such as weight or cholesterol levels with physical activity or step counts allows them to monitor their health.

B. Personal agency as a determining condition of individual empowerment

While certain legal and technical conditions regarding data portability rights are met, such as personal cloud tools, the ability for individuals to perform Personal Big Data, thus achieving empowerment, raises a key question: what kind of personal agency do individuals hold to *implement* their decision? Ensuring that users have the capacity to act entails checking whether they are able to assume the new power granted to them, i.e. in this case, responsibility for making decisions relating to how their own data is managed (knowing who to share it with, hence making informed decisions), and the ability to orchestrate the implementation and effectiveness of their decisions (being able to contribute to implementation and to assume control over it). We shall introduce here a new trust condition that individuals must be able to establish to exercise their personal agency, that we call *bilateral trust*. It will be illustrated by Example 1, and open questions regarding the technical processing architecture and the legal liabilities of the individual in this context shall be discussed.

1. Personal agency in the context of Personal Big Data: a new trust condition

Assuming management of one's own data in terms of Personal Big Data with personal agency would presuppose a capacity to (i) administer and secure one's data, (ii) stipulate and apprehend permissions to different applications and third parties, and (iii) define which processing is authorised and set up safeguards to ensure one's decisions are effective. We thus argue that transposing personal agency to the Personal Big Data context would lead individuals to secure bilateral trust whilst the personal data underlying their decisions is processed. On one hand, individuals must be assured that their data is handled in line with the decisions made and that the Personal Big Data computation will indeed be implemented faithfully and confidentially (i.e. the expected code is executed, and the personal data provided in inputs is not exposed). On the other hand, third parties and external applications need a reciprocal guarantee from the individual, that the Personal Big Data processing results are indeed using the right datasets and are run as expected. This means being able to settle a two-sided trust guarantee, which we call *bilateral trust*. In Example 1, we illustrate this condition in the simple case of computing an energy bill based on a customer's energy consumption traces.

Example 1: Personal Big Data to enable citizens to compute energy bills.

[61] T. Lehtiniemi, "Personal Data Spaces: An Intervention in Surveillance Capitalism?" *Surveillance & Society* 2017, vol. 15(5), p. 626-639

[62] N. Anciaux *et al*., "Personal Data Management Systems: The security and functionality standpoint", *Information Systems* 2019, vol. 80, p. 13-35

[63] Pilote MesInfos 2016-2018. Synthèse/Enseignements/Actions. G. Jacquart, S. Medjek, M. Molins. Available at: mesinfos.fing.org/wp-content/uploads/2018/06/LivrableA5_Synthese-Enseignements-Actions_VF_Web.pdf.

[64] Note that while the concept presupposes exclusive control by individuals over their data, Personal Big Data does not assume that individuals "own" their data (see p. 23 of the FING paper, op. cit.). Available at: mesinfos.fing.org/wp-content/uploads/2018/06/LivrableA5_Synthese-Enseignements-Actions_VF_Web.pdf

[65] For more details on the three promises of personal cloud systems summarized here, we refer the interested reader to section 2 "*Existing personal cloud solutions*" of the above mentioned paper (N. Anciaux *et al*.).

[66] For instance: CNIL, *Vie privée à l'horizon 2020*, Cahiers 2012, vol. 1, p. 55, CNIL, *Le corps, nouvel objet connecté*, *Cahiers IP* 2014, vol. 2, p. 23 et seq, A. Poikola, K. Kuikkaniemi, H. Honko, *MyData – A Nordic Model for human-centered personal data management and processing*, available on: http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf

More and more citizens are concerned about feeding personal data to external services. When calculating an energy bill, the energy consumption traces generated by a smart metre, which reveal details of the individual's activity, are sent to the energy supplier who then calculates the bill and charges the customer.

Big Data for the citizen: an ability to interact with third parties without revealing personal data. The PIMS alternative ensures that "services move to the data" rather than personal data being sent to services. Citizens can thus exercise data portability to retrieve traces of consumption from their smart metre, and a computation code/app provided by their energy supplier is downloaded on their PIMS to produce the bill. The issue related to personal agency is that the energy supplier must trust the individual.

Bilateral trust as a necessary condition for "Personal agency". Personal agency aims to empower citizens to perform such Personal Big Data computations on their own. This requires making individuals capable of *bilateral trust*, by means of two main new capabilities:
(1) First, the ability to guarantee to the individual that their raw personal data remains confidential. To the extent that detailed energy consumption trails may reveal the individual's activity, this is a prerequisite to trigger adoption.
(2) Second, the capacity for the individual to undertake that the final result was indeed computed on the expected dataset (the provider must be sure that the data subject has not truncated their data to lower the bill) and used the expected code (the one furnished by the energy provider). This is an essential issue if the client is to be charged according to the result.
The provider may only have access to the aggregated energy consumption result (needed to charge the client), or be allowed to hold a finer degree of data (for instance, in case of billing error or dispute).

In light of this necessary *bilateral trust*, the technical and legal conditions in which the PIMS solutions are offered shall be analysed in order to determine which are likely to ensure personal agency. In other terms, attempting to assess the personal agency of service users implies ascertaining which party enjoys actual agency. From a technical perspective, one must assess who is trusting whom and thus who the administrator is, therefore questioning the processing architecture. From a legal standpoint, liability issues are raised, for example in the case of error or dispute. An additional concern is to make the proposal appropriate (and acceptable) in practice, while avoiding to overburden the individual. In the following, we discuss these open questions.

2. Conditions and framework for personal agency *:* a call for a new vision

From a technical perspective, one notable feature of the personal cloud is that the processing and applications of Personal Big Data "are moving" to the relevant data, as opposed to personal data which migrates toward remote services, as it occurs with most existing cloud services[67]. Individuals' personal agency can hence be measured by their capacity to implement this type of application under their exclusive control, in a digital ecosystem that allows them to build the desired reciprocal trust. Personal agency would therefore depend on architectural choices for personal clouds, i.e. the technical solutions implemented.

With centralised approaches (for example, MyDex.org or MyData.org), data administration and security is based on the personal cloud platform provider. This type of centralised management built on delegation technically allows secondary uses (beyond the individual's control) and exacerbates the risk of large-scale attacks (affecting millions of individuals). This requires strong trust[68] from individuals to the platform provider and all the personal applications running on the system. In this respect, the Data Governance Act proposes conditions to reinforce trust in the intermediaries ensuring data portability, to be used in the different data spaces.[69]

[67] D. Mula, "The Right to Data Portability and Cloud Computing Consumer Laws", *in Personal Data in Competition, Consumer Protection and Intellectual Property Law, Towards a Holistic Approach?*, Springer, MPI Studies on Intellectual Property and Competition Law, 28, 2018 pp. 397-409, spec. p. 398 and 399

[68] Anciaux, N., Bonnet, P., Bouganim, L., Nguyen, B., Pucheral, P., Popa, I. S., & Scerri, G. (2019). Personal data management systems: The security and functionality standpoint. *Information Systems*, *80*, 13-35. See in Section 2.1: "*Typically, data leakage resulting from attacks conducted against the personal cloud provider or the applications (which could be granted access to large subsets of raw personal data), or resulting from human errors, negligence or corruption of personal cloud employees and application developers, cannot be avoided in practice. This is critical because such solutions rely on a centralized cloud infrastructure settings which exacerbate the risk of exposing a large number of personal cloud owners, and hence may be subject to many sophisticated attacks.*"

[69] See Chapter 3 of the Data Governance Act, European Commission, COM(2020) 767 final

Self-hosting is a solution based on decentralised architecture, where each individual manages their own personal data on domestic hardware (for example, Di.Me,[70] CloudLocker, Cozy Cloud, Databox or Tim Berners-Lee's Solid). This gives individuals physical control over the platform which, if properly implemented, gives very high overall security (the cost-benefit ratio of an attack is dissuasive because any one attack only reveals a single individual's data). But responsibility for administering this system might befall individuals, with the attendant risk of error, loss or theft of personal data[71]. The DynDNS attack in late 2016, which infected non-secure embedded systems like printers and internet boxes, points out the vulnerability of self-hosted solutions.

Intermediate architectural solutions to these two extreme approaches can pave the way for different compromises according to the level of personal agency sought.

From a legal view, several questions must be addressed. First, how can the individual's ability to make informed decisions relating to the use of their personal data be ensured? On this point, the legal framework needs clarification, particularly as regards the expression of consent by the agent (the first condition of personal agency). One must ensure that users have the technical and cognitive capacities to make informed choices (second condition of personal agency).

Second, are individuals with personal agency therefore called upon to bear all responsibility when it comes to processing their own data with these Personal Big Data solutions? This raises concerns about excessive responsibility, leading to a potential "boomerang effect". The regulators also stress that users must be informed of the risks they run in taking over management of their own data, in that they lose access to the data security solutions offered by data controllers and take responsibility for the data.[72] This is all the more true as the liability regime established by the GDPR does not seem designed[73] to take into account the shift in perspective caused by these new individual data management solutions. Some are bound to criticise a potential elusion of liability by operators offering these new individual data management services, who might claim that their individual users should be qualified as data controllers[74]. Yet, the latter might benefit from the purely personal or household activity exemption, excluding the application of the GDPR[75]. In this case, the GDPR applies to the provider of the means for such processing[76]. The issue is that those means must be essential[77], and not only technical and organisational, to trigger the qualification of data controller[78]. In the case of personal data management systems, the provider might be qualified as data controller[79]. However, as he merely supplies the means for individuals to "compute" their personal data, he cannot be liable for all the obligations under the GDPR[80]. This qualification

[70] M. Sjöberg *et al.*, "Digital me: Controlling and making sense of my digital footprint", 5th International Workshop "Symbiotic Interaction", p. 155-167, 2017. http://hiit.github.io/dime-server/

[71] S. Abiteboul, B. André, D. Kaplan, "Managing your digital life with a Personal information management system", *Communications of the ACM* 2015, 58 (5), pp. 32-35

[72] See the G 29 Guidelines on the right to data portability, WP 242 rev. 01, 5 Apr. 2017, p. 22 et seq. https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233

[73] For instance, liability for the conception and determination of means of processing cannot weight on the shoulders of a user processing data outside the scope of the domestic exemption through a medium furnished by an intermediary

[74] See WP29, Opinion 5/2009 on online social networking, Adopted on 12 June 2009, WP 163, p. 5, A. Debet, « La Commission des clauses abusives et la protection des données personnelles sur les réseaux sociaux : une incursion hésitante dans un territoire inconnu », *Revue des contrats* 2015, n°3, p. 496, N. Metallinos, *in* A. Debet, J. Massot et N. Metallinos, *Informatique et libertés, La protection des données à caractère personnel en droit français et européen*, Lextenso, coll. Les Intégrales, 2015, spec. n° 557, p. 271

[75] GDPR, art. 2(2)(c)

[76] GDPR, recital 18

[77] Such as the data to be processed, the duration of processing, who might have the right to access this data… See: WP29, 1/2010, p. 14, EDPS, *Guidelines on the concepts of controller, processor and joint controllership under Regulation (EU) 2018/1725*, Nov. 2019, p. 10, EDPB, *Guidelines 07/2020 on the concepts of controller and processor in the GDPR*, version 1.0, adopted on 2 Sept. 2020, p. 14, §38

[78] WP 29, 1/2010, p. 14: "*determining the means would imply control only when the determination concerns the essential elements of the means*." See, N. Metallinos *in* A. Debet, J. Massot et N. Metallinos, *op. cit*. spec. n°468, p. 234 stating that services providers generally qualified as data processors might be qualified as data controllers when offering their services to individuals

[79] S. Abiteboul et al., (2015), where the PIMS provider acts on behalf of end-users, while providing security and informing them on options for security and privacy, *see also* EDPB, 07/2020, p. 20, §62 and f. on joint controllership for the provision and use of a tool, *otherwise*, see GDPR, recital 18

[80] EDPB, 07/2020, p. 9, §12

must reflect the exact involvement of the provider so as not to exceed his personal responsibility[81]. Since users of technical means might be qualified as joint controllers alongside the provider, as set out by the European Court of Justice [82], the issue is still pending for individual users of PIMS. Consequently, the entire liability chain between individuals, providers, suppliers and third-party services should be clarified in relation to the relevant architecture.[83] Beyond the agent, the liability of third parties involved in this ecosystem should be thought over, whether they provide tools or control applications, once it is agreed that technical liability cannot be laid exclusively at the door of the individual, albeit one who has sovereign control over their own data. It must be shared between actors in varying degrees, according to the architectures and each party's level of intervention in the use of the data. These are the conditions to ensure that "strong empowerment" does not ultimately lead to the exclusion of all safeguards provided for by the GDPR to protect individuals in data processing.

More generally, we can notice that whether weak (ability to migrate from one service to another) or strong (ability to conduct third-party processing of one's own data with bilateral guarantees), empowerment as promoted heretofore is undeniably a prerequisite in building a new approach ensuring individuals a degree of personal agency over their own data. However, it should be noted that the current purposes of PIMS solutions, as propounded by their editors and considered by academics or associations, essentially entail strictly individual benefits[84]. Moreover, the uses and technologies to create bilateral trust between individuals and a third party remain poorly implemented. Lastly, self-hosting solutions are limited to individuals who are sufficiently concerned about protecting their privacy to acquire the necessary expertise (to the point of playing sorcerer's apprentice) to install and administer their own system. Thus, Personal Big Data nowadays is aimed mainly at users interested in self-tracking and in cross-checking data for their sole benefit. These advantages remain insufficient to trigger widespread adoption. Some proposals have attempted to provide stronger incentives, although they are still based on individual interests. In many cases, these entail enabling individuals to "monetise" their personal data. For example, the start-up Embleema offers a personal cloud system based on blockchain, where individuals can collate their healthcare data from hospital and laboratory sites, from DMPs or connected objects (like Fitbit watches) to monetise access.[85] Over the medium term, the aim is to set up a marketplace for healthcare data, giving stakeholders access to "real-time" data on patients in return for payment. This reopens the debate on whether individuals "own their data" which opposes proponents of liberal analysis[86] and defenders of the right to informational autonomy as the ultimate guarantee of individual freedom in the digital age.[87]

For data empowerment to genuinely take off, we need to broaden the ambitions. Indeed, the power derived by the individual from their personal data remains limited as long as the ability to establish bilateral trust is not provided. This is a necessary (unsatisfied) condition of agency as demonstrated by the first part of this paper. Moreover, full power over personal data requires the exploitation of the data of many individuals (and not just one), following the example of the personal data exploitation deployed by the major web players which is based on using personal data of millions of individuals. Individual empowerment could be enhanced through a

---

[81] EDPS, *Guidelines on the concepts of controller, processor and joint controllership under Regulation (EU) 2018/1725*, Nov. 2019, p. 9, determination of a sole operation of the processing is enough to be qualified as data controller

[82] *See for instance*: CJEU, Gr. Ch., June 5th, 2018, C-210/16, Wirtschaftsakademie Schleswig-Holstein GmbH, CJEU, July 29th, 2019, C-40/17, Fashion ID GmbH & Co. KG. Summing up these cases: EDPB, 07/2020, p. 19, §§61-66

[83] In this respect, the new model enshrined by the GDPR is based on the accountability of all intervening actors throughout the data lifecycle.

[84] Anciaux, N., Bonnet, P., Bouganim, L., Nguyen, B., Pucheral, P., Popa, I. S., & Scerri, G. (2019). Personal data management systems: The security and functionality standpoint. *Information Systems*, *80*, 13-35. See the conclusion of their survey of PIMS solutions in Section 2, where the authors explicitly state that the distributed computations step (i.e., what we call "Big Personal Data" in this paper) is "currently poorly covered".

[85] See Embleema's "PatientTruth" solution: "Store Data, Share Records, Earn Tokens". (https://www.embleema.com/fr/patienttruth/).

[86] G. Noto la Diega, "Data as Digital Assets. The Case of Targeted Advertising", *in Personal Data in Competition, Consumer Protection and Intellectual Property Law, Towards a Holistic Approach?*, Springer, MPI Studies on Intellectual Property and Competition Law, 28, 2018, pp. 445-495, spec. p. 452

[87] On this debate, see the CNIL activity report for 2017, p. 52. https://www.cnil.fr/sites/default/files/atoms/files/cnil-38e_rapport_annuel_2017.pdf

form of collective empowerment[88], to secure social or societal progress surpassing purely self-centred benefits. The following part considers the conditions for collective agency.

II. Drafting collective empowerment based on personal agency

This part is organised in three sections. First, we will appraise the current schemes to access and process vast amounts of personal data (Big Personal Data), in order to conclude that these tend to disregard personal agency. In a second section, we shall analyse potential alternatives, whose aim is to facilitate the collective exercise of portability rights in a regulatory framework, with a nascent sense of personal agency. Finally, we argue that collective personal agency opens up to new uses of personal data, which could be defined by citizens themselves, but not only as suggested by pre-existing services or organizations.

A. A global race for collective uses: approaches devoid of personal agency

Cross-checking personal data among vast populations has both individual and social advantages in many areas such as healthcare, banking, smart cities, social assistance, etc. This collective use of personal data is based on computation methods referred to as Big Personal Data[89], that involve Big Data processing on the personal data of thousands or even millions of individuals. The processes underlying Big Personal Data combine techniques ranging from simple statistical analysis (grouping, aggregation) through automatic information search (automatic classification, rule discovery) to learning (based for instance on neural networks). As noted by the task force "AI for Humanity" in France, led by Cédric Villani, a combination of these techniques and their rapid growth have given rise to fierce competition in the global race for Artificial Intelligence.[90] With data now seen as a "major competitive advantage", "data sharing between private stakeholders has been identified as one of the main levers to catching up with American and Chinese stakeholders, who have the advantage of having access to massive amounts of data".[91] This explains the new ambition at the European level to access huge amounts of data "particularly from major stakeholders, who have a de facto monopoly on the collection of certain categories of data".[92] However, compared to Personal Big Data (see the first part above), Big Personal Data introduces a new difficulty: that of gathering data from large sets of individuals and carrying out the required processing. What are the contemplated scenarios and what is the situation regarding the data subjects' personal agency?

1. Identified models: B to G, B to B and G to B

Firstly, some advocate an "open model", which involves enshrining the wider concept of "data of general interest", a category of data established in France since the Law for a Digital Republic in 2016.[93] This type of model allows to move forward in opening up private sector data; concurrently, the European Commission has also envisaged arrangements to facilitate access to data held by private companies.[94]

[88] On this, see also Peugeot, "Brève histoire de l"empowerment : à la reconquête du sens politique", 3 Nov. 2015, http://www.internetactu.net/2015/11/13/breve-histoire-de-lempowerment-a-la-reconquete-du-sens-politique/

[89] In this part of the paper, we use the term "Big Personal Data" (a Big Data processing using the personal data of a large number of individuals) as opposed to the term "Big Personal Data" (where data of only one individual is involved, see note 58) used in the first part. For a detailed definition of "Big Personal Data", we refer the reader to Section 2 of the paper: McDonagh, M. P. Data Protection in the Age of Big Data: The Challenges Posed by Big Personal Data.

[90] C. Villani, "For a meaningfull Artificial Intelligence. Toward a French and European Strategy", March 2018, p. 25

[91] Villani report, op. cit. p. 27

[92] COM(2020) 66 final, p. 26 and f

[93] L. Cytermann, "Le partage des données, un enjeu d'intérêt général à l'ère de l'Intelligence artificielle", *Rev. aff. eur.* 2018, no. 1, p. 65

[94] "Building a European Data Economy", COM(2017) 09 final and "Towards a common European data space", COM(2018) 232 final. While the recently amended PSI Directive does not impose this opening up, the text nevertheless acknowledges that Member States remain responsible in their decision to apply the requirements of the directive to private companies, in particular those that provide services of general interest (Directive 2019/1024 of 20 June 2019).

From this perspective, the Villani task force recommended gradually opening up datasets from private operators "on a case-by-case basis" and according to the sector "for motives of general interest".[95] This could take place in two different ways:[96]

- the opening up of private data for general interest purposes in favour of public authorities (Business to Government, or "B to G") to help the development of public policies. For instance, mobility data inferred from flows of people or vehicles could be obtained from operators such as Orange, Waze and Uber and processed by the government, in particular to conduct research into reducing road traffic accidents;
- data sharing in favour of other economic stakeholders (Business to Business, or "B to B") for economic purposes such as innovation, research, the development of new services or AI or to boost competition. The banking sector is cited as an example, where Directive PSDP2 requires banking institutions to provide access to their clients' data to encourage the development of innovative businesses ("Fintech").

However, data sharing should be subject to certain conditions: in addition to compliance to the GDPR, the principle of proportionality needs to be respected and the relevant companies' interests must not be adversely affected – which presupposes protecting business secrecy and the possibility to monetise data – and this in turn prohibits to subject such access to compulsory gratuity "for trade between companies for which there would normally be a charge". The principle of transparency must also be respected.

Another form of opening up consists in giving the economic sector access to data currently held and managed by state actors (Government to Business, or "G to B"). For instance, the "Health Data Hub"[97] task force was set up in France to investigate the provision of healthcare datasets held by the State to economic stakeholders. The task force concluded that "healthcare data financed through social welfare is a communal heritage and recommended that "this data should be fully exploited for the benefit of the largest number of people" once they have been matched and documented with metadata to facilitate exploitation.[98] Respect for privacy is based on personal data pseudonymity (where data is stripped of any directly identifying information).[99] To ensure overall economic viability, the proposal also suggests that access to the "hub" could be "charged for private stakeholders in the form of a fixed subscription fee and a variable charge depending on usage".[100] The European Commission's Data Governance Act also moves in this direction to promote the circulation of data in B-to-B and G-to-B models.[101]

2. Limitations: models devoid of personal agency

Fears could be raised about these three data sharing models (B to G, B to B, G to B[102]). These differ mainly in terms of the public or private nature of the recipient entity, which is in charge of managing the vast amounts of collected personal data. However, most agree on a personal data management model operated beyond the data subjects' control. Some of these models are comparable to the current cloud solutions, criticised for the issues they raise in terms of security, privacy and informational asymmetry as regards the individuals involved. Thus, the sophistication and frequency of cyberattacks increases alongside the rising volume of data that could potentially be disclosed. There is a further risk of re-identification if the data anonymisation techniques are too weak to provide appropriate protection; this risk is proven as regards pseudonymity, which is no longer considered an adequate anonymisation technique.[103] Moreover, the potential for secondary uses which are inconsistent with the initial purposes depends exclusively on the trust placed in the centralising entity and on all its providers. None of these approaches however contemplates obtaining the data subjects' consent. Access to Big Personal Data therefore seems focused on simply transposing the existing controversial method of managing

[95] Villani report, op. cit. p. 34

[96] États généraux des nouvelles régulations du numérique, consultation document, 2018, p. 16, https://cnnumerique.fr/files/users/user192/Synthese_EGNUM.pdf

[97] M. Cuggia, D. Polton, G. Wainrib, S. Combes, "Health-data-hub, Mission de préfiguration", Oct. 2018, https://solidarites-sante.gouv.fr/IMG/pdf/181012_-_rapport_health_data_hub.pdf

[98] More broadly on data and value created by public sector, see: Commission, COM(2020) 66 final, p. 6

[99] Cuggia, M., Combes, S. (2019). "The French health data hub and the German medical informatics initiatives: two national projects to promote data sharing in healthcare", Yearbook of medical informatics, 28(1), 195

[100] M. Cuggia et al, *op. cit.* p. 4

[101] See Chapter 2 and Chapter 3 of the Data Governance Act, European Commission, COM(2020) 767 final

[102] As developed in the previous section

[103] For instance, a recent article shows that a handful of attributes containing demographic information is sufficient to unambiguously identify almost 100% of (American) individuals in any dataset: L. Rocher et al., "Estimating the success of re-identifications in incomplete datasets using generative models", Nature Communications, 10: 3069 (2019).

personal data, to the detriment of any form of personal agency for individuals, even in the case of sensitive healthcare data.

B. Alternatives ensuring a form of personal agency

Since the aforementioned scenarios are devoid of personal agency, we shall examine some promising alternative proposals under discussion.

1. Collective portability as a condition of personal agency

Alternative proposals providing individuals with the means to collectively control the use of their personal data are being encouraged, in line with the development of "civic portability". The Villani task force has indeed suggested extending portability rights from an individual to a collective prerogative, particularly as regards AI.[104] Thus, groups of citizens sharing common values and willing to act collectively (on the model of class actions), could exercise their portability rights and share their data with a public authority for a specific purpose, related to a public service mission. In the field of healthcare for instance, patients could make their medical data available to a research institute to improve the detection or treatment of a pathology. The objective here would be to enable the creation of new databases in favour of public services by allowing the free movement of data "under the exclusive control of citizens".[105]

There again, portability could act as the cornerstone of this initiative, giving each individual the capacity to consent to processing and even to secure collective processing, thereby upholding another form of agency. This invites reflection on possible collective portability and empowerment. Meanwhile, some authors focus on the definition of group privacy and data protection[106], notably by the expansion of data groups[107]. These emerging theories highlight the need for data protection law to broaden its scope of application, taking into account its collective aspect[108].

2. Employment law, data trusts and data altruism as first steps to ensure personal agency

Insofar as combining individual portability initiatives is insufficient to enable individuals to jointly orchestrate the uses resulting from the collection of their personal data, it seems that portability on its own cannot guarantee collective agency. One must therefore analyse how data subjects may be empowered to conduct collective processing under their control.

A first solution would be to incorporate collective "civic" portability within the existing legal framework. Thus, since the relevant personal data may result from an individual's labour (for example, data from Uber drivers), one could foresee overlapping analogies between employment law and data protection law. This gives rise to new ideas:[109] terms of use negotiated along the lines of collective agreements, a collective portability exercised within associations or trade unions[110]. Another solution could induce reconsidering the personal data governance model, given that "consent-based models of data governance fail to protect the public against privacy violations and the unethical collection and use of personal data".[111] Some authors have explored the implementation of new governance based on data trusts, and investigated other ways of regulating data usage. Such work however reached mixed conclusions, in that such control on data processing is ultimately

[104] Villani report, *op. cit.* p. 37

[105] *Ibid*. – See also CNIL, "La plateforme d'une ville. Les données personnelles au cœur de la fabrique de la smart city", *Cahiers IP* 2018, no. 5, p. 48

[106] B. Mittelstadt, "From Individual to Group Privacy in Big Data Analytics", *Philos. Technol.* 2017, 30, p. 475-494

[107] U. Pagallo, « The Group, the Private, and the Individual : a New Level of Data Protection ? », *in* L. Taylor, L. Floridi, B. van der Sloot (*Eds.*), *Group Privacy: New Challenges of Data Technologies*, Dordrecht, Springer, 2017

[108] See: N. Purtova, "Do Property Rights in Personal Data Make Sense after the Big Data Turn? Individual control and Transparency", *Tilburg Law School, Legal Studies Research Paper Series*, n°21/2017,, spec. p. 17 "personal data cannot be considered as concerning just an individual anymore; data processing resulting from a decision of one person will inevitably have spill-over effects on others (…)".

[109] L. Maurel and L. Aufrère, "Pour une protection sociale des données personnelles", 5 Feb. 2018, https://scinfolex.com/2018/02/05/pour-une-protection-sociale-des-donnees-personnelles

[110] See L. Taylor, L. Floridi, B. van der Sloot (*Eds.*), *Group Privacy: New Challenges of Data Technologies*, Dordrecht, Springer, 2017

[111] *Data Trusts. A new tool for data governance*, ElementAI and Nesta, 2018, p. 30

based on the trust individuals place in their fiduciaries.[112] The EU Data Governance Act promotes, on the one hand, services of data cooperatives[113], and on the other hands, a new status for data voluntarily made available by individuals for the common good to non-profit organizations with the emergence of new "data altruism organizations" subject to a EU controlled registration process and a common European consent form.[114]

These studies reinforce the argument that the individuals' personal agency in data processing is closely linked to the ensuing empowerment prospects.

Although this is a step in the right direction, for example regarding the respect of indiviudals' consent, there is still one step missing: citizens organized into groups should be able to define and compute themselves results of interest for the community, without revealing their personal data, and with a proof that the results was indeed computed faithfully (and therefore could be used to enforce the rights of individuals).

C. Towards strong empowerment safeguarding personal agency for Big Personal Data

This section addresses the ways in which a group of individuals may be enabled to implement and control all the effects of a Big Personal Data processing, and which legal and technical framework is to be promoted. We shall analyse two (extreme) scenarios and their perspectives on personal agency. We will then introduce a new notion of *mutual trust* as a component of personal agency in the context of Big Personal Data, as illustrated by an example. We shall then discuss the underlying issues, in order to establish a new technical and legal framework for personal agency in this context.

1. The issue of personal agency in collective computations

Big Personal Data processing by a large set of individuals can be led according to various technical scenarios, with different perspectives in terms of personal agency. There are two different methods: on one hand, the centralised approach, which consists in bringing all the data to one entity for processing; on the other hand, the decentralised approach, where each contributing individual is treated as an autonomous entity, capable of interacting with all the others to operate the processing together. Although many technical solutions exist halfway between these two extremes, analysis of the latter allows to identify different prospects regarding personal agency.

The first approach requires a centralised controller, governed by a third party entity, to administer the digital environment in which the computations are to be performed. The effect on the appropriate security measures is colossal since the benefits of an attack on this centralised entity are very high (access to the personal data of millions of individuals). In addition, the trustworthiness of the central entity is key to avoid secondary use of the data. Personal agency thus resides solely in the trust individuals consent to place in a third party entity.

The second approach does not introduce a centralised control point. Instead, each individual can be seen as a computation node that bears responsibility for part of the processing. Accordingly, their control of each node gives individuals a role as an agent of the computation agent. This approach however poses distinct risks to individuals. Firstly, by its very nature, computation implies an exchange of personal data between participants, thus transforming each of them into a potential attacker. Secondly, the external infrastructure supporting the data exchanges (for example, internet gateways) can observe some of these exchanges. Lastly, the data processed at each node and data exchanges between nodes can neither be defined nor even understood or administered by a non-expert individual (without a specific framework). Thus, this approach offers a new perspective on personal agency, but also presupposes the definition of a legal and technical framework that allows individuals to exercise their rights freely.

2. Personal agency for Big Personal Data: *mutual trust*

Elevating individuals to agents in terms of Big Personal Data consists in enabling them to decide (for example through consenting) whether to contribute to such a processing with their own personal data. It also means providing assurances to all data subjects involved that the processing is conducted in line with the stated purpose, with integrity and confidentiality. While personal agency relating to Personal Big Data establishes bilateral trust between an individual and a third party, each individual agent in Big Personal Data processings must be able to establish ***mutual trust*** between all participating individuals and the third party entity to which the

[112] See also: T. Hardjono, A. Pentland, "Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management", available at: https://arxiv.org/pdf/1905.08819.pdf

[113] See Chapter 3 of the Data Governance Act, European Commission, COM(2020) 767 final

[114] See Chapter 4 of the Data Governance Act, European Commission, COM(2020) 767 final

results are sent. On one hand, each individual must have a guarantee that their own data cannot be disclosed during processing and that all the other participants will act in a trustworthy manner and implement the processing as expected, in accordance with what each has been consented to. If one of the agents noticed a failure or violation, no result should ever be produced. Conversely, the recipient of the final result must be assured that it complies with the expected processing, based on the right dataset from the stated number of participants. We illustrate this mutual trust condition through Example 2, which stages a collective of parents computing statistics based on the online gaming data of their children.

Example 2: Big Personal Data to help parents reduce their children's addiction to video games.
More and more parents are worried about their children's addiction to online video games, such as free-to-play cooperative multiplayer "Battle Royale" games. These concerns are justified when video games companies have at their disposal petabytes of data used to feed Big Data algorithms to make the games as addictive as possible. Indeed, the main source of income for this category of games (free-to-play) are in-game purchases and events, which explains the publishers' willingness to maximise the timethat millions of users spend playing. Confronted to this issue, parents may feel powerless. They can either prevent their children from playing games at the risk of isolating them, or do nothing and let them sink into addiction.

Big Data for the citizen: an ability to explore 'anti-toxic' conditions. A reasonable solution would be to analyse the playing habits of children populations, to help determine the attitude parents could adopt when their child seems to develop an addiction. Thus, just as games editors use Big Data to quantify the impact of new game features on increasing children's playtime, parents should be empowered with Big Data means to collectively help defining better conditions to prevent children from being addicted.

Mutual trust as a necessary condition for "Personal agency". The notion of personal agency introduced in this article aims to empower willing parents to jointly define and perform Big Data computations for their collective benefit. Making parents "agents" of such collective computations requires providing them with *mutual trust*, by means of three new capabilities:
(1) First, the ability to ensure that the children's personal data will remain confidential. This is a prerequisite to convince parents to supply children's data in the computation and trigger broad adoption.
(2) Second, the capacity to attest that the final result was indeed computed on the expected data, with the agreed code and the appropriate number of participants. This is a necessary condition if the result is to serve as a basis for future decisions and recommendations in order to enforce the rights of the children.
(3) Third, the capability to ensure compliance with the legal basis for processing, the legitimate obtention of the relevant personal data through the exercise of the data portability right, informed consent, with clear statements concerning purpose, minimal personal data collection and no further use of personal data.

The Manifesto-based framework[115]: a new legal-technical solution operated in three phases.

Step 1: Formulate a hypothesis to be checked. Consider a collective (or association) of parents of young players aiming to reduce their children's playtime. For example, one could allow children to play more frequently but for a shorter time, limit the amount of games instead of the playtime, provide a fixed amount of money to be spent in the game (rather than let the children "win" it in the game), organise collective sessions (e.g., with remote classmates) rather than playing alone, block videos related to these games with parental control on Youtube, etc. Would some strategies overcome others in terms of reducing playtime in the long run?

***Step 2: Express a Manifesto for the collective computation*.** To test some of these options, the parents may express a Manifesto, which is both a set of rules describing the computation and a formulation of the legal basis for the considered processing. The manifesto can act as a contract, drafted between all the participants, giving their consent to the collection and processing of personal data, and to the random attribution of an 'agent' role in the computation, such as data collector or data aggregator. The obligations can be deferred until the realisation of a future condition (e.g., reaching the required threshold for the processing to start). This manifesto must be validated by a regulatory body (e.g. CNIL in France) which certifies its compliance with privacy laws. This certified manifesto is then published so that parents who wish to participate can download it and give their consent.

[115] For further technical details on new secure distributed execution framework based on secure hardware, we refer the interested reader to: Ladjel, R., Anciaux, N., Pucheral, P., & Scerri, G. (2019). Trustworthy distributed computations on personal data using trusted execution environments. *TrustCom/BigDataSE* 2019, pp. 381-388

Step 3: Execution of the Manifesto. This phase starts when the number of consenting parents reaches the threshold specified in the manifesto. Any participating parent is endowed with the aforementioned three capacities. From a technical standpoint, the condition to enforce these abilities in recent proposals[116] is that the participating parents' personal computer is equipped with a processor implementing 'trusted execution environments' in hardware (which is the case for recent computers endowed with Intel or AMD processors).

3. Rethinking a legal and technical framework to secure collective agency

To achieve a generalisation of Big Personal Data, a framework needs to be defined, firstly to support the essential elements of personal agency and secondly to avert the risks of privacy breach as well as damage to the integrity of the computation.

Among the relevant issues, the first one is whether the regulation forbids individuals to collectively use the personal data they have recovered pursuant to the exercise of their portability right. If the GDPR allows such Big Personal Data processing[117], the second issue is on its relevant legal basis. In our view, the choice of consent as a legal basis for inter-individual processing is the best option, insofar as consent is the only legal basis empowering and providing agency to individual. The conditions for consent to be lawful in this context supports this point: the data subject must be "offered control and (be) offered a genuine choice with regard to accepting or declining the terms offered or declining them without detriment"[118]. Consent is the only legal basis allowing individuals to have a granular control over which data is being processed; it may also be retracted at any point, along with the data processed under such legal basis. The regulation requires consent[119] to be freely given, specific, informed and unambiguous, by a clear affirmative act[120].

It could be expressed for example by means of a manifesto stipulating the type of Big Personal Data processing to be performed, where the purpose, the collected data, the computation code distributed to the participants, the result produced and the recipient entity are laid out, as well as the minimum number of participants required to achieve a useful result. Next, the manifesto would need to be verified and validated by a regulatory authority (such as the CNIL, the French data protection authority, or ANSSI, the French national cybersecurity agency). In addition to validating each case, the issue for the regulator is whether to draw up data collection clauses clarifying the types of algorithms implemented and the permissible output. Once approved, the manifesto would be published and made available to adequate groups of people, who could then decide whether or not to sign up. The regulator's endorsement would provide various guarantees ensuring respect for their personal data, securing their personal agency and, in time, could give rise to the drafting of a sectoral code of conduct to delineate the responsibilities and undertakings of stakeholders in Big Personal Data. Finally, a secure mechanism to ensure the agency of participants should be able to allocate the processing across all participants and execute it, without deviating from the manifesto or revealing any data other than the final result, thus safeguarding the mutual trust outlined above.

This could be considered as a realistic objective in the current state of technology. The conventional computation techniques used in business systems cannot be applied here due to the unusual scale of distribution (the computation could in theory encompass a fraction of the population of a country). Some secure distributed computation protocols based on cryptographic techniques (called "secure multiparty computation") could be used in some cases but cannot yet perform satisfactorily if extended to support generic computations for a large number of participants.

However, new technologies are currently being developed and use trusted computing hardware – which one is usually already equipped with – to set up generic secure distributed processing on a large scale. Most smartphones and PCs belonging to individuals now have secure processors such as Intel SGX, ARM Trustzone,

[116] Such as: Ladjel, R., Anciaux, N., Pucheral, P., & Scerri, G. (2019). Trustworthy distributed computations on personal data using trusted execution environments. *TrustCom/BigDataSE* 2019, pp. 381-388

[117] J. Belo, P. Macedo Alves, "The right to data portability: an in-depth look", *Journal of Data Protection & Privacy* 2018, vol. 2, 1, 53-61, spec. p. 55

[118] WP29, *Guidelines on consent under Regulation 2016/679*, As last revised and Adopted on 10 April 2018, WP 259 rev.01

[119] GDPR, art. 6(1)(a)

[120] GDPR, recital. 32

AMD PSP, etc. A recent study[121] shows that the concept of mutual trust as defined above is compatible with these types of hardware.

Conclusion

In this paper, we have showed that the notion of personal agency, as set out by social sciences, can be transposed to the case of personal data processing in the Big Data context. We provided a general definition of personal agency, which offers a new angle to analyse the proposed approaches for personal data processing operations.

A first interpretation of this definition in the context of Personal Big Data allows individuals to transmit calculated results (i.e., personal information resulting from a calculation) to third parties of their choice, rather than their detailed personal data. Indeed, the entropy of information of an aggregated result (e.g., the total amount of electricity consumed in the last month) is much less informative than that of a complete personal data history (e.g., consumption values produced on a power line reveal the electrical appliances used at each moment[122] and can therefore be used to infer precise daily behavior of household residents). As a result, personal agency enhances the ability of individuals to further regulate the dissemination of their personal data.

As a second interpretation, we have shown that personal agency opens up to new collective uses of personal data, where citizens organized into groups, are able to confidentially and compliantely calculate and deduce information from the group's personal data, with the possibility of proving that the result was obtained on given sets of personal data of the participants, using appropriate code, faithfully executed.

In both situations, this leads to the formulation of new necessary conditions related to the degree of trust that individuals must be able to provide to each other to be considered as "agents" of the processing. In the case of "Personal Big Data", a *bilateral trust* condition must be established. In the case of "Big personal data", a condition of *mutual trust* is required. We therefore outlined a preliminary proposal for a legal-technical co-construction illustrated by examples, which reflect the feasibility of these conditions in the current state of technology, and discuss related challenges which remain to be addressed.

Of course, many details will still need to be ironed out. Some technical building blocks demonstrating the feasibility of such a solution remain to be established. Many other open-ended questions may be raised: should user consent be set up for each processing operation or for a group of processing operations? How can it be formally demonstrated that one can withstand a small number of users who tamper with their hardware to attack security features – though the hardware security technologies are difficult to attack, vulnerabilities can always arise in an environment where security amounts to a race between hackers and manufacturers? How should the mechanisms described be integrated in a real operating system, and more particularly within existing PIMS products? Combining the circulation of huge amounts of data with informational sovereignty for each individual means that they should be seen as agents of the ecosystem currently being set up. A new structure therefore needs to be built to ensure full personal agency, with underlying mutual guarantees for individuals and for the entire ecosystem in which they operate. The terms still need to be adjusted but this new way must be explored to avoid individuals to be seen as mere datafied objects in the future.[123]

[121] Ladjel, R., Anciaux, N., Pucheral, P., & Scerri, G. (2019). Trustworthy distributed computations on personal data using trusted execution environments. *TrustCom/BigDataSE* 2019, pp. 381-388

[122] See for example, Figure 3 p.15 in : Quinn, E. L. (2009). Privacy and the new energy infrastructure. Accessible at : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1370731

[123] This contribution is the fruit of interdisciplinary discussions conducted as part of the *GDP-ERE* project financed by *Data IA* Convergence Institute of the University of Paris Saclay and the ANR *Perso Cloud*.