European perspectives on big data applied to health: The case of biobanks and human databases

Itziar de Lecuona, http://orcid.org/0000-0002-5081-5756
María Villalobos-Quesada, http://orcid.org/0000-0003-4930-1982

ABSTRACT

The paradigm shift to a knowledge-based economy has incremented the use of personal information applied to health-related activities, such as biomedical research, innovation, and commercial initiatives. The convergence of science, technology, communication and data technologies has given rise to the application of big data to health; for example through eHealth, human databases and biobanks.

In light of these changes, we enquire about the value of personal data and its appropriate use. In order to illustrate the complex ground on which big data applied to health develops, we analyse the current situation of the European Union and two cases: the Catalan VISC+/PADRIS and the UK Biobank, as perspectives.

This manuscript advocates for stopping the unjustified accumulation and commercialisation of personal data, protecting the interests of citizens and building appropriate frameworks to govern big data projects for health. A core tool for achieving such goals is to develop consent mechanisms which allow truly informed but adaptable consent, conjugated with the engagement of donors, participants and society.

KEYWORDS: big data, knowledge-based economy, eHealth, human databases, biobanks, data protection.

1. INTRODUCTION

eHealth can be defined as 'the application of information and communications technologies across the whole range of functions that affect the health sector'.[1] The technologies applied to eHealth are quite broad, but big data is an element of great significance. Big data can be defined as the acquisition, management and use of massive digital data, which involves multiple disciplines (such as computer engineering, informatics and statistics) and is supported by computing technology

[1] Commission of the European Communities. (2004). e-Health – Making Healthcare Better for European Citizens: An Action Plan for a European e-Health Area. Available at: https://ec.europa.eu/digital-single-market/en/news/e-health-making-healthcare-better-european-citizens-action-plan-european-e-health-area [Accessed 26 Feb 2018]

(clusters, clouds, servers, etc.). Therefore, big data depends on software and hardware components. The magnitude of the data used in big data are often of terabytes or petabytes in magnitude. To put this in context, a terabyte equates to the capacity of approximately 1500 CDs, while a petabyte is the capacity of 1.5 million CDs. Big data is also considered to have high variety, high variability, high velocity and complexity.[2] The generation, storage, management and analysis of such incredible amounts of information is accompanied by theoretical and technological advances in areas such as electronic and computer engineering, informatics, mathematics and physics.

Data in the context of big data has no, or very low, intrinsic value. The true value is generated from a process called 'big data to knowledge' (BD2K)[3] that refers to the transformation of disconnected to analysed data that can inform decision-making. The analysis, based on mathematic algorithms, reveals correlations such as patterns related to medical condition, treatment, age, residence, gender, ethnicity, income and education.

An important issue regarding the use of big data is the risk of discrimination. By assigning a person to a group of particular characteristics, an individual becomes a component of a collective, which raises concerns about conscious and unconscious discrimination as a result of the use of big data in decision-making.[4] Several cases of discrimination or biases have been reported, for example regarding crime prediction algorithms and internet-based searches.[5] The risk of discrimination is even more pressing when decision-making becomes automated, which is key in the formulation of autonomous systems.[6]

A second focal point regarding the use of big data is the protection of privacy. The process of BD2K enables third parties to infer personal information from individuals who have never granted consent. In the hands of a competent party, personal information can be used by third parties in ways that may affect our decisions, the

[2] Gandomi A, Haider M. (2015). Beyond the Hype: Big Data Concepts, Methods, and Analytics. International Journal of Information Management. 35, 137–144.

[3] Margolis R, et al. (2014). The National Institutes of Health's Big Data to Knowledge (BD2K) Initiative: Capitalizing on Biomedical Big Data. Journal of the American Medical Informatics Association. 21(6), 957–958.

[4] Barocas S. (2014). Data Mining and the Discourse on Discrimination. Proceedings of the Data Ethics KDD Workshop: Conference on Knowledge Discovery and Data Mining, New York. Available at: https://dataethics.github.io/proceedings/DataMiningandtheDiscourseOnDiscrimination.pdf [Accessed 26 Feb 2018]

[5] Barocas S, Selbst AD. (2016). Big Data's Disparate Impact. California Law Review. 104, 671–732.

[6] Note: Examples of autonomous systems are: smart cities, driverless cars and autonomous health systems.

perception of ourselves and our social interactions, and might limit our access to health services, education, economic opportunities and the professional market, among others. For these reasons, the implementation of decisions based on big data analytics needs to follow a careful assessment of the individual and social implications; that is, an evaluation of micro and macro scenarios of big data.[7]

Big data has become highly valuable for health-related activities and biomedical research.[8] Recently big data has incorporated the use of new information and communication technologies and wearable technologies (mHealth),[9] which are the basis of the Internet of Things (IoT).[10] Through these technologies it is possible to extract data from everyday life, such as diet, alcohol consumption, location and vital signs, which have been integrated into a holistic concept of health, which rejects genetic reductionism and advocates for prevention and prolongation of health. Terms such as 'P4 medicine'[11] and 'systems medicine'[12] have been coined to refer to this integration. Here we focus on the role of big data applied to health in the EU. With that in mind, we will explore the challenges and efforts to regulate the use of personal information at the heart of big data activities, with specific reference to electronic medical records, human databases and biobanks. The cases of the UK Biobank and the Catalan PADRIS will be analysed as two independent European perspectives. Although both cases are coherent with the use of health-related data for research and innovation, the enactment of both proposals is very different, illustrating different approaches to secondary use of personal data and in the case of the UK Biobank, it will eventually lie outside EU jurisdiction.

[7] Mantelero A. (2016). Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection. Computer Law & Security Review. 32, 238–255.

[8] Gené Badia J, Gallo de Puelles P, de Lecuona I. (2018). Big data and data security. In Spanish. Atención Primaria. 50, 3–5.

[9] Note: Some examples of mHealth already released to the market are: motion trackers, devices that measure vital signs and body-worn smart clothing. They have a wide range of applications, such as: wellbeing (stress, sleep, exercise, meditation and pregnancy), health (blood pressure, weight, heart activity, body temperature, sugar levels, fertility, rehabilitation and pain management) and beauty (health of skin and hair).

[10] Bhatt Y, Bhatt C. (2017). Internet of Things in Healthcare. In C. Bhatt, N. Nilanjan, & A.S. Ashour (Ed.), Internet of Things and Big Data Technologies for Next Generation Healthcare (p. 13–33). Studies in Big Data (23). Cham, Springer International Publishing.

[11] Hood L, Flores M. (2012). A Personal View on Systems Medicine and the Emergence of Proactive P4 Medicine: Predictive, Preventive, Personalized and Participatory. New Biotechnology. 29(6), 613–624.

[12] Kirschner M. (2016). Systems Medicine: Sketching the Landscape. In U. Schmitz, O. Wolkenhauer (Ed.), Systems Medicine. Methods in Molecular Biology (p. 3–15). New York, NY: Humana Press.

2. THE CHALLENGE OF eHEALTH IN THE DATA MARKET

The EU has faced the challenge to balance the promotion of the data-driven economy agenda[13] and the development of frameworks to regulate the new digital business models,[14,15] in order to guarantee an economic development respectful of individual rights.

Controlling the European data market, which could amount to EUR 106 billion by 2020,[16] is a complex task that must be carried out at multiple levels. First of all, individuals must be well informed and understand the value of their data. Secondly, at an institutional level, bioethics committees have the challenge to act as well-informed advisors in a context of constant innovation.[17,18] Thirdly, researchers, developers and their institutions will have to integrate the concept of Responsible Research and Innovation (RRI), where governance, ethics, open access, public engagement and scientific education must be central to the creation and application of knowledge.[19] Finally, data protection, privacy, transparency and accountability must be rooted in local and international legislation.

The commercialisation of personal data has led to the recognition of the right of subjects concerning their personal information. Important milestones were marked in 2014, when the right of individuals to data protection was acknowledged as superior to the

[13] European Commission. (2017, January 10). Commission Outlines Next Steps Towards a European Data Economy. Press Release. Available at: https://ec.europa.eu/digital-single-market/en/news/commission-outlines-next-steps-towards-european-data-economy [Accessed 10 Jan 2018]

[14] EU Health Programme. (2014). Overview of the National Laws on Electronic Health Records in the EU Member States and their Interaction with the Provision of Cross-Border eHealth Services. European Commission, Milieu Ltd. & Time.lex. Available at: https://ec.europa.eu/health/ehealth/projects/nationallaws_electronichealthrecords_en [Accessed 26 Feb 2018]

[15] WHO Regional Office for Europe. (2016). From Innovation to Implementation. Copenhagen, Denmark. Available at: http://www.euro.who.int/__data/assets/pdf_file/0012/302331/From-Innovation-to-Implementation-eHealth-Report-EU.pdf [Accessed 26 Feb 2018]

Note: The WHO European Region is composed by 45 nations. Interactive results available at: http://portal.euro.who.int/en/data-sources/ehealth-survey-2015/ [Accessed 2 Feb 2018]

[16] IDC, Open Evidence. (2017). European Data Market, SMART 2013/0063. European Commission. Available at: https://ec.europa.eu/digital-single-market/en/news/smart-20130063-study-european-data-market-and-related-services [Accessed 26 Feb 2018]

Note: Here 'data market' is defined as: 'the marketplace where digital data is exchanged as "products" or "services" as a result of the elaboration of raw data'.

[17] Sospedra E, Gené M. (2016). Biomedical Research with Human Biological Samples: A New Challenge for Bioethics and for the Clinical Research Ethics Committees. Revista Española De Medicina Legal. 42(3), 89–92.

[18] de Lecuona I. (2017). The tendency of commodification of parts of the human body and intimacy in research with biological samples and data (small and big). In Spanish. In M. Casado (Ed.), De la Solidaridad al Mercado (p. 267–295). Col·lecció de Bioètica (8). Barcelona: University of Barcelona.

[19] Casado M, et al. (2016). Declaration on Research Integrity in Responsible Research and Innovation. Observatory of Bioethics and Law, UNESCO. Barcelona: University of Barcelona.

economic interests of data controllers and processors, and the right to be forgotten was recognised.[20] The right to be forgotten justifies the removal of online information when it proves inaccurate, inadequate, irrelevant or excessive for the purposes of data processing.

A year later, the European Union Agency for Network and Information Security (ENISA) recognised the risk that the use of big data poses for privacy and protection of personal data. ENISA shifted the discussions from 'big data or privacy' to 'big data with privacy', endorsing privacy and data protection as core principles. According to ENISA, the adoption of these principles will benefit individuals and 'the very prosperity of big data analytics'.[21]

The European General Data Protection Regulation (GDPR) came into force on 24 May 2016 and applies from 25 May 2018. It aims to harmonise European politics and mechanisms to protect citizens against the abuse of their personal data and safeguard their privacy. The Regulation recognises that 'the protection of natural persons in relation to the processing of personal data is a fundamental right'. [22] Along the same line, the GDPR assigns new responsibilities to data controllers. For example, the GDPR extends its jurisdiction to all companies that process data from individuals who reside in the EU, independent of the location of the head office. It introduces an obligation to carry out data-protection impact assessments of operations *a priori* and introduces the legal requirement of 'privacy by design' and data portability. Furthermore, it establishes that controllers have to immediately notify users and states as soon as they become aware of personal data breaches. Accordingly, compliance with the regulation will be reinforced by strong sanctions. In order to help data controllers comply with the Regulation, the figure of the Data Protection Officer is introduced, an expert on data protection law.

From the point of view of citizens, the GDPR recognises their right to access and receive information about the use of their data from data controllers and processors,

[20] Court of Justice of the EU. (2014, May 13). An Internet Search Engine Operator is Responsible for the Processing that it carries out of Personal Data which appear on Web Pages Published by Third Parties. Press Release No. 70/14, Judgment in Case C-131/12.

[21] D'Acquisto G, et al. (2015). Privacy by Design in Big Data. ENISA.

[22] European Parliament and of the Council. 2016. Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC [Accessed 30 May 2018]

acknowledges the right to be forgotten and improves the conditions of informed consent.[23] It is a turning point for empowering EU citizens through transparency, accessibility and information, in the face of a data-driven economy.

3. MEDICAL RECORDS, DIGITALISATION AND INTEROPERABILITY

Given that the integration of the EU has increased citizens' mobility, healthcare systems have the difficult task to provide the best care across the community as a whole. In parallel, advances in medicine, technology and biomedicine have transformed health care into a highly specialised activity, offered by interdisciplinary teams often spread across institutions. This mobile and highly-specialised health care requires an 'integrated and interoperable European health information space'.[24] An important requirement to building that space is the digitalisation of medical records. Electronic medical records are single, online data collection of personal medical information, which are accessible via multiple routes and have the potential to be linked to other databases.

Health-related data, such as that contained in medical records, is used for steering policymaking and improving public health, and it has an enormous potential for research and innovation, in which both public and private sectors currently participate. Especially regarding research and innovation, the nature of digital data poses legitimate questions about privacy, where old legal standards, applied to analogue medical files, become obsolete, as in the case of anonymisation.

The integration of the different EU systems in order to achieve interoperability faces relevant challenges. Although electronic health records are quickly being adopted, the EU member states have yet to come to agreements; for example, regarding the minimum content of electronic medical records or the consent required to create or access these files. Furthermore, specific legislation needs to be developed in order to govern the institutions involved, standards for data exchange, secondary uses, role-based access, and the right of citizens to access, erase or modify data.[25,26] Community

[23] Ibid.

[24] Commission of the European Communities. (2006). Connected Health: Quality and Safety for European Citizens. Luxembourg: Office of Official Publications of the European Communities. Available at: https://ec.europa.eu/digital-single-market/en/news/connected-health-quality-and-safety-european-citizens-report-pdf-1-mb [Accessed 2 Feb 2018]
Note: Here 'interoperability' is defined as 'systems and services that are connected and can work together easily and effectively, while maintaining patient and professional confidentiality, privacy, and security'.

[25] EU Health Programme, op. cit. note 14.

[26] WHO Regional Office for Europe, op. cit. note 15.

guidelines should become efficient, inclusive and transparent national frameworks which are required to immediately tackle subjects such as anonymity, secondary uses of data, mobile apps, social media and the role of the private sector. In the case of electronic medical records, the EU states will not only have to implement privacy by design, but privacy from interoperability.

In order to better understand the tensions related to the protection of personal data when applied to biomedical research, we analyse in the next section initiatives that involve the secondary use of health-related personal information. In section 4.1 the Catalan VISC+/PADRIS project will be discussed and in section 4.2 the case of the UK Biobank. Health-related apps and wearable technology (mHealth), though relevant to the subject, are beyond the scope of the present paper.

4. IT'S ALL ABOUT DATA: BIOBANKS AND HUMAN DATABASES

Biobanks, collections of biological samples mainly created for assisting biomedical research, have been widely used across the EU and important efforts to regulate them have been made.[27] Alongside scientific and technological advances, the material contained in biobanks diversified and the amount of data generated from those samples grew exponentially. Furthermore, biobanks have integrated other sources of information such as electronic medical records, 'omics' data,[28] and information about life-style. This information is organised in databases and is processed in the context of big data.

Nowadays biobanks and human databases generate intense ethical debate regarding the right to research, the advancement of science and biomedicine, the common good and individual rights. Because their stability largely depends on the perception of the general public, it has become a challenge for biobanks and human databases to earn and maintain citizens' trust. The Icelandic deCODE project,[29] and PADRIS and the UK Biobank discussed here, are only a few examples which show that ethical viability is at least as important as legal legitimacy.

---

[27] Gottweis H, et al. (2012). Biobanks for Europe. A Challenge for Governance. Luxembourg: Publications Office of the European Union.

[28] Note: 'Omics' is a suffix added to the name of different fields to indicate the incorporation of high-throughput technologies (large scale), e.g. genomics, transcriptomics, proteomics, metabolomics, etc. Omics has developed at an extraordinary rate after the Human Genome Project, which motivated the development of affordable technologies.

[29] Winickoff DE. (2006). Genome and Nation. Iceland's Health Sector Database and its Legacy. Innovations. Technology, Governance, Globalization. MIT Press. 80–105.

4.1. The Catalan Projects: VISC+ and PADRIS

VISC+ (in Catalan it stands for Adding Value to Health Information in Catalonia) was a big data project for biomedical research in the hands of the Agency for Health Quality and Assessment of Catalonia (abbreviated AQuAS in Catalan). The project was founded on primary care databases[30] and hospital databases, including the Shared Medical Records of Catalonia[31] (abbreviated HC3 in Catalan).[32,33]

The objective of the VISC+ project was to drive research and innovation in medicine and health sciences, based on the reutilisation of personal health-related information at big data scale, as encouraged by the EU.[34] Based on opt-out presumed consent, VISC+ had the goal of offering a unified Catalan health database to third parties for health-related research,[35] opening the door to those seeking economic gain.[36] This objective is in direct conflict with the principle of 'purpose limitation' recommended by the European Commission, which considers explicit consent necessary when personal data is used for commercial purposes,[37] like VISC+ claimed.

---

[30] Note: Such as the Information System for the Development of Research in Primary Care (abbreviated SIDIAP in Catalan).

[31] Note: The HC3 was built upon presumed consent and an opt-out option, through a process that was very limited in regards to information, transparency and inclusive participation. The unique requirement to access the HC3 is the presumed consent given by the patient when requesting health assistance. The implementation of such minimum requirement is apparently justified, since citizens have been afforded the rights to Access, Rectify, Cancel and Oppose (ARCO, Organic Law 15/1999 of 13 December on the Protection of Personal Data). Already by 2016, almost all health centres in Catalonia were connected to the HC3 network. For more information: Buisan L. (2013). The confidentiality in health care. In Catalan. Barcelona: Publicacions i Edicions de la Universitat de Barcelona.

[32] Solans Fenández O, et al. (2017). Shared Medical Record, Personal Health Folder and Health and Social Integrated Care in Catalonia: ICT Services for Integrated Care. In G. Rinaldi (Ed.), New Perspectives in Medical Records. TELe-Health (p. 49–64). Cham: Springer.

[33] CatSalut. (2016). Proceedings 2015. In Catalan. Catalan Health Service. Government of Catalonia, Department of Health. Available at: http://catsalut.gencat.cat/web/.content/minisite/catsalut/coneix_catsalut/memories_activitat/memories_catsalut/2015/memoria_catsalut_2015_integra.pdf [Accessed 26 Feb 2018]

[34] European Commission. (2014). Towards a thriving data-driven economy. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Brussels. Available at: https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/towards-thriving-data-driven-economy [Accessed 26 Feb 2018]

[35] AQuAS, Department of Health. (2015). More Value to the Health Information of Catalunya (VISC+). In Catalan. Government of Catalonia, Department of Health. Available at: http://aquas.gencat.cat/web/.content/minisite/aquas/projectes/antic_visc/memoria_visc_aquas2015.pdf [Accessed 26 Feb 2018]

[36] Llàcer MR, et al. (2015). Document on Bioethics and Big Data: Exploitation and Commercialisation of User Data in Public Health Care. Observatory of Bioethics and Law, UNESCO. Barcelona: University of Barcelona.

[37] EGE. (2014). Ethics of Security and Surveillance Technologies. Opinion No. 8 of the European Group on Ethics in Science and New Technologies. European Commission. Available at: https://publications.europa.eu/en/publication-detail/-/publication/6f1b3ce0-2810-4926-b185-

Despite the official support that VISC+ enjoyed, several groups, like Marea Blanca (a civil society organisation),[38] the Catalan Parliament[39] and the Observatory of Bioethics and Law of the University of Barcelona,[40] argued against its implementation. The Observatory published a Declaration on Bioethics and Big Data, which addressed the implications of the use of public healthcare information, with the goal of guiding policy-makers. The declaration was followed by the Catalan Parliament, which urged the project to be stopped and to instead develop a public program that would not allow the commodification of personal healthcare data.[41] This approach to biomedical research is the spirit of the Spanish Law 14/2007[42] and the Royal Decree 1723/2012,[43] which ensures a climate of altruism and solidarity.

After VISC+ was challenged by deficiencies and strong criticisms, it was abandoned in 2016. A new public-funded project replaced it: the Public Data Analysis for Health Research and Innovation Program (abbreviated PADRIS in Catalan).[44] The project is going through its initial phase of definition and establishment, but it is clear that it will fall under the responsibility of AQuAS, the only gateway of access to the project. Firstly, PADRIS wishes to participate in the improvement of the health system. Secondly, it responds to the determination to maintain and improve the position of Catalonia in the international scientific community, by providing health information for research and innovation.

According to early reports, PADRIS will be largely based on opt-out consent, excepting the use of genetic data which will require explicit consent since AQuAS considers it a 'personal and unique identifier' impossible to anonymise.[45] Under this premise and given the strong consensus that complete anonymisation of personal health data is impossible, PADRIS would be required to implement explicit consent for all its data,

54fc3225c969/language-en [Accessed 26 Feb 2018]

[38] Marea Blanca. Information available at: http://www.mareablanca.cat/tag/visc/ [Accessed 2 Feb 2018]

[39] Motion 150/X of the Parliament of Catalonia, about the project VISC+. (2014, October 30). Palace of Parliament of Catalonia.

[40] Llàcer MR, et al., op. cit. note 36.

[41] Motion 49/XI of the Parliament of Catalonia, about the project VISC+. (2016, June 30). Palace of Parliament of Catalonia.

[42] Spanish Law 14/2007 of 3 July on Biomedical Research. Available at: https://www.boe.es/buscar/doc.php?id=BOE-A-2007-12945 [Accessed 28 Feb 2018]

[43] Royal Decree 1723/2012 of 28 December, which Governs Activities of Procurement, Clinical Use and Regional Coordination of Human Organs for Transplant and Establishes Quality and Security Requirements.

[44] Government of Catalonia. (2017). Public Program of Data Analysis for Health Research and Innovation in Catalonia –PADRIS–. AQuAS. Barcelona. Available at: http://aquas.gencat.cat/en/projectes/analitica_dades/ [Accessed 2 Feb 2018]

[45] Ibid.

which is not the case. Under the current level of public engagement and general knowledge of risks and benefits of PADRIS, the opt-out consent would be inadequate. PADRIS has restricted access to the resource to public and certified research and health related institutions,[46] which may be an attempt to prevent the commodification of data. However, it did not take into account that some of the research carried out in these institutions is done in collaboration with the private sector. Secondly, big data analytics is an intense activity that requires infrastructure, software, hardware and highly specialised professional services, where usually private parties get involved. Therefore, we urged AQuAS to set clear strategies for PADRIS to appropriately deal with the role of the private sector in research and innovation, or explicitly restrict it. More importantly, PADRIS has the obligation to design a consent that will satisfactorily inform citizens about who may have access to their personal information and to which possible ends.

4.1.1. PADRIS, public participation and the common good

As a result of the approval of PADRIS, AQuAS was required to implement a participatory and deliberative public consultation process.[47] Accordingly, AQuAS organised three activities (two face-to-face meetings in Barcelona and an online questionnaire).[48]

The exact follow-up strategy is yet unknown, but since the activities mentioned before were implemented in Barcelona and not in the whole of Catalonia, and the online form only targets a fraction of the general public, it would be expected that a more inclusive second phase of public engagement would be carried out. To this end it is crucial to choose the appropriate channels of communication in order to broaden the consultation to the public as a whole.

Furthermore, PADRIS must implement a permanent working line to implement systems of public inclusive debate with a pragmatic approach, what we would like to call 'bioethics in action'. To include organised participation systems in the governance

[46] Ibid.
Note: PADRIS grants access to institutions certified by the Agency for the Research Centres of Catalonia (CERCA in Catalan) and to institutions associated with public hospitals or non-profit organisations from the Integrated Public Healthcare System of Catalonia (SISCAT in Catalan) or with the Catalan Association of Public Universities (ACUP in Catalan)
[47] Ibid.
[48] AQuAS. (2016, August 08). Participative and deliberative process of PADRIS. Available at: http://aquas.gencat.cat/ca/projectes/analitica_dades/proces-participatiu// [Accessed 24 Feb 2018]

frameworks, to share responsibility regarding data, to formulate clear protocols of ethical review and to ensure accountability, will generate the trust and transparency needed as foundation of PADRIS.[49] Accordingly, in cases of conflict, there must be protocols in place to guarantee that individual rights will never succumb to economic, scientific or political interests.

Additionally, PADRIS needs to incorporate an analysis of the individual and social impacts and a deep technical exploration to inform data security and privacy protocols. To this end, the participation of data scientists is vital to evaluate PADRIS and those projects seeking access to the resource.

Although the VISC+ project generated tensions and public distrust and there are unanswered questions about PADRIS, we should keep an open mind towards the use of personal information for the common good. In the next section, the case of the UK Biobank will be studied as a second example of highly sensitive personal data used as a resource for biomedical research.

4.2. UK Biobank

The UK Biobank is a non-profit charity supported by the National Health Service (NHS) that aims to support biomedical research and offer insights into the transition from health to disease, causes of death and the role of life-style.[50] The biobank was based on the construction of the 'altruistic self' in order to recruit participants, which amounted half a million.[51] Participants provided personal data, health information, and biological samples (blood, saliva and urine). Moreover, participants agreed that their medical records would be linked to their profile and indicated if they wished to be re-contacted.

Although the UK Biobank is the owner of the database and samples donated, the commitment of participants is not absolute. They have the right to withdraw partially or completely, with limitations associated with transferred samples or information, which may be impossible to destroy. Although clarification of the limits to control

[49] Nuffield Council on Bioethics. (2015). The collection, linking and use of data in biomedical research and health care: ethical issues. Medical Research Council, Nuffield Foundation and Wellcome Trust. Available at: http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf [Accessed 26 Feb 2018]

[50] UK Biobank. (2007). UK Biobank Ethics and Governance Framework, Version 3.0. Available at: http://www.ukbiobank.ac.uk/ethics/ [Accessed 26 Feb 2018]

[51] Tutton R, Prainsack B. (2011). Enterprising or altruistic selves? Making up research subjects in genetics research. Sociology of Health & Illness. 33, 1081–1095.

samples and information is important for donors, traceability protocols must be under constant improvement. Biobanks often offer the possibility of complete opt-out as the ultimate donors' control instrument, but should continue to improve governance mechanisms that empower participants and directly involve them in decision-making, if they wish to do so.

For storage purposes, the UK Biobank will reversibly anonymise data and samples, and when not, it guarantees to protect them through severe access restrictions. Accordingly, data and samples will be anonymised before they are provided to researchers. These efforts are reinforced by an agreement with researchers to refrain from identifying the participants. The UK Biobank considers that these measures are sufficient to ensure privacy, and categorises the reidentification risk as 'relatively small'.[52]

The biobank's strategy is to finance its activities through the fee charged to access the resource. Those fees cover administrative and application costs, and depend on the samples and information requested. Access to the resource will be granted to *bona fide* private and public researchers, as a limited, non-transferable licence, without ownership rights. Researchers can hold intellectual property (IP) rights over the databases they generate, but they are bound by the obligation to allow the incorporation of those secondary databases back in the UK Biobank's database. The UK Biobank will not accept IP rights to be used unreasonably, nor in ways that restrict research or access to health care. For example, IP rights over naturally occurring genetic sequences, biomarkers, proteins or biochemical processes.[53] This list should be continuously revised in order to avoid misinterpretation or loopholes. The policy of not holding restrictions on the private sector has caused doubts, but so far, the approach towards research and IP rights seems to have discouraged those who attempt to aggressively appropriate the results which derive from the resource.

4.2.1. UK Biobank's object: 'bona fide' research

The UK Biobank is uniquely interested in supporting 'health-related' and '*bona fide*' research. Projects are required to be of 'public interest', and consistent with the Ethics and Governance Framework and the informed consent. The body responsible for

[52] UK Biobank. (2012). Summary de-Identification Protocol. Available at: http://www.ukbiobank.ac.uk/resources/ [Accessed 26 Feb 2018]

[53] UK Biobank Coordinating Centre. (2011). Access Procedures: Application and Review Procedures for Access to the UK Biobank Resource. Version 1.0. Available at: https://www.ukbiobank.ac.uk/wp-content/uploads/2012/09/Access-Procedures-2011.pdf [Accessed 26 Feb 2018]

evaluating the research proposals is the Coordinating Centre, which is the only access gateway. Between 2012 and 2017, almost a thousand research projects have been approved.[54]

If necessary the Coordinating Centre can call upon the assistance of the Steering Committee, the Access Sub-Committee, or the Ethics and Governance Council. The latter is an independent council with the authority to refer to the Coordinating Centre, even if not called upon. Additionally, information can be sought directly from those participants who agreed to be re-contacted, or the general public.[55]

Once the biobank approves a research proposal, it does not wish to obstruct the principle of freedom of research. In the same line of thought, the UK Biobank requests researchers to share their results in advance and notify if their findings are controversial, in order to be prepared to inform the public and the media.[56] However, the possibility remains that published results could affect the interests of the British people, or particular groups. To respond to this situation, it would be of great value to find a scheme or figure associated to the biobank, by which a more proactive consultancy position could be adopted, providing advice about how to best present the data generated, without interfering with the freedom of research.

4.2.2. The focal point: trust, privacy and consent

In general, biobanks' Achilles heel is to reach an informed consent that balances the protection of participants, and supports a wide range of research. The UK Biobank found a solution using an explicit broad consent, purpose limitation and a multilevel control mechanism. Actually, broad consent is the most common type of consent used by human databases and biobanks. From the point of view of the participant, re-consenting to every new project is impractical and undesirable. In the case of the UK Biobank, a participant could have received more than 900 research applications. From a more technical standpoint, the need to re-consent could cause administrative challenges and delays, and could render data difficult to compare if the cohort changes continuously.

[54] UK Biobank. (2018). Approved Research Summary. Available at: http://www.ukbiobank.ac.uk/approved-research/ [Accessed 26 Feb 2018]
[55] UK Biobank, op. cit. note 50.
[56] UK Biobank Coordinating Centre, op. cit. note 53.

However, the adequacy of broad consent is challenged when research not explicitly included in the informed consent causes controversy. For instance, in response to a new regulation on somatic-cell nuclear transfers (hSCNT)[57] in 2008, the UK Biobank indicated that explicit consent will be requested when proposed research activities did not fall within the existing consent, but concluded that 'it is not its role to second guess science or social attitudes at an as-yet-undetermined time'.[58] That response was rather imprecise and caused criticism that could have been avoided.[59] In order to maintain trust upon the use of broad consent, biobanks must be predictable by taking clear positions regarding determined types of research.

Since the biobank aims to understand causes of death and disease including those which potentially cause mental incapacity, participants accepted that their data would be used also under these conditions. However, if a participant changes his mind and notifies the biobank, his wish to withdraw upon death or mental incapacity will be honoured. The biobank can be notified by a family member or representative or act directly according to the routine follow-up system.[60]

Ultimately, the broad consent of the UK Biobank functions as a representativeness agreement, by which participants accept that decisions regarding the use of their data will be made by a commission. Those who surrender their information to the biobank not only trust that it will be kept secure, but that the biobank's activities will not harm them. This vision is remarkably similar to the representativeness of modern democratic societies.

Two key elements that directed and shaped the establishment of the UK Biobank are privacy and trust. During the eleven years needed to get off the ground,[61] the UK Biobank went through a long process of public consultation.[62,63] Organisations such as the House of Lords, the Consumers' Association and GeneWatch (a non-profit policy research and public interest group) presented serious criticisms about the nature and

[57] UK Human Fertilisation and Embryology Act 2008. Available at: https://www.legislation.gov.uk/ukpga/2008/22/contents [Accessed 2 June 2018]
[58] Laurie G. (2009). Role of the UK Biobank Ethics and Governance Council. The Lancet. 374, 1676.
[59] Jones DA, MacKellar C. (2009). Consent for Biobank Tissue in Somatic-Cell Nuclear Transfer. The Lancet. 374, 861–862.
[60] UK Biobank, op. cit. note 50.
[61] Árnason G. (2007). On Human Genetic Databases. In M. Häyry et al. (Ed.), The Ethics and Governance of Human Genetic Databases. Cambridge: Cambridge University Press.
[62] HGC. (2001). Public Attitudes to Human Genetic Information. People's Panel Quantitative Study Conducted for the Human Genetics Commission. London.
[63] MRC. (2000). Public Perceptions of the Collection of Human Biological Samples. Medical Research Council, the Wellcome Trust. London.

applications of the resource, and the quality and validity of the public consultation undertaken.[64] A similar process was discussed before, regarding the Catalan projects.

It is not only important to earn the initial trust of the participants and the public, but to maintain it. Over time, the UK Biobank will have to demonstrate that the control mechanisms established are able to protect the interests of participants and third parties. The UK Biobank should live up to its commitment to transparency and to maintaining a solid 'engagement with participants and society in general'.[65] Concerning the scientific breakthroughs, the UK Biobank will have to continuously re-examine its goals and the spirit of the consent given by participants, a task that will require openness, transparency and involvement of both participants and the general public.

5. DISCUSSION AND RECOMMENDATIONS

Until a couple of decades ago, the economic system was based on tangible articles or services that were easily controlled. Today, that model has changed into a knowledge-based data-driven economy which is organised around data, and not around processes. Data is difficult to contain; it can be easily published, shared, copied, transferred, analysed, formatted, edited, updated, changed and deleted. On the Internet, data tends to flow freely, available to anyone who can afford the technology, and technology is becoming increasingly accessible. Before, people used to say that money speaks; nowadays, data speaks. Consequently, we need to discuss how we will protect individual rights and privacy in a data-centred world.

Until very recently, it was an extended practice that personal data was generated, capitalised and used without the consent or knowledge of the data holder. The GDPR is expected to articulate the means to give citizens the control over their personal information and protect them against spurious interests and commodification. Nevertheless, the post-Brexit position of the UK regarding the GDPR remains to be seen. The first challenge as society is to undergo a process of education and empowerment regarding the value of data. The individual's autonomy, manifested as written consent, can never serve as justification for projects which may come in conflict with individual rights. An example is the scheme to use personal health-care data for

[64] House of Lords. (2003). Science and Technology. Select Committee on Science and Technology, Third Report. Science and Technology Committee Publications. Available at: https://publications.parliament.uk/pa/cm200203/cmselect/cmsctech/132/13202.htm [Accessed 2 Feb 2018]

[65] UK Biobank, op. cit. note 50.

economic gain of third parties, as exposed here in the case of VISC+. Another example is the use of personal data for research that falls outside the informed consent, a common challenge for biobanks and databases.

In order to succeed, biobanks and human databases depend on the trust of the general public. Therefore, this relationship has been the focus of European authorities and European countries. For example, reports showed that approximately 54% of Spanish citizens[66] and 55% of British citizens[67] think that the benefits of science and technology outweigh the risks. Furthermore, 81% of citizens in the UK consider that science make their lives easier.[68] This positive standpoint is supported by the fact that in Spain, medical doctors, scientists and professors are the professionals most valued,[69] while in the UK 90% of citizens consider that scientists make valuable contributions to society.[70]

At a European level, the last Science and Technology Eurobarometer showed that citizens have a positive view about science and technology and consider that they should be consulted about decisions taken in this regard. Furthermore, they do not trust science funded by the private sector and it is considered that it limits or negatively influences scientific activities,[71] this view is evident in both the UK Biobank and PADRIS. Projects based on personal health-related information should take these attitudes towards science as pillars of their construction and functioning. The interests of donors or participants, and the society they represent, should not be given a secondary or optional status, but a central and integral part of the establishment, goals and activities of biobanks and human databases.

Here we analysed the Catalan PADRIS and the UK Biobank as case studies, which reflect the current status of the use of personal data in the EU. These projects are very different in their processes of establishment, management, goals and governance. However, they both exemplify approaches to the use of health-related personal information. Specifically, PADRIS aims to use public health data at a national level, under a presumed consent, while the UK Biobank is based on donors who have agreed to a broad consent supported by a representation system.

---

[66] FECYT. (2016). Social Perception of Science and Technology. In Spanish. Government of Spain, Ministry of Economy, Industry and Competitiveness.

[67] Castell S, et al. (2014). Public attitudes to science. Ipsos Mori, Social Research Institute. Department for Business, Innovation and Skills.

[68] Ibid.

[69] FECYT, op. cit. note 66.

[70] Castell, et al., op. cit. note 67.

[71] European Commission. (2010). Special Eurobarometer 340. Science and Technology Report.

Throughout this paper, we demonstrate the necessity of actively incorporating ethics from the formulation of projects based of adding value to personal data, and to be reflected in their governance mechanisms. The experiences in Europe, including the UK Biobank and PADRIS, point towards the construction of broad consent assisted continuously by feedback mechanisms and permanent revision of the objectives of the consent provided with the participants and society. They call for the use of personal data as a project for society and not uniquely as a research and innovation initiative. European responsible research and innovation (RRI) initiatives, such as RRI tools,[72] are currently starting to enable this kind of change in society, and the scientific and innovation sectors.

Although accountability should be strictly enforced, it is more important to introduce changes from within the use of personal data; for example, through the implementation of the principles of privacy by design and privacy from interoperability. Above all, infringements to individual rights or our private and family lives should not be treated as collateral damage that can be financially compensated.

What is at stake is the name of science, research and innovation, which has been in the core of the EU economic agenda. Additionally, fundamental values and rights such as liberty, free development of personality and privacy are also at risk. Freedom of research and enthusiasm for technological advances cannot be invoked against individual rights. Therefore, research and innovation based on personal data, must be performed following the highest standards and respecting the values of the society at stake, ensuring data security, transparency and privacy.

Through this work we advocate the incorporation of ethics into big data projects from their outset. We support the movement of responsible research and innovation promoted by the EU, but highlight that the interests of those individuals from whom data is generated and society in general are of foremost importance. It is necessary to develop governance frameworks that embrace respect for individual rights, responsibility and accountability. We cannot forget that citizens are the owners, providers, and consumers or potential beneficiaries of the most valuable primary material: data.

[72] RRI tools. Information available at: https://www.rri-tools.eu/ [Accessed 2 June 2018]