Challenges for Citizen Science and the EU Open Science Agenda under the GDPR

*Anna Berti Suman and Robin Pierce**

Present discussions on the implications of the GDPR for medical practice and health research mostly target the passive collection of health data. This article shifts the lens of analysis to the scarcely researched and rather different phenomenon of the active sharing of health data within the framework of Citizen Science projects. Starting from this focus, the article queries whether data processing requirements under the GDPR impacts the advancement of Citizen Science for health research. A number of tensions between the two aims are identified both in abstract terms and 'in practice' by analysing three Citizen Science scenarios and drawing parallels with the experience of 'collective' Clinical Trials. The limited literature on the topic makes this article an exploratory reflection on key tensions, with the aim of opening the way for further research. This discussion is inspired by the need to guarantee that opportunities of Citizen Science will not be unduly curtailed by the advent of the GDPR but also to ensure that Citizen Science is implemented in ways that are consistent with the GDPR.

Keywords: Citizen Science, Open Science, GDPR, Secondary Use of Health Data, Research Exemption

I. Introduction

In recent years the impeding advent of the EU General Data Protection Regulation (GDPR),[1] currently entered into force, stimulated a lively discussion on the implications of the new Regulation for data processing in medical and health research.[2]

The debate focuses on the concern that the new consent requirements would severely restrict medical and health data research.[3] However, such discussion mostly targets the *passive* use of wearables and sensors by patients and citizens, and the related challenges. The present contribution shifts the lens of analysis, bringing the focus on the scarcely researched and rather different phenomenon of the *active* use of such devices, in the instances that the users thereof voluntarily and actively decide to share their health data by engaging in so-called 'Citizen Science' projects.

We define Citizen Science and analyse it as inserted in a broader transition toward a (more) 'open' science, historically inspired by the free-software move-

---

DOI: 10.21552/edpl/2018/3/7

* Anna Berti Suman, PhD researcher at the Tilburg Institute for Law, Technology, and Society. Her PhD project aims at investigating the influence of 'Citizen Sensing' practices on environmental health risk governance. She has work and research experience in health and technology, environmental law, and Citizen Science.
Robin Pierce, JD, PhD, is associate professor at the Tilburg Institute for Law, Technology, and Society (TILT, Tilburg University, The Netherlands). She has taught in Data Protection and Privacy Law and currently teaches a course on Regulation for the LLM Law & Technology. Her current research addresses the themes of data protection law and health data and AI in medicine.
For correspondence: <A.BertiSuman@uvt.nl>.

[1] EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

[2] Menno Mostert and others, 'Big Data In Medical Research And EU Data Protection Law: Challenges To The Consent Or Anonymise Approach' (2015) 24 European Journal of Human Genetics.

[3] John Mark Michael Rumbold and Barbara Pierscionek, 'The Effect Of The General Data Protection Regulation On Medical Research' (2017) 19 Journal of Medical Internet Research.

ment and more recently captured by the EU Open Science agenda. Within this context, Citizen Science is presented as an 'enabler' of openness in science. Further, we juxtapose the new data processing requirements under the GDPR and specific manifestations of Citizen Science and Open Science to identify a number of tensions. These tensions led us to the definition of an overarching question: we query whether the data processing requirements under the GDPR represent a possible hindrance to the advancement of Citizen Science and Open Science for health research.

In trying to answer this question, we dedicate attention to the perspective of the *active* and voluntary sharing of health/medical data through the citizen/patient's engagement in Citizen Science projects aimed at realizing the values underpinning Open Science. The lens we use is consequently that of the voluntary element characterising the participation in such projects. Throughout the analysis of selected empirical scenarios, we explore whether this voluntary consent can justify the processing of citizens/patients' data in Citizen Science projects or, rather, we still miss some guarantees to ensure that processing is legitimate under the GDPR.

The article further evolves in the assessment of possible tensions between the Open Science agenda brought forward by the EU Horizon 2020 strategy, which includes promotion of Citizen Science, and the current EU data protection goals crystallized in the GDPR. Being both goals highly prioritized in current EU policies, the need to reconcile the two aims seems indispensable to enhance the EU research agenda, without harming data protection. We subsequently investigate such tensions 'in practice' by analysing three Citizen Science scenarios derived from the Scistarter platform and by assessing the experience of 'collective' clinical trials.

We considered it appropriate to contextualise such dilemmas in the broader debate on the need for a more participatory science. Arguably, Citizen Science and Open Science bring the promise to improve both the quality and the social acceptance of science. By engaging in this discussion, we provide a justification of Citizen Science and we pinpoint its legitimacy. We refer to a hypothetic 'right to science' calling for more open and participatory research processes.

We consider the GDPR as one of the possible metrics against which to assess the legitimacy of health data processing in Citizen Science projects. Other possible angles, equally valid, could and should be explored in future research. The answer to our question and 'solutions' to the tensions we present in the paper cannot clearly or exclusively be found in the GDPR. Nonetheless, considering the existence of current tensions between the EU Open Science agenda and the EU data protection agenda, it seems valuable to verify whether the GDPR sheds some light on these dilemmas.

Acknowledging that the questions raised by participatory forms of science entailing health data collection are numerous yet not sufficiently discussed in the literature, this article provides an exploratory reflection that paves the way for further research. While not providing answers, the article examines whether a conflict between Citizen/Open Science and the GDPR does exist and, if so, what should be the role of future research in harmonising the two aims. The reflection is ultimately inspired by the need to ensure that potential benefits of Citizen Science will not be compromised by an inability to comply with the stricter data processing rules enshrined in the GDPR.

II. Setting the Scene: Citizen Science and Open Science

1. Citizen Science

Citizen Science can be defined as 'the active participation of lay people in scientific research'.[4] It is a relatively old concept, traceable back to the publication in 1995 of Irwin's book titled *Citizen Science as a study of people, expertise and sustainable development*,[5] which is considered the first mention of Citizen Science in the academic literature. The practice developed through the experience of Cornell Lab of Ornithology where the first 'citizen scientists' were engaged in birds' monitoring. Subsequently, the practice developed along three main streams, as illustrated by Kullenberg and Kasperowski.[6] If the first

---

[4] Lea Den Broeder and others, 'Public Health Citizen Science; Perceived Impacts On Citizen Scientists: A Case Study In A Low-Income Neighbourhood In The Netherlands' (2018) 2(1) Citizen Science: Theory and Practice 1 <http://doi.org/10.5334/cstp.89> accessed 1 June 2018.

[5] Alan Irwin, *Citizen Science* (Routledge 1995).

[6] Christopher Kullenberg and Dick Kasperowski, 'What Is Citizen Science? – A Scientometric Meta-Analysis' (2016) 11 PLOS ONE.

stream, ie the engagement of lay people in the domains of biology, conservation and ecology (as the Cornell Lab experience), and the second, ie the engagement of citizens in the collection of geographic data, are not of interest for the present research, the last stream identified by Kullenberg and Kasperowski[7] is of particular relevance for this article. The third category of Citizen Science relates to social sciences and epidemiology, and considers the practice a tool to foster citizens and patients' participation in monitoring health (and the environment). We refer here to Citizen Science as inserted in a broader trend, namely the progressive promotion - especially at the EU level - of a more 'open' science, through a number of policies such as the Open Science and Open Data strategies (pillars of Horizon 2020) and by the 'Science with and for Society' aims of the current EU research agenda. Citizen Science can be related to the broader concept of Open Science as it 'enables openness'[8] in science. Indeed, the majority of Citizen Science projects make their data, metadata and methods publicly available and freely reusable. In addition, by making (lay) people the protagonists of research processes, Citizen Science challenges traditional ways of producing scientific knowledge and opens it up to the public contribution but also scrutiny.

2. Open Science, Open Data and Science with and for Society

The widespread push for an *open* science finds its historical roots in the more dated 'free-software movement' and 'Open Source Initiative' (OSI). The movement is here briefly outlined as it is regarded as having inspired the recent Open Science agenda. In connection with developments in software technology, a growing push for the sharing and collaborative improvement of software source code emerged. Computer programmers and developers started sharing software with the aim to mutually learn and improve computing. Examples of software released freely were the TeX typesetting system, the Netscape Communicator Internet suite (which subsequently led to Mozilla Firefox) and Linux. In order to encourage and support this open source movement, in 1998 – shortly after the Netscape source code was released – the Open Source Initiative (OSI) was founded, as an educational, advocacy, and stewardship organisation. The ideal behind the initiative was that a participatory approach to creating and improving software technology would ultimately benefit the progress of computing. This ideal highly resonates with the values underlying the Open Science transition, as it will be explained further. The OSI key actions in the field were to draft the Open Source Definition and to release in 1999 a constantly updated list of approved open source licenses.[9]

Similarly to what will be discussed for Open Science (and Citizen Science) in relation to the GDPR, also the free-software movement was confronted with some of the legal challenges deriving from *openness*. In particular, open software ideals risk to clash with protection of copyright, as emerged in the famous case *Jacobsen v Katzer* (US District Court for the Northern District of California, 2006) on copyright infringements, patents' invalidity and breach of the US Digital Millennium Copyright Act.[10] The free software scenario challenges the concept itself of copyright. Indeed, it allows software developers to modify and improve an initial software product, while the initial distributor can control the code's presentation and further dissemination. However, instead of a 'copyright', a 'copyleft' effect covers the product as the code is considered open and free for all to use in innovation and development of software. This clash within the legal system between openness and protection shows interesting parallelisms with the Open Science and Citizen Science debate presented in this article.

Moving to Open Science at the EU level, it can be affirmed that Citizen Science is one of the results of a push towards a more open science at the European level. Consequently, the Open Science EU agenda is here inspected. On the EU website, Open Science is defined as 'the ongoing transition in how research is performed and how knowledge is shared'.[11] An example of this transition is represented by the Euro-

---

[7] ibid.

[8] See European Citizen Science Association, 'Citizen Science & Open Science - Policy Brief is out!' <https://ecsa.citizen-science.net/blog/citizen-science-open-science-policy-brief-out> accessed 20 June 2018.

[9] History of the OSI by Opensource.org <https://opensource.org/history> accessed 25 August 2018.

[10] See 'Jacobsen v. Katzer' (*Creative Commons*, 2010) <https://wiki.creativecommons.org/wiki/Jacobsen_v._Katzer> accessed 25 August 2018.

[11] See European Commission, 'Open Science' <https://ec.europa.eu/research/openscience/index.cfm> accessed 6 June 2018.

pean Open Science Cloud initiative (EOSC). The EOSC is defined as 'a federated environment' aimed at enabling 'trusted access to services, systems and the *re-use* of shared scientific data across disciplinary, *social* and geographical borders', these latter including not only European but also global borders.[12]

Open Science is also closely related to the push for Open Data in Science within the European context. To this regard, it is worth to mention the Statement by the European Members of the International Council for Science,[13] recently formulated during the workshop 'Open Data in Science: Challenges and Opportunities for Europe'.[14] The Statement provides a list of key recommendations addressed to all research to be conducted on the European arena, in order to align with the Open Science and Open Data aims. At a general level, it is stressed that 'publicly funded scientists make their research data *available* in reusable format in order to *enhance the quality and effectiveness of science* and as a contribution to help address societal and environmental challenges'.[15] Making research data available is considered as a mean to a higher end: strengthening research quality and effectiveness.

The FAIR Principles are enumerated: 'Findable, Accessible, Interoperable, Reusable' research is presented as a means to enhance verifiability of research. Overall, the key message is the need to change the research culture and to create incentives that recognize the value and reward Open Science and Open Data in scientific activities. This is demonstrated by the closing words of the Statement where researchers are invited to cooperate to build 'a research culture in their communities in which data and tool sharing are the norm and promoted as part of scientific practice.'

Another important pillar of Open Science is 'Open Access' (OA), which has been defined by the European Commission as 'the practice of providing online access to scientific information that is free of charge to the user and that is *re-usable*'[16] referring both to publications and to scientific research data. Open Access has been strongly supported by Horizon 2020. Article 29.2 of the Horizon Model Grant Agreement mandates that all projects funded by the programme must ensure that any peer-reviewed journal article published is made accessible openly.

The Open Science agenda promoted by the European Commission and the Citizen Science discussion intermingle with what, since 2001, exists as the 'Science and Society' Action Plan aimed at better connecting science and citizens within the EU. In 2007, under the 7th Framework Programme for Research and Technological Development (FP7), the action plan became 'Science in Society', thus stressing the importance of public engagement of civil society in science. Under Horizon 2020 (Part V), Science in Society became 'Science with and for Society', having a particular focus on developing innovative ways of connecting science to society.[17]

The breadth of this transition towards 'Open Science' and 'Science with and for Society' cannot fit within a single paper. However, this article will account for a narrower Open Science aspect, that of involving laypeople in contributing to a more open science through Citizen Science practices. These practices represent a good example of 'Science with and for Society' and reasonate with the EU Open Science and Open Data agenda. The connection of Citizen Science with the EU agenda is underlined by the wording of the 'Citizen Science & Open Science: Synergies & Future Areas of Work' policy brief.[18] It is stated that 'Citizen Science and Open Science together can address grand challenges, respond to diminishing societal trust in science, contribute to the *creation of common goods* and shared resources, and facilitate knowledge transfer between science and society to stimulate innovation' (emphasis added). The hype is high.

However, Citizen Science and Open Science also bring a valid example of how the aims of openness and participation in science may conflict with the protection of fundamental guarantees currently recognized by the GDPR. In addition, it will be shown

---

[12] The Commission High Level Expert Group on the European Open Science Cloud, *Realising the European Open Science Cloud* (Publications Office of the European Union 2016).

[13] European ICSU Members, *Open Data in Science in Europe Statement by the European Members of the International Council for Science* (2018) <http://www.euro-icsu.org/thematic_work/opendata/documents/European_ICSU_Members_Open_Data_Statement_2018.pdf> accessed 5 June 2018.

[14] The workshop brought together the European Members of the International Council for Science (ICSU) and the All European Academies (ALLEA) in Brussels on 31 January 2018. A number of other participants, from science organisations, public research funders, the publishing sector, researchers and policy makers.

[15] European ICSU Members (n 13) (emphasis added).

[16] See <https://ec.europa.eu/research/openscience/index.cfm?pg=openaccess> accessed 6 June 2018 (emphasis added).

[17] See European Commission, 'Science With And For Society (Swafs)' <https://ec.europa.eu/research/swafs/index.cfm?pg=about> accessed 6 June 2018.

[18] See European Citizen Science Association (n 8).

that under the rhetoric of 'democratising' medicine and scientific knowledge production, market interests endangering the participants/patients may be hidden. In the following section, such possible tensions existing between the aims of openness and of participants/patients' interests, especially related to data protection, will be disentangled. The ultimate aim is to ascertain whether data processing requirements under the GDPR represent a possible hindrance to the advancement of Citizen Science for health research and, if this is the case, how this conflict could be resolved.

Literature elaborating on a possible contrast between Open/Citizen Science and data protection is currently scarce. Consequently, the discussion that follows has been inspired by similar reflections on comparable contrapositions between data protection and other goals. Gellert,[19] for example, discusses the risk to privacy posed by smart grids and how the aims of environmental protection may clash with privacy and data protection guarantees. Such approaches are transposed to the present discussion, with the aim to build the frame for identifying a possible conflict between two sets of interests.

III. Tensions between Scientific Openness and Data Protection

In confronting the Citizen Science and Open Science strategies at the EU level and the recent developments brought forward by the GDPR, we identified a number of tensions that can be summarised along the following lines:

- A possible need to distinguish between the nature of consent when data are actively and voluntarily disclosed in the framework of Citizen Science projects, and the nature of consent for the passive data collection in case of the use of wearables and other tracking technologies;
- Related to the previous point, the limits of 'voluntary' consent and – as spelled out in the empirical scenarios – to what extent voluntary participation ensures awareness over possible risks of data processing;
- Whether the 'safeguards' mentioned in the GDPR, eg pseudonymisaton could diminish the value of some Citizen Science projects. While explicit consent could be understood as foundational to the voluntary nature of participation in Citizen Science, it is not always clear that participants know and understand the full nature of their participation and the risks involved in disclosure of identifiable or re-identifiable data. Thus, the use of identifiable data, even voluntary, in the absence of explicit informed consent, and without employing safeguards of de-identification, would likely be in violation of the GDPR. This could also be in violation of national laws governing health data and medical research if, as France does, explicit consent is required for the processing of identifiable data for research.[20] This is a defensible position in view of longstanding norms of research ethics. However, it does add an additional layer to coordination of Citizen Science projects if documentation of informed explicit consent is required in the same way that is required of institutional research:
- A possible tension between the long lasting and iterative nature of many Citizen Science projects which gather consent only at an initial stage, which leads us to question the effectiveness and legitimacy of consent over an extended time;
- Arguably the most serious tension, the decentralised nature of Citizen Science's activities which may be considerably hindered by the GDPR requirement for a data controller;
- The importance for the amplification of Citizen Science and Open Science of a wide and cross-border research data sharing, possibly clashing with the more stringent GDPR requirements for trans-border data sharing among research units, labs and across countries.

In a sense, the EU Open Science and Citizen Science strategies seem to account for the existence of such tensions and for the need to balance possibly conflicting interests. For example, the above mentioned Statement by the European Members of the International Council for Science of 2018 underlines the demand for FAIR metadata based on 'trustworthy data repositories' which can ensure '*long-term* preservation of Open Data'(emphasis added).[21] Interestingly,

---

[19] Rapahael Gellert, 'Redefining the smart grids' smartness. Or why it is impossible to adequately address their risks to privacy and data protection if their environmental dimension is overlooked' (2015) 24(1) Journal of Law, Information and Science.

[20] art 54, Act no 17 of 6 January 1978 on Information Technology, Data Files and Civil Liberties, France.

[21] European ICSU Members (n 13).

this recommendation entangles with GDPR provisions, such as the new limits to data storage ex Article 5(1)(e) and Recital (39) of the GDPR. However, in line with the data protection and privacy aims promoted by the GDPR, the Statement stresses that such repositories should 'adequately caters for privacy, security and intellectual property issues, especially in relation to personal data'.[22]

Furthermore, the promotion of a 're-use culture', illustrated in the preceding section, may conflict with some of the goals pursued by the GDPR, such as that of 'purpose limitation' [Recital 28; Article 6(1)(b); Recital 50; Article 5(1)(b); Article 89] and of 'data minimisation' [Recital 28; Article 6(1)(c); Recital 39; Article 5(1)(c)]. There may be a clash of two possibly conflicting cultures, that of opening and that of protecting research data. This conflict, as reflected in the report 'Implementation Roadmap for the European Open Science Cloud',[23] is *social* rather than technical.

In line with the present discussion aimed at 'reconciling' these tensions is the recent 'Open Research Data' (ORD) pilot,[24] a programme devoted to make 'research data generated by Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access'. The goal of optimising sharing, reuse and openness is balanced with other possibly conflicting interests such as protection of scientific information, intellectual property rights, data protection and privacy concerns. The example of the ORD pilot shows that the two aims can and should be reconciled. In fact, only their combination can lead to the production of sound and ethical science.

IV. Dilemmas in Citizen Science: Three Scenarios

Citizen Science is increasingly being used for health and medical research. In order to provide an overview of the existing Citizen Science projects in these areas and to verify tensions 'in practice', we screened the SciStarter platform,[25] an online repository of Citizen Science projects. We used the keywords 'health & medicine'. We did not add a geographic area as the majority of the projects do not operate only on a country but interact with patients/users from several countries, among which EU countries.

The first project we analysed from the SciStarter platform is called 'Rethink Fertility', a global-scale project run by the Department of Sociology at the University of Oxford, which we categorise as an 'institutional' actor. [26] The goal of the project is to 'generate insights into modern infertility issues' drawing on personal DNA reports. The Rethink Fertility team on the project's platform asks to 'people around the world, of all ages and nationality' help 'to uncover major social and genetic contributors to infertility'.[27] This help can be provided by participants via the filling of a questionnaire and the uploading of personal reports, in case they had their DNA genotyped (either through 23andMe, Ancestry.com or MyHertiage; there is also a link to Gencove.com in case people wish to request a report). The project coordinators ensure that the Rethink Fertility web portal on which the data need to be uploaded is 'secure'.[28] However, no clear information is provided on the platform regarding what it is meant for *secure* and which standards are applied for securing the data. In addition, the reliance on market-oriented services such as 23andMe, heavily contested in the literature,[29] raises additional concerns in terms of the awareness and freedom of the users in sharing their health data with the Rethink Fertility platform.

Another project we analysed is the 'Cochrane Crowd', a Citizen Science initiative aimed at categorising and summarising healthcare evidence in order to 'make better healthcare decisions'.[30] The project page explicitly affirms: 'No experience necessary!' and adds a slogan 'Trusted evidence. Informed decisions. Better health'.[31] The project exists only online, and asks participants 'just five minutes a week [that] will make a difference'. The project seems 'grassroots-driven' as initiated by a collective

---

[22] ibid.

[23] *Implementation Roadmap for the European Science*, SWD(2018)83.

[24] See (n 16) and OpenAIRE, 'What is the Open Research Data Pilot?' (14 November 2017) <https://www.openaire.eu/what-is-the-open-research-data-pilot> accessed 6 June 2018.

[25] See <https://scistarter.com/> accessed 7 June 2018.

[26] See SciStarter, 'Rethink Fertility' <https://scistarter.com/project/19992-Rethink-Fertility>; Rethink Fertility website <http://www.rethinkfertility.org> accessed 10 June 2018.

[27] ibid.

[28] ibid.

[29] See for example Anna Harris, Susan Kelly and Sally Wyatt, *CyberGenetics Health genetics and new media* (Routledge 2016).

[30] See SciStarter, 'Cochrane Crowd' <https://scistarter.com/project/12379-Cochrane-Crowd>; Cochrane Crowd website <http://crowd.cochrane.org/index.html> accessed 11 June 2018.

[31] ibid.

of people interested in crowd-sourced health evidence. It appears rapidly growing in terms of social uptake. As a matter of fact, the website describes a community currently including over 4,000 people from 88 countries that managed to categorise over 1 million research records since 2016. The Cochrane Crowd collects and details health evidence from 'the crowd' and appraise, assess, and synthesize them in systematic reviews.[32] Although a project like this would require extensive empirical analysis, some questions can already be outlined. For example, is the crowd really an even group of equally informed and equally powerful actors or, rather, there are hidden forces and interests pushing the discourse for a wider data sharing. We also question whether the noble aim of 'better health' is the only or true driver of the initiative or market interests, such as those underpinning 23andMe, act behind the scene. Lastly, with regard to the voluntariness of the data sharing, to what extent are participants actually persuaded to join by the overenthusiastic rhetoric of improving health through new evidence? How conscious are they of the possible risks entailed by sharing their health data on the platform? This all heavily affects the nature of consent, which could be seen as shifting from free and informed to induced and even misinformed consent.

The last scenario we considered is not explicitly falling within the Citizen Science category (and it is not included in the SciStarer platform) but nonetheless presents interesting parallels with the present Citizen Science discussion as it again shows instances of participatory science involving volunteered health data collection. It revolves around the use of 'collective' clinical trials exemplified by experiences such as that of the Apple Research and Care Kits,[33] which recently entered the arena of research on medical products. On the Apple website, the two kits are respectively defined as a 'software framework for apps that let medical researchers gather robust and meaningful data' and a 'software framework for apps that let you better understand and manage your medical conditions'. The first is of higher interest here as it involves, similarly to Citizen Science, research based on health data fed by users of apps, sensors and smartphones. The potential of scaling up health research to a crowd of connected users recently attracted the interest of big pharmaceutical companies, such as Norvartis, which decided to adopt the Apple Research Kit for its trials.[34]

The hype, similarly to Citizen Science, is high. On the Apple website, we read 'Doctors around the world are using iPhone to *transform* the way we think about health. Apps created with ResearchKit are already producing medical insights and discoveries *at a pace and scale never seen before*'.[35]

The promise is that digital technology will allow opening up clinical trials to a great number of patients without the need for their physical presence, thus making recruitment easier and reducing the costs. The Apple Research Kit and similar projects can gather health/medical data remotely from participants using iPhone apps, with several possible applications such as research on asthma, breast cancer, cardiovascular disease, diabetes and Parkinson's disease.

In an Apple's press release on the project, it is stated that 'users *decide* if they want to participate in a study and how their data is shared'.[36] However, considering the inciting discourse of a health research transformation which will tremendously improve health care, it is natural to inspect to what extent users freely choose to join or are not persuaded to join by hidden actors leveraging on the hype. Although these collective forms of research are more 'structural' than some of the Citizen Science projects we described, being inserted in companies' formal research schemes, the questions brought to the table do not differ substantially.

While the aim of making medicine and science more transparent and participatory should be welcomed, the other view of the described scenarios would inquire about the implications of this type of participatory research in the broader commercial context where companies (eg the famous case of 23andMe) can exploit the participants/patients' valuable informational assets. The risk is that behind the

---

[32] ibid.

[33] See Apple, 'ResearchKit and CareKit' <https://www.apple.com/researchkit/> accessed 12 June 2018.

[34] See Andrew McConaghie, 'Novartis and Apple to scale up clinical trial collaboration' (*pharmaphorum.com*, 24 January 2018) <https://pharmaphorum.com/news/researchkit-novartis-apple-scale-clinical-trial-collaboration/> accessed 11 June 2018.

[35] ibid (emphasis added).

[36] See Apple, 'Apple Introduces ResearchKit, Giving Medical Researchers the Tools to Revolutionize Medical Studies' (Press release, 9 March 2015) <https://www.apple.com/newsroom/2015/03/09Apple-Introduces-ResearchKit-Giving-Medical-Researchers-the-Tools-to-Revolutionize-Medical-Studies/> accessed 12 June 2018 (emphasis added).

hype of a science more open to the citizen and thus more democratic, the 'winners' are the market actors rather than the people. In light of these concerns, restrictions imposed by the GDPR may be seen – on one side – as a possible hindrance to the flourishing of such projects. However, on the other side, they should be regarded as a possible way to reconcile the progress of science with the need to ensure data protection.

V. Does the Nature 'From Below/From Above' of Citizen Science Change the Game?

In our investigation outlined in the preceding section, we found some examples of very 'institutionalised' Citizen Science projects, namely initiatives launched by academic/scientific research centres or public bodies (such as the Rethink Fertility project), but also examples of grassroots-driven initiatives (such as the Cochrane Crowd) or created by the private sector (such as the Apple Research Kit). In all these projects, patients/users were invited to voluntarily and actively share their health data by contributing to Citizen Science/collective research projects with different goals. These goals ranged from personal learning over a health/medical topic, the obtaining of a sort of reward (eg virtual 'coins' to be spent for free medical check-ups), the support of new scientific discoveries and the contribution to the 'common good' (eg by reporting effects of air quality on health conditions of the population in a specific neighbourhood) etc.

Overall, we categorised these initiatives as more or less 'from below' or 'from above', the first category standing for groups of *uninvited* patients/users which organise themselves in collectives to set up a Citizen Science project, whereas the second indicates instances of *invited* patients/users that are asked to join a specific project by institutional/market actors. In both cases, however, the participants actively decide to contribute to the initiative as a 'free' choice. Yet, their level of awareness on possible risks related to their sharing of health/medical data may vary substantially from the first to the second category. In order to answer the outlined dilemmas, one hypothesis may be to argue that in the first instance citizens have a certain level of awareness needed for setting up the Citizen Science project, which would arguably 'warn' them against adverse effects of data sharing. In the second instance, contrarily, the invited participants may decide to join with a lower awareness on potential risks, 'tempted' by attractive discourses such as the contribution to make scientific discoveries, which may even worsen in cases of private companies offering the service (eg the well-known case of the genetic service 23andMe).

However, another equally possible hypothesis may be that of claiming that, in more 'institutionalised' Citizen Science projects, the participants are made aware of possible risks by competent parties, which also set up the project in a way that minimises harm to users. Yet, this second hypothesis could sound naïve as it misses the understanding of the vested interests that may be hidden behind a Citizen Science project 'from above', such as mass surveillance, nudging aims and market purposes.

Both lenses led us to question to what extent the voluntariness of the sharing is a result of a free, conscious and informed choice. In the following sections, we will go through this and other tensions identified previously in Section III to assess whether also 'in practice' a conflict does exist between aims to openness and to data protection.

VI. The Tensions in the Light of the Empirical Scenarios

Through the inspection of the three scenarios, we identified a number of problems in the confrontation between Citizen Science 'in practice' and the GDPR. Our aim was to verify whether under the GDPR the processing of health data in Citizen Science projects (and adjacent CCT) can be considered safe and legitimate.

First, we ask whether Citizen Science projects can be covered by Article 89 of the GDPR 'Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes'. We believe that Citizen Science may fall under the scientific research category (in particular considering the provisions of Recitals 157 and 159 related to 'studies conducted in the public interest in the area of public health'). Yet paragraph 1 of Article 89 stresses the need - also for this type of processing - of *appropriate safeguards for the rights and freedoms* of the data subject. We underlined above the freedom of the sharing subject can be questionable in some of the scenarios discussed. This is also subject to relevant national laws pertaining to permissible use of archived personal data.

Recital 26 of the GDPR states that 'the principles of data protection should [...] not apply to *anonymous* information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable' (emphasis added). Citizen Science, if using anonymous data, could be considered 'safe' as it is not classified as personal data. However, the first problem we pinpoint is that often Citizen Science initiatives include the disclosure of sensitive personal data, such as ethnic group (eg indigenous nationality) or a particular status (eg status of HIV patient), which are functional or, in certain cases, indispensable to the aims of the initiative (eg engaging indigenous women in a Citizen Science project targeting HIV coping strategies). Anonymisation of this data may, in some instances, lessen the scientific value of the data and hence the initiative, particularly if personal trajectories are relevant to the study objectives. Pseudonymisation, a possible safeguard specifically mentioned in the GDPR, may preserve some value in such instances, contingent upon national regulatory frameworks. For example, France requires that identifiable data can only be used in research where explicit consent has been given.[37] This can give rise to a conflict between preserving scientific value and complying with the GDPR, creating unforeseen burdens for the initiative coordinators.

An example of a tense relation between openness and data protection can be the use of Citizen Science projects in schools. Citizen Science has indeed grown as a method for active and participatory learning for students. Yet students in school-age have often not yet reached the major age. In order to collect data from children under 16, parental consent is needed under the GDPR. Given the need to preserve a minor's long term privacy interests, the requirement of parental consent is a reasonable restriction. Nevertheless, these additional measures to be taken by the project coordinators may discourage the diffusion of Citizen Science in schools.

Analysing further the presented Citizen Science projects under the lens of the GDPR, it is unclear who would be data controller for the participants' data processing. Often, Citizen Science projects are geographically scattered and rely on collaborations between formalised and informal actors and groups. Their model of governance could be captured as a decentralised, federate model. Such a model, while creating a number of opportunities for innovation from various independent players, may create confusion in terms of *who is ultimately controlling* the data that are at stake. A hypothetical solution would be that of providing for more joint controllers, given that they can be identified clearly in this panorama of actors. Consequently, Article 26 GDPR on Joint Controllers would apply to the actors that jointly determine *the purposes and means* of the processing. Paragraph 1 of the article mandates that 'they shall in a transparent manner *determine their respective responsibilities* for compliance with the obligations under this Regulation, *in particular as regards the exercising of the rights of the data subject and their respective duties* to provide the information [..]' (emphasis added). Paragraph 2 states that 'the essence of the arrangement shall be made available to the data subject' and paragraph 3 adds that 'irrespective of the terms of the arrangement referred to in paragraph 1, the data subject may exercise his or her rights under this Regulation *in respect of and against each of the controllers*'(emphasis added). The discussion here may intersect with the ongoing debate on whether the GDPR is appropriate to regulate decentralised technologies. As argued by the Privacy and Data Protection subsection of the German Blockchain Association (Bundesblock),[38] the GDPR may already be *outdated*, since it fails to account for decentralised technologies such as Blockchain. The Bundesblock showed how, in Blockchain ecosystems, different stakeholders intervene in the data processing. The organisation defended that 'although they do not process data, developers of Blockchain protocols also play a role in defining how the data is processed by way of the protocol.' In such scenarios, which also reflect some of the characteristics of Citizen Science networks, a form of indirect responsibility of the developer of the infrastructure could be hypothesized. Clearly, the distribution of responsibilities envisaged by the GDPR reflects a centralised

---

[37] Act n 17 (n 20) art 54.

[38] See Blockchain Bundesverband, 'Blockchain, data protection, and the GDPR' (25 May 2018) <https://www.bundesblock.de/wp-content/uploads/2018/05/GDPR_Position_Paper_v1.0.pdf> accessed 24 August 2018.

model and misses the ability to capture the new parties involved in decentralised data processing structures.

Furthermore, the identification of one or more data controllers in a decentralised system could be based as much on accountability as on actual control. Given that most Citizen Science projects must be coordinated at some point, likely candidates for the role of data controller could be identified near to or at a point of coordination, as long as these are persons with some element of control. This would preserve the decentralized nature that is integral to Citizen Science, but provide identifiable mechanisms for accountability, thus also furthering legitimacy. However, as this would essentially be an artefact in view of the fact that control is decentralised, the option may present downstream complications. Yet it moves toward a signal of accountability that could be crucial to the sustainability of Citizen Science.

In addition, it should be noted that one of the most important responsibilities of the data controller is to ensure that processing of personal data is compliant with the GDPR and, where appropriate, to conduct a data protection impact assessment ex Article 32 GDPR. This also entails instructing every processor on how personal data should be used. In a decentralised system, it is difficult to identify who could or should take on this responsibility. Where this is a research initiative 'from below', as Citizen Science often is, an obvious locus of the data controlling responsibilities may be absent. Nevertheless, when a Citizen Science project collects and processes personal data the fact that there may not be an entity with central control over these activities cannot excuse the project from the legal obligation to have a data controller. Indeed, the participants are data subjects and enjoy the same protections provided by the relevant laws. While they may be able to give explicit consent to the use and processing of personal or sensitive data, they may not consent to the non-existence of a data controller. This is a legal requirement not subject to the choice of the Citizen Science participants.

There must be some locus for accountability for ensuring minimal risk to rights and interests of data subjects. This would include assessing risk and ensuring that adequate safeguards are in place or provided for (eg as instructed to a processor). This presents an interesting dilemma because control may not be the determinative aspect of the entity holding the responsibility of data controller with regard to assessment and minimisation of risk. In fact, the reach necessary to conduct a data impact assessment may be as diffuse as the Citizen Science project itself, presenting a virtually impossible task for any of the lay persons involved in the project. Thus, the possibility of joint data controllers has appeal, but to the extent that this requires a level of expertise, none of the participants, jointly or individually, may be suited to hold this responsibility. For this reason, it may be in the interests of Citizen Science projects to hire a data controller. This presents a clear challenge to the advancement of Citizen Science in that it introduces an external party who then is required to assume a degree of control over others in the project, will incur costs, and mediates the risk that the participants may be willing to assume.

Even with one or more persons taking on necessary tasks of coordination, these individuals, if charged with the role of data controller, may find themselves or the project liable for penalties when the rights of data subjects are not protected in compliance with the GDPR. This could have the undesirable effect of discouraging assumption of this responsibility or the participation in Citizen Science projects altogether. In the case of school projects, the natural candidate for data controller may be the school itself. However, school officials may find this an onerous and poorly rewarded undertaking that may ultimately place the school at risk of financial penalty.

Consequently, like Blockchain, Citizen Science as a decentralised activity may be considerably hindered by the GDPR requirement for a data controller. Where the coordination is done by an institutional entity, like a school, the nature of the responsibilities of data controller may function to discourage school uptake of Citizen Science. Where commercial entities are involved, the solutions are less complex as resources and expertise should be readily available. The biggest challenge comes in initiatives 'from below', where one or more of the participants will be required to assume this responsibility, while not having true *control* or sufficient oversight of the project. This could be regarded as the GDPR's biggest hurdle to Citizen Science. Further research is needed on the limits of the GDPR vis-à-vis decentralised networks, in particular with regards to Citizen Science data collection infrastructures.

Lastly, the GDPR imposes new rules governing how data can be transferred among researchers in different units, labs and across countries (for example the need for data treatment agreement between research centres in different countries). If, on one side, this development can be welcomed as a guarantee for a respect of the EU standards for data processing even when the data are transferred abroad from an EU country; on the other side, such provisions may threaten research collaboration, openness in science and research speed. Provisions regarding necessary assurances and policies such as the 'Privacy Shield' can be expected to assume a similar role in Citizen Science as in other research and, in this way, places Citizen Science on an even playing field with other research involving transborder transfers and protects data protection interests equally.

Considering the preceding discussion, our question on whether data processing requirements under the GDPR represent a possible hindrance to the advancement of Citizen Science for health research can be answered positively. Yet this outcome can be prevented by a work of harmonisation (which is partially ongoing, for example the mentioned ORD pilot) that should orient the future research agenda.

VII. Justifying the Processing on Rights: The Right to Science

Despite possible contrasts between the rights protected by the GDPR and the aims of openness in science, it seems important to stress that Citizen Science is not only a set of practices having to comply with data protection regulations to be legitimate. It is in fact also a 'rightful' practice in itself as it is based on a specific right, as it will be discussed.

First, Citizen Science is a manifestation of a broader discussion on the role of expert and lay knowledge in health practice and research of democratic societies.[39] Citizen Science is also connected to the discussion aimed at challenging the exclusive reliance on expert opinion in solving scientific problems. It is prompted by the scrutiny of the argumentation scheme of expert opinion and the dogmatic appeal to expert knowledge as the ultimate source of authority (the so-called 'argumentum ad verecundiam' by Walton).[40] Walton defines the 'ad verecundiam fallacy' as a 'device to force premature closure of the dialogue' by using expert opinion as the unquestionable end of the discussion.[41] Phenomena such as that of Citizen Science inserts in the broader trend of questioning this 'end of the discussion' by challenging expert authority and showing that other legitimate sources of knowledge are entitled to occupy the research arena.

Within the context outlined, we argue that Citizen Science can be justified on the basis of an overarching right, namely the 'right to science'. The first mention of the right in an official document was in 2012, when Farida Shaheed, Special Rapporteur for the United Nations (UN), submitted a report to the UN Human Rights Council on the scope and application of the right to science.[42] The Special Rapporteur frames the right as that to enjoy the benefits of scientific progress and its applications. Particularly timely for the present research is part (b) of Shaheed's reasoning, where she defines the right as entailing the 'opportunities for all to contribute to the scientific enterprise and freedom indispensable for scientific research' and part (c) where she envisages for 'participation of individuals and communities in decision-making and the related right to information'.[43] The Rapporteur stresses the need for further work aimed at conceptualizing and clarifying the right and inspecting its applications. These recommendations stimulated an academic debate on the need for the UN Committee on Economic, Social and Cultural Rights to officially recognize the right and provide guidance on its implementation for governments and for the scientific community.[44]

The status of Citizen Science may substantially change if the practice is considered as an implementation of a right, the right to science. Under this perspective, the rights recognized by the GDPR, in case posing obstacles to the development of Citizen Science, may need to be balanced with the enjoyment of the right to science. Like other types of scientific research, a right to science or the deeply entrenched right to freedom of expression are not unqualified rights, and must operate within certain parameters.

---

[39] Wiebe E Bijker, Roland Bal and Ruud Hendriks, *The Paradox of Scientific Authority: The Role of Scientific Advice in Democracies* (Wiley 2011).

[40] Douglas Walton, *Appeal to expert opinion* (Penn State Press 1997).

[41] Douglas Walton, *Argument evaluation and evidence* (Cambridge University Press 2016) 13.

[42] Farida Shaheed, 'The right to enjoy the benefits of scientific progress and its applications' (A/HRC/20/26, HRC 2012).

[43] ibid 1.

[44] Audrey Chapman and Jessica Wyndham, 'A Human Right to Science' (2013) 340(6138) Science.

However, just as with other types of research, if the gain for the public interest and for the scientific community appears higher than the possible threats to data protection, a data controller may mediate the GDPR in a risk-based way, and permit research where the benefit outweighs the risks and the risks are found to be acceptable. Under a rights-based reading, such a balancing may be more difficult to defend.

VIII. Conclusion

The article shows that a tension does exist between openness in science and research data protection. This may, in the near future, produce an undesirable disincentive to engage the public in health research and broadly share research data, resulting in a hindrance to the progress of Open Science and Citizen Science. The literature currently available does not target sufficiently the questions raised by this article. Our reflections tried to fill – only in minimal part – this gap by providing an exploratory reflection for further research.

The need to reconcile the aims of openness with that of data protection led us to suggest that there may be the need to investigate the different nature of consent when data are actively and voluntarily disclosed in the framework of Citizen Science projects and when, on the contrary, they are passively gathered through sensors, smartphones and apps. We also suggest that the quality of consent should be deeply scrutinised in cases of Citizen Science projects where market interests may be hidden. Furthermore, we reflected on the limits of data anonymisation in Citizen Science projects requiring for their aims the collection of personal information. A need to reconcile scientific value and data protection emerges. In addition, we observe that initial consent may present limits when participants are engaged in Citizen Science projects spanning over a considerable period and evolving over time. We also pinpoint what can arguably be regarded as the most serious tension between Citizen Science and the GDPR, deriving from the decentralised nature of Citizen Science's activities and the GDPR requirement for a data controller. Lastly, we identify the need to ensure that research collaboration is stimulated and not hindered by the stricter rules under the GDPR for the sharing of research data among different scientific hubs and countries.

These reflections seem particularly timely as they do not only apply to Citizen Science but also, potentially, to other practices and regulations such as those of 'collective' Clinical Trials, as presented in the third scenario analysed. Documents such as the Article 29 Working Party 'Guidelines on Consent under Regulation 2016/679' can provide a blueprint on how to answer these questions by interpretation of the new GDPR, but only to a limited extent. The GDPR in fact cannot be seen as a source of solutions for *all* the described tensions. Further research should look for inspiration in the GDPR but also in other documents, for example those recognizing a right to science and thus providing the foundation of Citizen Science. Future work exploring comparison and harmonisation can ensure that the 'Science with and for Society' goal of the EU do not harm nor are harmed by the new GDPR provisions, in view of achieving an EU scientific agenda close to the people, their interests and their rights.