Computer Law & Security Review 52 (2024) 105911

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

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

Citizen scientists as data controllers: Data protection and ethics challenges of distributed science

Nadezhda Purtova [a,*], Robin L Pierce [b]

[a] Molengraaff Institute for Private Law, Utrecht School of Law, Faculty of Law, Economics and Governance, Utrecht University, Newtonlaan 231 3584 BH Utrecht, the Netherlands
[b] Law School, University of Exeter, United Kingdom

ARTICLE INFO

Keywords:
Citizen science
GDPR
Data controller
Research ethics

ABSTRACT

Citizen-science is a rapidly expanding approach to knowledge production that increasingly involves the collection of personal data in various forms. This processing of personal data invokes relevant data protection laws and, specifically, the designation of data controller, the person(s) or organisation(a) who determine if and how personal data is to be processed and hence are charged with the legal responsibility for compliance with the General Data Protection Regulation (GDPR). Traditionally, in the context of research, professional researchers would be designated controllers, and research participants whose data was processed would be "data subjects" and hence enjoy the GDPR's protections. Yet, citizen-scientists adopt a dual role, acting both as participants and as researchers. This paper maps the implications this dual role has from the perspective of data protection law and research ethics. We explain how the data protection concept of controller has been interpreted very broadly. As a result, in their dual role, citizen scientists can be both data subjects entitled to protection and data controllers, sometimes of their own data, tasked with data protection compliance obligations. If citizen scientists share the objectives of research projects they participate in or co-shape those objectives, it is likely that they – together with the professional researchers - will be considered controllers, and held responsible for the processing of personal data in compliance with the GDPR. The paper discusses how this can affect both the quality of protections provided to participants (including participant-researchers), thus undermining the fundamental goal of research ethics, generally, as well as the practice of citizen science itself. We analyse this question of citizen scientists as data controllers as both a matter of law and research ethics. We conclude with policy recommendations that can be applied both on the level of data protection law (to reconsider how the role of controller is assigned) and research ethics guidelines that should take a nuanced approach to the circumstances of assignment of the status of data controller in citizen science projects as an important step toward responsible and ethical participatory research.

1. Introduction

This paper explores an intersection of citizen science, a relatively new and rapidly expanding approach to generating scientific knowledge, and data protection law to examine the implications of this law for citizen science when the knowledge generation process involves personal data. The particular focus of analysis lies on the concept of controller. Controller is a crucial concept in the basic mechanics of European data protection law. This term refers to the individuals or organisations who determine if, why and how personal data is collected and used. Under the General Data Protection Regulation ('the GDPR'), the data protection principles and data subject rights are effectuated by the corresponding obligations of controllers.[1] Controller has influence on the purposes and means of data processing, unlike a processor, who does not have data processing purposes of its own but acts on instructions. Therefore a controller is designated to bear the principal load of data protection compliance.[2]

For the GDPR to work in practice, there must be at least one

* Corresponding author.
E-mail address: n.n.purtova@uu.nl (N. Purtova).

[1] After the data protection reform and the entry into effect of the GDPR processors also bear some data protection obligations and responsibilities. However, these are minor compared to the share of responsibility carries by the controllers.

[2] "The controller shall be responsible for, and be able to demonstrate compliance with [the general principles of data protection]" Art 5(2) GDPR.

https://doi.org/10.1016/j.clsr.2023.105911

Available online 3 November 2023

controller so that sanctions can always be imposed and remedies sought.[3] Therefore, the concept of controller has been interpreted broadly by the EU Court of Justice in Luxemburg. The purpose of the broad interpretation was to ensure "complete and effective protection" of data subjects.[4] Yet, the data protection scholarship has been criticizing the resulting meaning of controllership as too expansive.[5] That is, that in the effort to ensure that data subjects are absolutely protected, the (reading of) GDPR may be overly-inclusive in defining the parties who bear the responsibility of controller.

This paper subjects the concept of controller to the stress-test of "distributed science". We apply the current broad interpretation of controller in the context of citizen science projects that involve the processing of personal data. We conclude that under some circumstances citizen scientists will likely be regarded as joint controllers together with the researchers who may be leading a project, as well as controllers of their own personal data and that of each other . This may serve to weaken the protection of the citizen scientists as data subjects and thus reaffirms the already voiced concerns about the current case law of the EUCJ on the concept of controller. Furthermore, this broad understanding of "controller" also leads to an outcome that is at odds with fundamental principles of conducting ethical research that holds researchers responsible for the protection of participants. Instead, citizen scientists, who play a double role of "scientists" and research participants, are now charged with responsibility for their own (data) protection under the GDPR. Not only does this lead to a peculiar and counterintuitive outcome. It also yields a concerning result should lay "scientists" fail to comply with the GDPR and be subject to sanctions that are difficult to bear by lay individuals rather than research institutions. While ethics of research literature has addressed some issues of power imbalance and information asymmetries in the context of citizen science, the issues of responsibility for data protection compliance have remained unexplored. This paper aims to address this gap. Among others, we identify three issues that emerge on the point of contact of data protection law and citizen science. First, there is a concern of harm to citizen scientists as research participants. Second, there is a problem of responsibility for harm not corresponding to the actual control over harm. Third, there is a concern of exclusion of underprivileged from participatory science.

To do this we first present some background of citizen science, its benefits, contexts of use and various configurations. We then explain the meaning of controller in data protection law, including the authoritative interpretation by Article 29 Working Party and the European Data Protection Board, advisory bodies in EU data protection under respectively the 1995 Data Protection Directive and the GDPR, as well as the binding case-law of the European Court of Justice. The analysis further proceeds to sketch the implications of the broad interpretation of controller for citizen science from the perspective of data protection law and research ethics. The paper concludes with a summary of findings and some policy recommendations.

2. Background: citizen science

The threshold for generating scientific knowledge is now easier in many ways than it has ever been.[6] The proliferation of digital technologies, mobile devices, as well as the availability of information on the internet has contributed to the expanded reach of laypersons engaging in the production of science.[7] At its core, citizen science, as a form of participatory research democratizes the practice of knowledge production,[8] engages lay people who have not undergone the traditional training in the scientific method to discover or produce science, thus taking it out of the exclusive domain of professional scientists. While citizen science made its early inroads largely in the area of environmental phenomena,[9] the scope of participatory science projects by non-experts has expanded significantly. The prevalence of digital technologies has allowed for broader participation in the "self-quantification"[10] phenomenon in which individuals can keep track of information pertaining to themselves. This not only allows for tracking data relevant to one's own health, but also enables laypersons to aggregate their personal information to generate new insights pertaining to relevant communities, locations, health status, and effects to answer questions that they consider to be important. This has led to an increase in participatory science in such fields as epidemiology, genetics and genomics,[11] and specific disease-related research. Citizen science is practices in other contexts, too, such as environmental activism.[12]

Citizen science can take a range of forms, including bottom-up, with questions, goals, and methods originating from the community of laypersons as well as research institution-led initiatives that recruit lay participants to collect or contribute data, with numerous variations along this spectrum. A study could involve any of these forms and seek to collect a variety of personal data, such as citizen scientists' reports of their experiences and observations, personal, as well as facilitated by technology such as sensors, genetic information, bio-samples, results of various types of bio-tracking, e.g. heart rate, weight, pulse, and other health or fitness indicators. There is recognized value in these kinds of

[3] Christopher Millard 'At this rate, everyone will be a [joint] controller of personal data!' (2019) 9(4) International Data Privacy Law 217.

[4] E.g. *Google Spain SL, Google Inc v Agencia Española de Protección de Datos and Mario Costeja González* Case C-131/12 [2014] ECLI:EU:C:2014:317 [34], [53]; *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH (Wirtschaftsakademie)* Case C-210/16 [2018] ECLI:EU:C:2018:388 [28].

[5] Millard (n 3). For more recent critique see Michele Finck 'Cobwebs of control: the two imaginations of the data controller in EU law' (2021) 11(4) International Data Privacy Law 333–347.

[6] Barbara Prainsack. 'Understanding participation: the 'citizen science'of genetics.' In Barbara Prainsack, Silke Schicktanz, Gabriele Werner-Felmayer (eds.) *Genetics as social practice* (Routledge, 2016), 163-180.

[7] J. Patrick Woolley, Michelle L. McGowan, Harriet J.A. Teare, Victoria Coathup, Jennifer R. Fishman, Richard A. Settersten, ... & Eric T. Juengst 'Citizen science or scientific citizenship? Disentangling the uses of public engagement rhetoric in national research initiatives' (2016) 17(1) BMC Medical Ethics, 1-17; Amelia Fiske, Barbara Prainsack & Alena Buyx, 'Meeting the needs of underserved populations: setting the agenda for more inclusive citizen science of medicine.' (2019) 45(9) Journal of Medical Ethics, 617-622.

[8] J. J. Schensul 'Democratizing science through social science research partnerships' (2002) 22(3), Bulletin of Science, Technology & Society 190-202.

[9] See e.g. Michiel van Oudheusden, Joke Kenens, Go Yoshizawa, & Nozomi Mizushima. 'Learning from Citizen Science after Fukushima: Probing the Role and Potential of Citizen Science in Nuclear Science and Technology Governance in Japan and Belgium.' Workshop report (SCK CEN 2019) available online https://researchportal.sckcen.be/en/publications/workshop-report-learning-from-citizen-science-after-fukushima-pro last accessed 2 October 2023.

[10] Dawn Nafus & Jamie Sherman, 'Big data, big questions| this one does not go up to 11: the quantified self movement as an alternative big data practice.' (2014) 8 International journal of communication, 11; Manal Almalki, Kathleen Gray & Fernando Martin Sanchez. 'The use of self-quantification systems for personal health information: big data management activities and prospects.' (2015) 3(1) Health information science and systems, 1-11.

[11] Stacey Kuznetsov, Aniket Kittur, & Eric Paulos, 'Biological citizen publics: Personal genetics as a site of public engagement with science.' (June 2015) In Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition available online at https://dl.acm.org/doi/abs/10.1145/2757226.2757246 last accessed 2 October 2023, 303-312.

[12] Anna Berti Suman. *Sensing the risk: A case for integrating citizen sensing into risk governance*, (Tilburg, Open Press TiU 2020), available online at https://digi-courses.com/openpresstiu-sensing-the-risk/ last accessed 2 October 2023.

participatory research that extends beyond the political consideration of democratization. In principle, citizen science can enhance the scientific enterprise by conceivably accessing information that may be difficult to obtain by traditional research institutions because of access, lack of priority, or lack of funding. Thus, the conduct of research by non-professionals can serve important ends.[13]

Issues of responsibility in the conduct of research are complex. A completely bottom-up research project would arguably locate responsibility for risks with the lay researchers since there is no outside institution or other actor on whom responsibility could be ascribed. However, once the form starts to move further along the spectrum to involve a research institution, the assigning of responsibility is arguably less clear, absent specific agreement.[14] That is, where a research institution or professional researchers collaborate with laypersons to develop the purpose, means, or determine access to data collected, these actors all operate in a "directive" or "responsible" capacity. It is precisely this shared decision-making that characterizes much of participatory research, and is heralded as among its benefits, that triggers shared responsibility, including responsibility for the handling of any personal data that might be collected. This raises the question of whether a citizen scientist, in addition to being a participant, is also a controller under the meaning of the GDPR.

3. The meaning of controller under the GDPR

Data protection law knows three key actors: data subject, a living natural person to whom personal data relates, who would potentially suffer injury should data protection law be violated, and who thus enjoys data protection rights; data controller, a natural or legal person who alone or jointly with others determines the means and purposes of data processing and carries the main load of the data protection obligations,[15] and processor, a natural or legal person who processes personal data on behalf of a controller.[16] Traditionally, the stakes in establishing the status of a controller or processor are high, since the status of a controller comes with the data protection obligations, and the boundaries between the two are blurry. The sections that follow review how the concept of controller is understood in the authoritative opinion of the Article 29 Working party and in the case law of the EU Court of Justice, pre- and post its Judgment in *Fashion ID*.

3.1. Article 29 working party and European data protection board guidelines

The Article 29 Working Party, the former EU advisory body under the 1995 Data Protection Directive, produced guidelines for determining the status of controller and processor (WP169).[17] While not formally binding, the guidelines in practice bear undeniable persuasive authority and were a primary reference point for compliance with the data protection law. The WP169 has retained its significance also after the GDPR came into effect in place of the 1995 Directive, since the definitions of a controller and processor did not undergo any significant changes between the two legislative instruments. According to WP169, the dividing line between controller and processor lies along the *factual* influence over purposes and means of processing, arising out of explicit or implicit legal competence (e.g. a competence conferred by a statute vs a competence necessary to fulfil explicit authority, but not explicitly named), contractual arrangements, but also, and importantly, out of other circumstances determining the factual ability to determine the purposes and means of processing, even when these factual circumstances contradict the statutory or contractual arrangements.[18]

The European Data Protection Board, which replaced the WP29 after the GDPR came into effect, issued own guidelines on the concepts of controller and processor under the GDPR.[19] The guidelines – although update the WP169 - do not deviate from the WP29 opinion in the main lines of interpretation, emphasising the importance of the factual influence over the purposes and means of processing.[20]

3.2. The EU court of justice case law – pre-Fashion ID

The EUCJ's line of case law on controllership started in 2014 with its decision in *Google Spain.*[21] The question was if a search engine operator should be considered a controller with regard to personal data published on third party websites and processed in the context of activity of a search engine. The Court ruled that the search engine operator determines the purposes and means of data processing in the context of that activity[22] and thus is a controller. To rule otherwise on the ground that the search engine operator does not exercise control over personal data published on the websites of third parties would be contrary to the objective of the relevant provision of the Directive "to ensure, through a broad definition of the concept of 'controller', effective and complete protection of data subjects."[23] A decisive criterion determining the status of a controller was that the data processing carried out in the context of its activity "can be distinguished from and is additional to that carried out by publishers of websites,"[24] as the activity of search engines plays "a decisive role in the overall dissemination of those data" in that it makes the data searchable to each user.[25] The main *Google Spain* legacy relevant for the concept of controller is that the concept ought to be interpreted broadly, in light of the objective of the data protection law to ensure effective and complete protection of data subjects.

In the 2018 *Wirtschaftsakademie* the Court continued developing its caselaw on controllership. Wirtschaftsakademie offered some educational services via its fan page set up on Facebook. It was established that as a part of non-negotiable conditions of use set by Facebook, administrators of fan pages receive anonymous statistical information on the page visitors collected by means of cookies installed on visitors' devices, containing a unique user code, making the data processed personal (the Facebook Insights function).[26] The page visitors were not notified of the placement and functioning of the cookie and subsequent data

[13] See David B. Resnik, Kavin C. Elliott, & Aubrey Miller. 'A framework for addressing ethical issues in citizen science.' (2015) 54 Environmental Science & Policy, 475-481.

[14] See Alena Buyx, Lorenzo Del Savio, Barbara Prainsack, & Henry Völzke. 'Every participant is a PI. Citizen science and participatory governance in population studies.' (2017) 46(2) *International Journal of Epidemiology,* 377–384.

[15] See Art 5(2) GDPR prescribing that "[t]he controller shall be responsible for, and be able to demonstrate compliance with, [general data protection principles]".

[16] As defined in Art 4(1), (7) and (8) GDPR respectively.

[17] Article 29 Working Party "Opinion 1/2010 on the concepts of 'controller' and 'processor'" WP169, adopted on 16 February 2010.

[18] WP169, 8-9 (e.g. "[b]eing a controller is primarily the consequence of the factual circumstance that an entity has chosen to process personal data for its own purposes.")

[19] European Data Protection Board "Guidelines 07/2020 on the concepts of controller and processor in the GDPR" adopted on 7 July 2021.

[20] Ibid, e.g. p. 3

[21] *Google Spain SL, Google Inc v Agencia Española de Protección de Datos and Mario Costeja González* Case C-131/12 [2014] ECLI:EU:C:2014:317.

[22] *Google Spain* 33.

[23] *Google Spain* 34; see also 58 where the Court says that "ensuring effective and complete protection of the fundamental rights and freedoms of natural persons, and in particular their right to privacy, with respect to the processing of personal data" is an objective of the 1995 Directive and not only of the definition of controller.

[24] *Google Spain* 35.

[25] *Google Spain* 36.

[26] Case C-210/16 *Unabhangiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH* (Judgment, 5 June 2018) 15.

processing, which was in violation of the data protection rules.[27] The national courts went back and forth between ruling Wirtschaftsakademie a controller jointly with Facebook, or Facebook alone,[28] since the latter "alone decided on the purpose and means of collecting and processing personal data used for the Facebook Insights function" while the former only received anonymous statistics.[29] The Court ruled in favour of considering the Wirtschaftsakademie a controller, jointly with Facebook. The Court reaffirmed the purpose of data protection law "to ensure a high level of protection of the fundamental rights and freedoms of natural persons"[30] and cited the need to interpret the meaning of controller broadly in view of the goal of the definition to ensure effective and complete protection of the data subjects.[31] The Court acknowledges that while Facebook is indeed the actor "primarily determining the purposes and means of processing"[32], using it to serve its system of advertising,[33] Wirtschaftsakademie itself "must be regarded as taking part" in determining the purposes and means of data processing, and hence a joint controller.[34] This follows from the examination of the contribution of Wirtschaftsakademie "to determining, jointly with Facebook ..., the purposes and means of processing"[35]. Any administrator of a fan page on Facebook concludes a contract with Facebook and thereby subscribes to the conditions of use, including the cookie policy.[36] By creating a fan page, its administrator enables Facebook to install cookies on the devices of the page visitors, including those without a Facebook account.[37] When setting up a fan page its administrator for its own objectives of managing and promoting its activities, can set parameters determining production of statistic,[38] e.g. request for demographic and other data of its target audience.[39] Finally, while the administrator has no access to personal data collected by Facebook and only receives anonymised audience statistics, one does not have to have access to the personal data in order to be a controller.[40] Importantly, the Court – for the first time - brought up the issue of distribution of responsibility between joint controllers. It ruled that responsibility between joint controllers does not have to be equal, but needs to be assessed on a case by case basis,[41] since joint controllers may be involved at different stages of processing and to different degrees.[42] This issue played an important role in its subsequent jurisprudence.

In the same year the Court had to deal with another case regarding the definition of controller. The relevant dispute in the *Tietosuojavaltuutettu v Jehovan todistajat* case concerned, among others, whether or not The Jehovah's Witnesses Community (*Jehovan todistajat*), even though it had no access to the relevant data, should be regarded as a joint controller along with its members who, in the course of their door-to-door preaching, made notes containing names, addresses and other personal data relating to the people they visited. The Court answered in the affirmative. It reaffirmed that in view of the objective to provide effective and complete protection of the data subjects the meaning of controller should be construed broadly, and that joint responsibility does not mean equal responsibility.[43] Similarly to WP29 position, the Court noted that the determination of the purposes of processing does not have to be in the form of written guidelines or instructions.[44] The Court restated that "[t]he joint responsibility … does not require each of [the multiple controllers] to have access to data".[45]

According to the Court, while the Jehovan todistajat members and not the Jehovan todistajat itself are deciding if and when they collect the data, the preaching is "organised, coordinated and encouraged" by the Community.[46] Data collected serves as a memory aid for further preaching. The community members engage in preaching for the purposes of the Jehovan todistajat. The Jehovan todistajat is also generally aware of the data processing taking place. It organizes and coordinates the preaching,[47] and hence "encourages its members who engage in preaching to carry out data processing".[48] Thus, "by organising, coordinating and encouraging the preaching activities of its members, … [the Community] participates, jointly with its members … in determining the purposes and means of processing"[49] and should be considered a controller.[50]

The resulting approach of the Court to understanding controllership has been described as "sweeping",[51] potentially making everyone a controller,[52] and thus laden with undesirable consequences,[53] including the "actual impossibility for a potential joint controller to comply with valid legislation".[54]

3.3. Fashion ID

In *Fashion ID,* the latest occasion where the Court dealt with the meaning of controllership, the Court tempered its broad approach somewhat. The case is significant for the issue of the degree of responsibility of joint controllers first raised in *Wirtschaftsakademie*. It involved a clothing retailer who placed the Facebook "like" button on its website, resulting in personal data of the website visitors being transmitted to Facebook.[55] The question was if the operator of a website that embeds a social plugin causing the personal data of the visitor to be transmitted to e.g. Facebook is a controller, even though this operator is unable to influence the processing of the data transmitted to that provider as a result.[56] The Court again answered affirmatively. The Court restated the existing case law on controllership.[57] It reaffirmed that

[27] *Wirtschaftsakademie* 15, 16.
[28] *Wirtschaftsakademie* 17-23.
[29] *Wirtschaftsakademie* 21.
[30] *Wirtschaftsakademie* 26.
[31] *Wirtschaftsakademie* 27, 28.
[32] *Wirtschaftsakademie* 30.
[33] *Wirtschaftsakademie* 34.
[34] *Wirtschaftsakademie* 39, 42.
[35] *Wirtschaftsakademie* 31.
[36] *Wirtschaftsakademie* 32.
[37] *Wirtschaftsakademie* 35.
[38] *Wirtschaftsakademie* 36.
[39] Wirtschaftsakademie 37.
[40] *Wirtschaftsakademie* 38.
[41] *Wirtschaftsakademie* 43.
[42] *Wirtschaftsakademie* 43.
[43] C-25/17 *Jehovan todistajat* (Judgment, 10 July 2018, ECLI:EU:C:2018:551) 66.
[44] *Jehovan todistajat* 67.
[45] *Jehovan todistajat* 69
[46] *Jehovan todistajat* 70
[47] *Jehovan todistajat* 71
[48] *Jehovan todistajat* 72
[49] *Jehovan todistajat* 73
[50] *Jehovan todistajat* 75
[51] Case C-40/17 *Fashion ID* (Opinion of Advocate General Bobek) ECLI:EU:C:2018:1039, 72.
[52] e.g. Millard (n 3).
[53] e.g. Opinion of AG Bobek (n 51), 73 et seq, but also Lilian Edwards, Michele Finck, Michael Veale, and Nicolo Zingales 'Data subjects as data controllers: a Fashion(able) concept?' (2019) Internet Policy Review published on 13 June 2019 available online at https://policyreview.info/articles/news/data-subjects-data-controllers-fashionable-concept/1400 accessed 2 October 2023, pointing to the risk of considering data subjects as controllers.
[54] Opinion of AG Bobek (n 51), 84.
[55] *Fashion ID* 26, 27.
[56] *Fashion ID* 64.
[57] The Court referred to the data protection law objective to ensure a high level of protection of the fundamental rights and freedoms and the broad interpretation of controller in view of the effective and complete protection of data subjects. Several actors can be controllers and bear data protection obligations at the same time. The status of a controller does not require access to personal data, and a person with influence over data processing for his own purposes may be considered a controller. *Fashion ID* 65-68.

multiple controllers can be involved in different stages of processing and to different degrees, and hence the joint responsibility does not mean equal responsibility and the level of liability of each controller has to be assessed given all the circumstances of each case.[58] The case's significance lies in how it developed the latter point.

The Court appears to have seen the problems with the broad meaning of controller that resulted from its previous caselaw and pursued the path of narrowing it down, as laid out in the opinion of AG Bobek. The Court noted that the meaning of data processing includes a variety of operations,[59] and that a processing instance may consist in one or a number of operations, relating to one of the different processing stages.[60] An actor "may be a controller, … jointly with others only in respect of operations … for which it determines jointly the purposes and means. By contrast, … that natural or legal person cannot be considered to be a controller … in the context of operations that precede or are subsequent in the overall chain of processing for which that person does not determine either the purposes or the means."[61] In the case at hand, the Court considered that Fashion ID was only able to jointly determine means and purposes of processing for the stage of collection and disclosure by transmission of the personal data of visitors to its website, and not for the processing by Facebook that occurred later. Hence, while Fashion ID is certainly a joint controller for the operation of data transfer, it cannot be considered to be a controller in respect of the subsequent operations, such as processing by Facebook for the purposes of advertising.[62] This "chain of processing" or "processing stages" approach to joint controllership has narrowed down the application of the concept of controller in the context of complex data processing involving multiple actors, such as social networks and digital service providers. Yet, it has already received criticism for "creating more problems than it solves"[63] by "losing sight of the bigger picture", in particular, of "the societal risks posed by complex, networked, personal data processing systems such as … Facebook."[64] Indeed, the risks of data processing in such systems are more than the sum of the risks of the individual stages of processing, yet the responsibility of (joint) controllers, a.o. to inform about those risks, as well as provide for data subjects' protection, are reduced to the latter.[65]

4. Data protection context: controllership in citizen science

4.1. Citizen scientists as controllers (of their own data)

How is the role of a controller – given the current state of law – assigned in the context of citizen science, and in particular, what role does a citizen scientist have?

As discussed earlier, the citizen science research can take a number of configurations, ranging from absolutely centralized to absolutely decentralized, and the many degrees of de-centralization in between. In the scenario of absolute centralization, the "professional scientists" lead and the citizen scientists follow their instructions and have no influence on the course of a study, including the purposes and means of (personal) data processing. The opposite scenario is of absolute decentralization, where citizen scientists are the true drivers of research, determining among others the purposes and means of data processing, and the professional scientists are not involved. Based on the current state of law on the concept of controller, citizen scientists will likely be considered controllers in all these scenarios, although the range of stages of processing for which they are responsible may differ and be more limited in some circumstances. This being said, the concrete outcomes will depend on the circumstances of each particular case.

In all contexts, the citizen scientists will be considered joint controllers for the data processing in the entire project if, as commonly practiced and required by the standards of ethical research with human participants in terms of responsibility of researchers for protections,[66] at the stage of being recruited they are informed about the purposes of the project and data processing and agree with them, similar to a Facebook page administrator who subscribes to Facebook's conditions of use, including the cookie policy.[67]

In decentralized distributed science projects, citizen scientists are by design given a real role and influence over the project design, for instance when research is closely linked to their interests and living environments, e.g. research into pollution and can benefit from the citizen scientists' knowledge of the situation. In this case, citizen scientists are given influence over purposes and sometimes means of processing personal data. For instance, they may co-determine what types of data will be collected and participate in discussions about and co-steer the purposes of data processing.

Moreover, as it was the case with the search engine provider in *Google Spain*, Facebok page administrators and administrators of a website with a Facebook "like" button in *Wirtschaftsakademie* and *Fashion ID,* when joining distributed science projects, citizen scientists will often have their own purposes different from those of the professional scientists, e.g. use gathered data to understand a phenomenon of relevance in their personal lives like the environmental conditions or online tracking,[68] support their position or defend their interests in their relations with public authorities,[69] and others. Even when a project to a larger or lesser degree is coordinated and steered by a professional scientist, and the factual influence over the purposes of data processing is varying, but present, they will likely be joint controllers together with the coordinating professional scientists (if involved) and their fellow citizen scientists.

Even in case of absolutely centralized citizen science project, it is fairly certain that the professional scientists who control the project, including determining the purposes and means of processing personal data, will under some circumstances not be the only controllers and citizen scientists will sometimes be considered joint controllers too. This will be the case if they have their own purposes served by processing personal data in the project as described earlier, e.g. investigating and documenting pollution and its impact, e.g. on the health of the citizen scientists themselves, as well as other community members. This will also most likely be the case where citizen science is a form of citizen activism, and its results serve purposes of civil initiatives pursued by the citizen scientists.

Even more so, in all these scenarios, where the personal data processed is their own, citizen scientists will be data subjects and (joint) controllers of their own data at the same time. While this may seem counterintuitive, there is nothing in the GDPR that explicitly prevents

[58] *Fashion ID* 70.
[59] *Fashion ID* 71.
[60] *Fashion ID* 72.
[61] *Fashion ID* 74.
[62] *Fashion ID* 75-76.
[63] Rene Mahieu and Joris van Hoboken, 'Fashion-ID: Introducing a Phase-Oriented Approach to Data Protection?' (2019). European Law Blog, available online at https://ssrn.com/abstract=3548487 ;last accessed 2 October 2023, 4.
[64] Mahieu & Van Hoboken (n 63) 3.
[65] Mahieu & Van Hoboken (n 63) 3. See also Edwards, Finck, Veale, and Zingales ().
[66] See e.g. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. DHEW Publication No. (OS) 78-0012; World Medical Association, Declaration of Helsinki: ethical principles for research involving human subjects (2008) (in the case of health research); and Resnik et al, (n 13) (for ethical framework drawing on established ethical principles)..
[67] *Wirtschaftsakademie* 32
[68] e.g. CSI-COP project https://cordis.europa.eu/project/id/873169
[69] e.g. Berti Suman (n 12).

data subjects from being controllers with regard to their own data. The former Article 29 Working party has alluded to the possibility of data subjects being controllers with respect to their own data in the context of mobile health apps,[70] and the French data protection authority CNIL has explicitly recognized such a possibility in the context of blockchain.[71] The ethical implications of a citizen scientist being a data subject and controller with regard to his or her data will be explored further in the paper. Here it suffices to say that this state of affairs may be morally quite controversial, as the aspiration behind the data protection law is to protect the data subject from harm. The dual role of participant and researcher, shifts the responsibility for protections to the person to be protected.

Following the "stages of processing" approach in *Fashion ID*, the degree of responsibility of citizen scientists may be limited in a few cases. For instance, in cases where professional researchers are subject to obligations not applicable to citizen scientists, such as archiving raw research data to enable verification of study results and ensure scientific integrity,[72] and process personal data for these purposes, the citizen scientists have no influence over such processing which has to take place regardless of their wishes, and thus will likely not be considered controllers for this processing. In the rare cases of the fully-centralized citizen science projects where the involved citizen scientists do not have any influence over the purposes and means of data processing within the project, e.g. they were not informed of the project purposes, and have no interest in the study outcomes, they will likely be considered controllers only for the stage of transferring personal data to the professional researchers, as they enabled data transfer to the professional researchers.[73] They will be fully responsible for the data processing in the project when their interests align with the purposes of the project, as described above.

Finally, the so-called "household exemption" under Art. 2 GDPR which often exempts from the GDPR the data processing "by a natural person in the course of a purely personal or household activity" will not apply here and create no exceptions for the citizen scientists. This is because to qualify for such an exemption, the processing must be carried out "in the course of private or family life of individuals," e.g. the data must not be shared with an indefinite number of people,[74] and processing must not be "directed outwards from the private setting of the person processing the data".[75] Data processing by the citizen scientists does not meet either requirement, as participating in research projects does not fall within their private or family life, and takes place outside of the private settings.

4.2. Data protection implications

The major implication of assigning the role of controller to the citizen scientists is that they will bear responsibility for compliance with data protection law, including respecting data protection rights and bearing data protection sanctions, jointly with the professional researchers if they are involved, and their fellow citizen scientists. This creates serious difficulties, both from the perspective of the practicality of compliance as well as the perspective of protection of a citizen scientist as a data subject.

Among others, in his capacity of controller, a citizen scientist is expected to determine an appropriate legal ground of data processing (e.g. if a legitimate interest would be appropriate or the affected rights and interests of the data subjects outweigh) and exercise complex balancing exercises and make normative calls to establish, e.g. if the data processing is fair, lawful, and proportionate to what is necessary for the purpose of processing, if that purpose is legitimate. The citizen scientists and professional researchers as joint controllers have to agree on and inform the data subjects of the division of their respective compliance responsibilities, in particular as regards the data subject information rights, and their roles and relationships of the joint controllers vis-à-vis data subjects.[76] These are not easy tasks, particularly from the perspective of the citizen scientists. While the professional researchers – although often ignorant of their data protection responsibilities – are expected to be aware of this aspect of their profession, especially when personal data processing is a core part of their research, this is less so for the citizen scientists. Indeed, while acting outside of household or domestic context, they are not professional players on the research field and lack the necessary resources, institutional support and expertise. Their knowledge of data protection requirements and especially of their role as a controller responsible will more often then not depend on the information provided by the professional scientists, e.g. in the course of recruitment. Their compliance effort will be heavily defined by support of the professional researchers. Given the complexity of this data processing context and the slow uptake of the data protection expertise outside of the big tech, it is highly likely that such information and support will be lacking or inadequate. The situation is further exacerbated where citizen scientists do not have access to personal data and hence are even less aware of the nature of data processing and the associated obligations, and yet are considered controllers,[77] e.g. the personal data of their fellow research participants directly supplied to the professional researchers. What results is high expectations of data protection on paper met by the reality of the non-professional actors such as citizen scientists unable to deliver on the expectations in the context of the highly-complex data protection regime where compliance demands institutional effort, resources and expertise.[78]

At the same time, even in cases where the professional scientists do provide the necessary information and support and take the bulk of the data protection obligations on their account, a data subject may exercise his or her rights with regard to any of the joint controllers, regardless of any agreements made.[79] Among others, a data subject has a right to claim compensation of damages suffered as a result of GDPR infringements against each and any joint controller, and one joint controller can be held liable for the entire damage,[80] with a right, upon payment of full compensation, to claim back from other controllers their respective parts.[81] Paying such a compensation for the damage as a

[70] The Working Party suggested that when health data is processed by a health app on the user's device, the data protection law does not apply to the user since he falls within the household exemption, while if the data is processed remotely, the provider of the app processing data for his own purposes does not fall under such exemption. This may be construed as recognition that were the household exemption not applicable, the user would be a controller of his own data. Article 29 Working Party, 'Letter to the Director of Sustainable and Secure Society Directorate of the European Commission,' published 5 February 2015, Annex I, available online <https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20150205_letter_of_the_art_29_wp_on_sh_transfers_tools_and_surveillance_annex_1.pdf > last accessed 5 may 2021, p. 5 and fn 6

[71] CNIL, Blockchain et RGPD : quelles solutions pour un usage responsable en présence de données personnelles? 24 September 2018, available online https://www.cnil.fr/fr/blockchain-et-rgpd-quelles-solutions-pour-un-usage-responsable-en-presence-de-donnees-personnelles, accessed 5 May 2021, cited in Edwards, Finck, Veale, and Zingales (n 53).

[72] According to the Dutch Code of conduct, raw scientific data has to be archived for a minimum of 10 years.

[73] *Wirtschaftsakademie* 35.

[74] Case C-101/01 *Lindqvist* (2003) EU:2003:596, 43.

[75] Case C-12/13 *František Ryneš v Úřad pro ochranu osobních údajů* EU:C:2014:2428, 33.

[76] Art 26 (1) and (2) GDPR.

[77] *Wirtschaftsakademie* 38.

[78] Christopher Kuner, Fred H. Cate, Christopher Millard, Dan Jerker B. Svantesson, 'The business of privacy' (2013) 3(2) International Data Privacy Law, 65–66, https://doi.org/10.1093/idpl/ipt003.

[79] Art 26(3) GDPR.

[80] Art 82(4) GDPR.

[81] Art 82(5) GDPR.

result of processing involving many fellow citizen scientists and professional scientists and research institutions, especially when the violations are serious and damages significant, may appear impossible for a single citizen scientist, as well as disproportionate in light of the unequal relationship between the lay citizen scientists and professional researchers. This provision was clearly written with professional controllers in mind and did not account for the situations where complex distributed data processing will involve controllers who are regular citizens. While data subjects wishing to claim such compensation may strategically opt to go after the professional researchers and the research institutions with deeper pockets,[82] it will provide little relief in case a citizen scientist is the only controller known to the data subject. The only exemption from the liability is if the controller "proves that it is not in any way responsible for the event giving rise to the damage."[83] The extent to which this will mitigate a citizen scientist liability depends on what "responsible for the event giving rise to the damage" means. The EUCJ case law on controllership seems to equate responsibility for processing to the status of a controller.[84] Thus, it seems impossible for a citizen scientist considered a controller with regard to a certain data processing operation to obtain exemption from liability in case such operation results in damages.

Finally, considering a citizen scientist as a joint controller with regard to his or her own data raises a question if this broad reading of the concept indeed leads to "effective and complete protection of data subjects" as intended.[85] Indeed, a status of a controller implies that the data subject carries at least some of the data protection obligations towards himself (in case of joint control, responsibilities for the obligations, in particular, to inform data subjects, should be divided between the controllers). As far as obtaining compensation for damages caused by data protection violations, citizen scientists who are controllers with regard to their own data jointly with professional scientists should still be able to obtain some compensation from the professional scientist. This is because the liability under the GDPR is joint and several (Art 82 (4) GDPR), i.e. applies to the professional scientist also. Yet, the liability of the latter may be limited if the professional scientist demonstrates he was not responsible for the event leading to the damage (Art 82(3) GDPR), e.g. if some of the "fault" may be attributed to the citizen scientist himself. Considering a data subject as a joint controller will likely enable the professional scientists avoid some responsibility for the consequences of their professional activity, which, from the perspective of the ethical standards of research, should be their concern, as discussed next.

5. Implications for ethical research

This section exploress the problems of research ethics that occur as a result of the broad approach to the notion of controller in the context of participatory science. Traditionally, research ethics is based on a clear distinction between a researcher and a research participant. Research ethics has as its main focus the protection of research participants from risks of harm, and charges the researcher with providing these protections. Generally, this responsibility is allocated to the "investigator"[86] researcher or physician or scientist in the case of medical research.[87] This responsibility essentially aims to ensure the integrity of the research.[88] Citizen science challenges this dual distinction, as citizen scientists assume both roles. The question of whether a citizen scientist is a controller and, as such, bears corresponding responsibilities arises from the very aspect of participatory research that delivers many of its most valued benefits – the layperson as both participant and researcher. However, research ethics has yet to provide for how to fulfil necessary obligations regarding protections when participants assume this dual role.[89]

Scholarship in the field of research ethics has already addressed some issues raised by the shift in the role of research participants that participatory science brings. Resnik and colleagues have devised a framework for addressing ethical issues in citizen science that aims to address the imbalance in expertise, education, and relevant knowledge about research practices between professional and lay researchers.[90] This useful framework targets 1) data quality and integrity 2) data sharing and intellectual property 3) conflict of interest and 3) exploitation.[91] This list captures many of the otherwise unaddressed risks associated with the conduct of citizen science research. However, it does not adequately reflect the ethics challenges posed by the status of a citizen scientist as a (joint) controller. Some researchers have compared the role of participant-researcher to that of "research assistants"[92] which recognizes the dual role of participant and agent. This, of course, raises issues of unequal positioning that may be captured by Resnik et al's issue of exploitation. While exploitation casts a wide net on practices that take advantage of unequal positioning, the specific issue of responsible parties in the form of controller raises a different type of dilemma for actors engaging in participatory research.

First and foremost, there is a concern of harm to citizen scientists as research participants. The assignment of joint controller to citizen scientists in circumstances of shared purposes essentially imposes responsibility for the protection of harms arising from the processing of personal data on persons who may be least knowledgeable about what those obligations are or how to execute them in compliance with the law. There are two disturbing ethical reverberations from this. One, the lack of knowledge can result in an absence of protections for citizen scientists in that if a lay person is the responsible party for the handling of data in compliance with the GDPR, but does not know the law or does not sufficiently understand how to apply this law, the personal data of participants may not be properly protected (including their own). This consequence may actually serve to negate the most fundamental purpose of research ethics of protecting humans from harm as a result of participating in research. Two, citizen scientists may be held liable for damages caused by data protection violations which can also amount to harm.

Second, there is a problem of responsibility for harm not corresponding to the actual control over harm. Given the distributed nature of many types of citizen science projects, the assignment of joint controller role may fall where the lay researchers actually have very little or no control. This not only can result in diminished protections for participants (or participant-researchers), but also results in imposition of responsibility on persons who have no means by which to actively assume that responsibility. In other words, mere participation in a project with certain shared features (e.g. purpose) can result in the imposition of liability on a lay participant who may have no knowledge of the scope of

[82] Indeed, those institutions will most likely be joint controllers as per *Jehovan todistajat.* Although they do not directly determine the purposes and means of each data processing operation taking place within it, have research as core of their business, encourage and organize it and are aware that it often involved personal data processing (*Jehovan todistajat* 70).

[83] Art 82(3) GDPR.

[84] e.g. *Wirtschaftsakademie* 43, *Fashion ID* 70.

[85] *Google Spain* 34; see also 58 where the Court says that "ensuring effective and complete protection of the fundamental rights and freedoms of natural persons, and in particular their right to privacy, with respect to the processing of personal data" is an objective of the 1995 Directive and not only of the definition of controller.

[86] See e.g. The Belmont Report (n 66).

[87] World Medical Association, Declaration of Helsinki (n 66).

[88] Lisa M. Rasmussen, 'Confronting research misconduct in citizen science.' (2019) 4(1) Citizen Science: Theory and Practice, 1.

[89] K.M. Oberle, S.A. Page, F.K. Stanley, & A.A. Goodarzi, 'A reflection on research ethics and citizen science.' (2019) 15(3-4) Research Ethics, 1-10.

[90] Resnik et al (n 13).

[91] Resnik et al (n 13).

[92] Kuznetsov, Kittur & Paulos (n 11).

his or her obligations or the ability to execute them effectively. The flip side of this is that where this takes place in the context of a professional researcher-layperson collaborative research project, the professional researchers may escape sole liability for a responsibility for which they alone may have the ability to bear effectively.

Third, there is a concern of exclusion of underprivileged from participatory science. The characterization of citizen scientists as joint controllers may ultimately have an impact on who participates in participatory research. That is, persons least likely to be able to bear or execute the responsibility of a joint controller in terms of sufficient knowledge about the law and what it requires of processing in a particular project or financial wherewithal to bear any sanctions ensuing from breach, may refrain from participation. This privileges the elite in society, already a concern in citizen science participation[93], in an enterprise that claims to "democratize" the generation of knowledge. This is problematic on multiple levels, not least for this undermining effect, but also for the robustness of research results, to the extent that a research-participant base does not sufficiently represent the relevant demographic or the demographic who may be affected by any research results.

The existing standards of ethical research need to be updated to be able to adequately address these dilemmas.

6. Conclusions and policy recommendations

In this paper we examined how the data protection law and research ethics interact in the context of participatory science and assigning responsibilities for protection of personal data between professional researchers and citizen scientists. We have demonstrated that – in order to ensure complete and effective protection of the data subject - the meaning of controller in data protection law has been construed very broadly both in non-binding authoritative interpretations and binding case-law. Even a small part in determining if and how personal data will be collected and used render an individual or an organisation a (joint) controller. This state of affairs has already received criticism in the data protection field for making "everyone a controller" and assigning responsibilities for processing personal data where they do not always belong or are practically possible to respect, e.g. in distributed computing or interactions of individuals and small organisations with technology giants on their platforms. The context of citizen science presents yet another case where this unbalanced distribution of data protection responsibilities manifests itself and creates dilemmas of research ethics so far not addressed by the ethics literature or guidelines.

We identified three such dilemmas. First, as long as it is possible to designate citizen scientists as (joint) controllers, participatory science may cause harm to them. Assigning status of a data controller to a citizen scientist who is a lay player in the field of research makes a citizen scientist responsible for protection of his or her own rights and thus jeopardizes the level of protection of the fundamental right to data protection. This may also lead to the citizen scientist being held liable for data protection violations when other data subjects are harmed, which is another form of harm. Second, the degree of responsibility for data protection compliance and violations will often not correspond to the degree of actual control citizen scientists have over the processes of compliance and violations. The *Fashion ID* Judgment addressed some of this problem, but not in part where citizen scientists are fully involved in formulation of the research objectives, or have their own objectives for research outcomes. Third, if the citizen scientists are adequately informed about their potential roles and responsibilities as controllers during the recruitment, this may discourage many of them, especially coming from less privileged groups, from taking part in participatory science and make participation in this form of science only accessible to the elites. It is well beyond the scope of this paper to map out a detailed plan of action to address these dilemmas. We will only sketch some general recommendations here, addressed to the EU legislator and courts, as well as the local research ethics committees, leaving more thorough explorations to further research.

The problem of citizen scientist as a controller regarding personal data of him/herself and others is a consequence of a more general problem that the data protection law is facing, namely, how to balance on the one hand complete and effective protection of the data protection rights and not allow responsible actors to avoid responsibility for data protection violations, while on the other hand still assigning responsibility where it belongs. The pre-*Fashion ID* approach has been criticized as over-inclusive, making "everyone a controller". The more restrictive approach the Court of Justice took in *Fashion-ID* has been criticised for drawing the boundary according to the "stages of processing", thereby oversimplifying the modern data processing which is more often then not a complex multi-stage and multi-actor phenomenon. Reducing responsibility for any impacts of processing on the data subject to individual stages of processing neglects this complexity and thus reduces rather than promotes protection of the data subject. This reductive approach does not significantly help the case of citizen scientists either. The Court should consider assigning responsibility based on a different criterion, e.g. the purpose of processing. To illustrate the impact in the context of participatory science, a citizen scientist (and any other joint controller) will be held responsible for data processing only to the extent it was done for the purposes the citizen scientist formulated. If professional researchers – in addition to the research purposes – also are archiving personal data for an extended period of time in order to ensure verifiability of the research results, the citizen scientists should not be considered joint controllers for this processing, since – although they may be aware of and accept this purpose – they did not formulate it. At the same time, if the citizen scientists process personal data for the purposes other than research, e.g. to support legal claims in court or other forms of civil activism, the professional researchers should not be joint controllers and responsible for data processing done for these purposes.

Yet, the change in (case) law may take a long time. In the meantime, action can be taken on the level of self-regulation and ethics guidelines for professional researchers engaging citizen scientists. Such guidelines should at least address the following points.

The first and foremost point of attention for such guidelines is an obligation of the professional researchers to fully inform the potential citizen scientists of their possible role as joint data controllers at the stage of recruitment, as well as the corresponding obligations and possible liability.

Second, professional researchers should run data protection impact assessment of their envisaged research projects to fully map and become aware of the data processing where citizen scientists will be (joint) controllers, along with the data protection obligations associated with such processing.

Third, under law the citizen scientists and professional researchers when acting as joint controllers have to agree on and inform the data subjects of the division of their respective compliance responsibilities, in particular as regards the data subject information rights, and their roles and relationships of the joint controllers vis-à-vis data subjects. The ethics guidelines should include a model agreement between professional researchers and citizen scientists where the responsibilities are divided between the two parties based on their actual roles and capacities to comply with data protection obligations.[94] Because of the gravity of both the responsibility and the consequences for non-fulfilment, these points should be addressed formally. For instance, such a model agreement can contain a clause on the obligation of the researchers to

[93] Prainsack (n 6).

[94] Resnik and colleagues raise a similar point in Resnik, et al. (n 13). That there should be an understanding between the researchers and participants at the outset about how certain rights and responsibilities are distributed.

indemnify citizen scientists against any data protection liability that may arise as a result of the project, in the context of the jointly formulated research purposes. This can partially address the hesitance of less well-off citizen scientists to join participatory science projects. Another model clause could contain an obligation for the professional researchers and their institutions to provide secure infrastructures through which personal data will be collected and stored.

Fourth, where citizen scientists process data beyond joint research purposes, they, alone, will carry responsibilities for such processing. However, researchers must inform citizen scientists of this.

Fifth, where professional scientists are aware of the citizen scientists' own purposes of data processing, they may wish to take measures to ensure that no personal data in an identifiable form is shared with the citizen scientists for those purposes, both to facilitate their own compliance and to shield the citizen scientists from possible liability.

Sixth, in cases where citizen scientists are joint controllers of their own personal data together with professional researchers, the model agreement should vest professional researchers with the data protection obligations.

Finally, to mitigate any data protection violation risks both to the citizen scientists and data subjects, professional researchers should offer citizen scientists training on the basics of data protection law and data security.

There is growing recognition that citizen science presents as a potentially valuable activity, some of which, takes place in the lacunae of traditional research ethics. The question of whether citizen scientists are data controllers under the GDPR is not an insignificant matter. Such a designation assigns substantial responsibility for the protection of fundamental rights under the law to persons who may not fully appreciate or be equipped to execute these obligations. Given the recognized benefits of citizen science, in principle, attention to nuance in the allocation of rights and responsibilities may serve to promote the positive yields of this form of research, also about the awareness about data protection. That citizen science may further the democratization of knowledge generation and enhance the broader scientific enterprise through local and lay lenses are reasons to promote sensible, ethical, and responsible compliance with the law. Taking a nuanced approach to the circumstances of assignment of the status of data controller in citizen science projects is an important step toward responsible and ethical participatory research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Aknolwledgement

The authors wish to acknowledge grant funding ID 873169 from H2020 project, "Citizen Scientists Investigating Cookies and GDPR Compliance"