Citizen-gathered data to support public services under emergencies: promises and perils of openness

Anna Berti Suman[a]
[a] *The European Commission's Joint Research Centre, Ispra, Italy*

Abstract. Society is experiencing sudden and sometimes unexpected crises multiplying at a fast pace. The emersion of these crises requires a swift intervention under the form of providing public services. However, institutions often struggle to cope with crises. In such confronting situations, local people, networks of non-governmental organizations, and associations mobilize. They do so setting up initiatives and producing (open access) data that could fall within the category of 'citizen-generated data'. This contribution draws on theoretical notions such as 'technology appropriation' and 'collective intelligence' in the context of citizen science initiatives to set the scene. It then dives into the analysis of two concrete examples of phenomenon discussed. Subsequently, it explores relevant stakeholders' position on the matter, relying on insights gathered during a focused workshop. In the discussion, the article explores how the current EU legal framework and international standards for data sharing could ensure that these data flows flourish and are used by the public sector, while preserving potentially conflicting interests.

Keywords: Crises, Citizen Science, Open Data, Collective Intelligence

1. Introduction

Initiatives resulting from agile and decentralized efforts from the grassroots often mobilize by producing data in response to crises. The data are then frequently made available in open access on freely accessible platforms on the web. Strategies enacted 'outside' the institutional system in the form of data purposely collected by crises-affected people could be scientifically enriching, as the experience of citizen science practices demonstrates (Berti Suman, 2019; Berti Suman, 2021). Some initiatives may be aimed at gathering information on the spread of a virus and analysing it, such as the Folding@Home example discussed, others at collecting geo-located data on settlements affected by a flood or an earthquake and at predicting future evolutions of the matter, such as the example of the MapSwipe discussed in Section 3.

The data at issue could be considered a basis for 'collective intelligence', i.e., the enhanced capacity created when people work together and join efforts (Guclu, 2004). This knowledge base may be crucial when competent authorities have to intervene on matters they are only partially prepared for. Institutions in charge of providing services to people under stress turn to these practices as opportunities for meaningful public interventions. In addition, researchers interested in studying how crises unfold and how people react to them find precious insights in such data. By contributing data and time, people demonstrate how a certain issue affects and matters to them. They embed values in the data they share (Berti Suman, Heyen, Micheli, 2023) and when they make data openly available, they inevitably disclose also important information on why they gathered such data, how and for whom. All this informational asset could offer opportunities for research and for interventions but could also expose the people engaged in gathering and reporting data, and their communities, to unintended and adverse consequences. This may occur especially when the data include special categories of personal data, such as personal data revealing racial or ethnic origin, pursuant to Article 9 of the General Data Protection Regulation (GDPR).

In light of the perils of openly sharing citizen-gathered data to address emergencies, for example with researchers and institutions, regulating such practices seems advisable. In other words, such decentralized and informal data flows which at the moment do not follow specific regimes for data sharing and storing may need to fit existing legal provisions, or new legal instruments may need to be formulated to regulate them. Already in 2020, Micheli et al. (2020), in an article on emerging data governance models, and Berti Suman and Pierce (2018), more specifically on health data sharing in the context of citizen science initiatives, pointed to a possible legislative and regulatory vacuum. Today in the European Union (EU), which is mainly the geographical scope of this study, we are confronted with developments in policy and legislation addressing data sharing for beneficial purposes such as research and crisis responses. This includes the array of EU policies on Open Science[1], of which projects involving citizen-gathered data can be considered a manifestation (Schade at al. 2020). Furthermore, the European strategy for data, of which the European Data Spaces, the Data Act, and the Data Governance Act (DGA) are key pillars, could offer an umbrella framework for situating, understanding and regulating informal data flows from the grassroots and channel them in a way that can be useful to better address crises. The discussion part of this article will address this legal and regulatory angle.

The article essentially argues that three research questions should be answered to fully unleash the potential of civic data gathered and shared openly in response to crises, without putting at risk the (personal) data that are at stake:

---

[1] See the EU Strategy 2020-2024 for Open Science, https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en.

- Can decentralised data flows coming from spontaneous civic initiatives help innovate the public sector in particular in relation to offering services after crises?
- How to apply citizen science methods and approaches to the field in discussion to ensure that the data are shared with institutions and researchers in a legally compliant and ethical way?
- What are the pitfalls of jointly sharing data as beneficial for contributing to collective intelligence, especially when there are hidden (e.g., market) interests that aim to profit from the data?

The article does not aim at offering exhaustive answers to these complex questions but offers a starting point for researchers and other interested actors wishing to navigate this matter. The article begins illustrating four theoretical lenses that are useful for framing the problem. Subsequently, it provides an insight into cases where the theoretical lenses manifest in practice. Lastly, it discusses data collected during an interactive workshop with pitches and live drawing (i.e., scribbling) to elicit reactions from participants on the matter. The workshop is useful as it offers a sense of what relevant stakeholders in the field, such as citizen science practitioners and researchers, think about the matter. The visuals resulting from the workshop are included in the text as figures to illustrate the key arguments emerged from the session. Relevant notions resulting from the literature review and from the discussion with workshop participants are juxtaposed to extract key themes that are analysed throughout the article. In the final part of the article, conclusions, limitations and future research avenues are outlined.

2. Theoretical lenses

2.1. Citizen science faced with ‘wicked problems’

The Covid-19 pandemic, the Ukrainian war and the climate crisis are examples of ‘wicked problems’. These are problems difficult to solve as facts are uncertain, values in dispute, stakes high and decisions urgent (Rittel and Webber, 1973). Their consequences are disruptive and threaten our health and socio-economic systems. We also live through structural transitions (for example, our increasing reliance on digital media) that change societal communication and interactions, and consequently the way we react to stressors and shocks. To address wicked problems, we cannot rely only on technocratic and top-down solutions.

In critical situations civic actors take action, also by producing data and enacting strategies (e.g., spreading geo-located alerts on an approaching flood; sharing personal data of missing people) that could be foundational for a form of 'collective intelligence' (Glucu, 2004), as discussed later.

Citizen science, broadly defined as "the active participation of lay people in scientific research" (Den Broeder et al., 2017), often mobilized in response to crises (one can think to the famous radiation citizen science initiative, Safecast[2], launched in response to the 2011 Fukushima disaster). Citizen science can be regarded as an umbrella concept encompassing manifold instances when ordinary people gather and analyse data on a matter of concern. Citizen science examples span from institutionalized practices designed by scientists or policy-makers to engage the civil society; practices aimed at educating the citizens on scientific matters; amateurial mapping of biodiversity; informal and conflictive forms of civic monitoring of pollution from industry etc. The forms of civic data gathering, analysis and sharing here discussed could fit within a broad understanding of citizen science.

Citizen science dates long back in history but has recently received a boost by recent developments in science policies (e.g., the EU endorsement for Open Science), and technology (e.g., progresses in mapping, sensors availability, Artificial Intelligence and data platforms). Kullenberg and Kasperowski (2016, 1) argue that citizen science evolved along three main 'focal points'. The first focal point is identified with the engagement of lay people in the domains of biology, conservation and ecology (such as the Cornell Lab experience, at Cornell University, in the United States - U.S.), where the citizen scientists contributed to the collection and classification of data for supporting official research. The second strand of citizen science is related to geographic information research, where citizens are engaged in the collection of geographic data, which is already closer to the matter analysed here. Lastly, a third strand is related to social sciences and epidemiology, where citizen science becomes a tool to foster public participation in monitoring health and the environment. Among third-strand citizen science initiatives, there are projects combining environmental and health data in response to crisis, which are the type of initiatives that this article encompasses. However, the geographic information aspect in strand two of Kullenberg and Kasperowski's conceptualization is equally relevant for this article as the ability of ordinary people to collect data that are also geo-located is key for designing responses to crises.

---

[2] See https://safecast.org/.

Lastly, Gabrys, Pritchar and Barratt (2016) stress the political dimension of the act of gathering data when institutional actors struggle to grapple with unfolding crises. They describe “the challenging claims that citizens can make with data” (Gabrys, Pritchar and Barratt, 2016, 1). The authors draw a distinction between a long-standing typology of citizen science that mostly entail amateur monitoring and instances of people “deliberately invoking the political possibilities of this data” (Gabrys, Pritchar and Barratt, 2016, 1). The data flows here at issue are often expression of this political dimension of data offered by the authors.

2.2. Appropriating technologies for resisting the market

Citizen initiatives that gather and share data can be (also or in some cases) an act of resistance. Making their own monitoring technologies (so called “do it yourself” - DIY), ordinary people try to contest mainstream technologies available on the market. By imaging and enacting different ways to share data, they innovate dominant models for data sharing (for example, between citizens and the public sector). The discourse that can be found at times in these initiatives is a reaction to forms of digital and data colonialism (Kwet, 2019). This colonialism finds its roots in the design by powerful actors (e.g., companies from the Silicon Valley) of the tech ecosystem, offered to users across the globe and in particular in the (relative) Global South, as a data extractive machine for the purpose of profit. The auto-production of technologies by civic actors and the construction of open access platforms for data sharing emerge as counterforces to this power imbalance.

An interesting phenomenon is the ‘Non-Aligned Technologies Movement’ (NATM)[3] which advances a collective claim, both from academic circles and from civil society, to provide alternatives to mainstream technologies (Mejias 2019). NATM is a global movement that aim to tackle the challenges of digital and data colonialism, and tries to draw imaginaries of people-centric, rights-focused technologies and data governance models. NATM goes beyond pure technology discourses, arguing that the problem also regards techno-social models that are engendering new forms of ‘extractivism’ (including data extraction) and exacerbating historic inequalities. The movement is global and multilateral as such inequalities are witnessed around the world and across sectors. It mainly adopts a decentralised anarchistic orientation. By claiming back technology, the movement wants to take back policy, regulation and governance.

---

[3] See the NATM page, https://nonalignedtech.net/index.php?title=Main_Page.

In some of the initiatives manifesting the phenomenon here at issue, the participants make use of ready-made technologies and platforms, which could reinforce this extractivist problem, whereas in other instances they build or assemble the technologies that they will use to gather data and construct their own platforms for sharing the resulting data. Movements of this type could become important stakeholders if we want to understand and regulate forms of civic data flows in response to crises.

2.3. Joining forces: sharing data to build forms of collective intelligence

Glucu (2004) discusses the notion of collective intelligence[4] in light of the behaviour of social insects that can survive in a changing world responding to evolving challenges by quickly gathering information and jointly finding solutions. The underlying idea is that the whole is more than the sum of its parts, which Johnson (2001) qualifies as the 'principle of emergence'. Entities that result from the collective forces of individual people (the parts) create a 'whole' that has properties that their single individuals do not have which can be valuable to tackle crises.

Similarly, the initiatives studied here are 1) decentralized models of data sharing; 2) reactive initiatives (they respond to crises in agile ways that often do not fit existing and regulated data sharing schemes); 3) solution-oriented efforts (people gathering data do so having in mind a certain problem and solution that they wish to find); and 4) socially-oriented (they generally have some forms of social purpose behind the sole data gathering and sharing). Lévy back in 1999 argued for the revolutionary potential of a greater societal reliance on collective intelligence. More recently, Verhulst (2018) suggested that coupling the potential of collective intelligence with progresses of Artificial Intelligence (AI), we can innovate how we govern and respond to crises. AI-based initiatives such as AlphaFold[5] (a project aimed at predicting a protein's 3D structure from its amino acid sequence), can offer precious pool of information to researchers but also to citizens for addressing unfolding public health needs. Some civic initiatives aimed to tackle global crises relied on AI and machine learning.[6] However,

---

[4] For drafting this section – in particular for developing the notion of collective intelligence and for the review of AI-based citizen science projects – the author would like to acknowledge the contribution of Marisa Ponti, University of Gothenburg.

[5] See https://alphafold.ebi.ac.uk/.

[6] See the case of Folding@Home, https://foldingathome.org/?lng=en. For all the infectious disease outbreaks addressed by the initiative, see https://foldingathome.org/diseases/infectious-diseases/?lng=en.

the majority of citizen (science) projects relying on AI and machine learning pertain to non-crises scenarios (McClure et al. 2020; Anton et al. 2021; Franzen et al. 2021[7]). A great deal of projects belong to fields such as astronomy and astrophysics (e.g., Galaxy Zoo, Gravity Spy, and Supernova Hunters), biology (e.g., EteRNA, EyeWire, and Project Discovery), and ecology and biodiversity (e.g., eBird, Bat Detective, Koster Seafloor Observatory).

Albrecht (2016, 12) in his provocative piece "Exiting the Anthropocene and entering the Symbiocene" argues that in a reality of crises and planetary collapse – which he frames as "the new abnormal" – it is important to synchronize human efforts to shared objectives, such as collective survival or, more simply, wellbeing. Albrecht suggests that society should embrace new foundations: in his view, the next era in human history should be named the symbiocene (from the Greek sumbiosis or companionship), this is, living together for mutual benefit. In the natural world, symbiosis is visible in the way different organisms living in close physical association interact. For example, macrofungi and flowering plants establish mutually beneficial and symbiotic associations. The metaphor of the symbiotic forest ecosystem overturns the assumption that evolution has mostly been based on competitive struggle between species. Applying these considerations to politics and human societies, Albrecht hopes for human intelligence to replicate the described symbiotic and mutually supportive relationships, including joining forces to tackle crises in a data-informed manner.

What is particularly relevant for this article is that an explicit mention to citizen science is made in Albrecht's reasoning where he affirms that the model he imagines foresees "symbiocratic governance by scientifically and traditionally informed humans (including those practicing citizen science)". Under this perspective, faced with crises, people respond together sharing the data that are needed, finding strength and mutual aid in the community and not fighting one against the other for taking advantage of a crisis (a manifestation of 'symbiocratic data governance').

Among the mutual relationships and interactions that are particularly interesting for the Open Science angle, there are those synergies that people affected by a crisis build with the scientific world. In several occasions (lately, with the Covid-19 pandemic), in the aftermath of crises, scientists offered their knowledge and resources (such as platforms for data sharing among researchers) to ordinary people engaged in crisis response. Such scientists and science institutions become 'stewards' of the civic-gathered data and can mediate between these actors

---

[7] See in particular Table 10.1 Examples of ML in citizen science projects.

and competent authorities to ensure that the data are used properly to inform responses (which may or may not happen in practice unfortunately). By discussing with citizens methods of data collection, analysis and results, they help ordinary people to ensure credibility and rigor of their monitoring. This can signal a broader phenomenon of hybridization between the world of research and that of informal civic data sharing practices. The legal implications of this hybridization should be considered as data sharing from 'traditional' science actors is a heavily regulated practice, differently from the regulatory gap that characterizes forms of spontaneous civic responses to crises through data sharing.

2.4. Sharing data for public benefit

The civic initiatives studied here share data for the 'public benefit'. The literature is divided on the notion of public benefit and different terms such as public benefit, public interest, public good and social value are often used interchangeably (Ballantyne and Schaefer, 2020, in relation to health data research). This creates social and arguably legal uncertainties on the conditions under which such public benefit takes place. Here I borrow from the field of health data, embracing Ballantyne and Schaefer (2020)'s definition of public benefit as the "additional benefit produced by research that would enhance the current knowledge (. . .) of a community." The authors note that public benefit is assessed weighing envisaged harms against expected benefit. The different but adjacent notion of public interest instead would "requires consideration of the trade-offs between competing common goods" according to the authors.

Cheung (2020) highlights how the lack of a legal definition of public benefit and public interest could lead to exploiting the notion to push for more data sharing with market actors, for example. Mészáros and Ho (2018) stress that the GDPR foresees different levels of public interest (general, substantial. . .) and clarification of these different levels would be needed as they can differently affect data subjects' rights.

Relevant for the cases here at stake is the notion of 'data altruism' and its link to conceptions of public interest (Comandè and Scheider, 2022). This concept is reflected in the DGA where the open sharing of data for the purposes of the public interest, such as for research purposes or (very relevant here) to improve public services, is foreseen. The term 'altruism' offers a crucial limit to the sharing: the principle of data altruism mandates that the data should be used for non-profit purposes, which should protect civic initiatives from the risk of market capture.

Public benefit is often understood way beyond an individual level, encompassing indirect benefits such as enhanced knowledge on a certain matter for a (more or less broadly understood) community. It is certainly not an easy outcome to demonstrate as it requires tangible evidence such as, e.g., after an earthquake, faster and more efficient provision of aid to individuals in need. Hutchings et al. (2020) argue that the public benefit discourse can change perceptions of the value of privacy, or – in other words – societal benefit may outweigh concerns regarding privacy. This may explain the drive of crises-affected or concerned people to mobilize sharing (at times also sensitive) data. In the case of the initiatives discussed in this article, there are instances of people that decide to 'trade' privacy for the public good. This means that the preoccupation towards a certain matter that had to be urgently addressed (e.g., the Covid-19 pandemic) outweighed the concerns for the possible privacy risks for the personal data shared. A similar outcome is discussed also in a study focused on health citizen science (Berti Suman and Pierce 2018) where we reviewed projects were patients/users were invited to voluntarily share their health data to contribute to medical research on pressing matters of concern, relying on a strong 'common good' argument.

The illustrated concern about participants putting aside privacy concern moved by a willingness to contribute to the common good is also found in Balestrini et al. (2021). In the study, the authors discuss opportunities and barriers posed by the up-scaling and spreading of citizen-generated data projects. In particular, the authors underline "the tensions that arose regarding data ownership when people collect data that reveal personal behaviours (. . .) and share it openly for the common good" (Balestrini et al. 2021, 12). The study recommends that is key to tackle matters related to data ownership from the start of the citizen initiative, as this "can have a significant impact on citizen engagement, trust, and commitment" (Balestrini et al. 2021, 12). This is not easy as it requires a great deal of flexibility by the project initiators, due to the fact that "perceptions about privacy may change as participants become more knowledgeable about technology and begin to make sense of data" (Balestrini et al. 2021, 12). Therefore, it is fundamental to shape citizen initiatives in a way that they become able to adapt their protocols in response to evolving participants' concerns about data ownership, as Balestrini et al. (2021) suggest.

Beyond privacy concerns, it is worth to highlight that a study (focused on biodiversity monitoring citizen science) revealed criticalities on the extent to which the surveyed projects concretely adhered to the Open Science principles, in particular in terms of "accessible data, code, software, publication, data management plans, and preregistrations"

(Suter et al. 2023; 1). The study notes that, despite claiming to contribute to openness in science, in practice several projects scored low in their realization of this objective. This clearly has "implications for how the public can interact with the research that they play an active role in contributing to" (Suter et al. 2023; 1). The authors thus encourage the definition of systematic recommendations to better implement Open Science principles across citizen science initiatives.

3. Cases insights

In this section, I discuss two real-world examples of grassroots-led civic responses that specifically responded to needs for public services producing data from 'below'. One of these initiatives entails the handling of personal data and foresee a further sharing of the data with public institutions. The initiatives here used as examples have different levels of civic engagement: they are managed by the organized or unorganized civil society, often coupled with a role of universities and research centres. They respond to needs for public services by addressing the need of data on risks with which decision-makers struggled to cope, in particular the Covid-19 pandemic and sudden natural disasters.

Among the cases worth of discussion, the Covid-19 pandemic gave new momentum to the already existing Folding@Home initiative, "a distributed computing project for simulating protein dynamics, including the process of protein folding and the movements of proteins implicated in a variety of diseases".[8] The initiative "brings together citizen scientists who volunteer to run simulations of protein dynamics on their personal computers and scientific researchers. Insights from these data are helping scientists to better understand biology and providing new opportunities for developing therapeutics".[9] While providing an almost technical support to the initiative, participants enacted forms of citizen science and joined efforts producing data that contributed to collective intelligence on infectious disease such as the Ebola Fever, the Chagas Disease and the Covid-19 pandemic.[10] Participants conceivably expected benefits for their community and society at large, beyond their individual interests. In this case, no personal data that could be considered potentially sensitive is at issue as the project does not require volunteers to share, for example, their own medical information.

---

[8] Folding@Home web page, `https://foldingathome.org/about-2/?lng=en`.

[9] Quotes taken from Folding@Home web page, `https://foldingathome.org/about-2/?lng=en`.

[10] For all the infectious diseases addressed by the initiative, see `https://foldingathome.org/diseases/infectious-diseases/?lng=en`.

Clearly, the benefits of participating outweigh any possible privacy concern.

Another crises-driven civic initiative for providing (also) public services is MapSwipe, an open-source project that aims at helping first responders working with communities affected by disasters, disease and conflicts with data that are needed to provide relief.[11] In particular, as first responded have to "cover large areas, but lack the data necessary for an efficient, effective response", the initiative provides an app to volunteers to "pinpoint where critical infrastructure and populations are located, allowing mappers to focus only on areas where they know features need to be mapped".[12] The data are collected for the public benefit, i.e., not for profit, with the aim of improving disaster relief for the affected people. In addition, in this initiative the discourse on technology appropriation discussed at a theoretical level is echoed. Here, potentially sensitive data could be shared for example medical conditions of certain people under stress. However, receiving assistance after a disaster seems to be a consideration winning over possible privacy concerns.

4. Stakeholders' views on the matter

During the European Citizen Science Association Conference - ECSA (October 5-8, 2022, Berlin), we[13] ran an interactive workshop titled "Rethinking public services provision: citizen science to support public and environmental health services",[14] with pitches from us as co-conveners, and live drawings from a contracted illustrator to elicit participants' reactions. Around 20 people voluntarily decided to attend our session among the others running in parallel at the conference. The participants could read the session outline from the conference program but had not received any preparatory material to the session.

We invited participants to reflect on theoretical foundations and highlight actual examples in which a certain civic initiative complemented or even substituted an official public service through data collection and sharing in crisis scenarios. We offered as examples the cases illustrated in the preceding Section 3. After introducing the topic, we asked participants to brainstorm in three groups, each assigned

---

[11] MapSwipe web page, https://mapswipe.org/en/index.html.

[12] Quotes taken from MapSwipe web page, https://mapswipe.org/en/index.html.

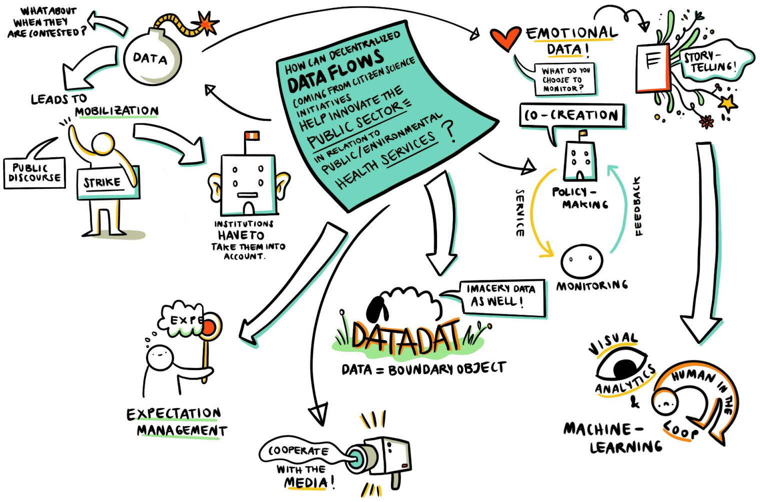
[13] With co-conveners Marisa Ponti, University of Gothenburg; Nils Heyen, Fraunhofer Institute for Systems and Innovation Research ISI; Sven Schade, European Commission's Joint Research Centre, and Alice Toietta, live illustrator.

[14] See https://2022.ecsa-conference.eu/startseite.html.

with a question. The participants were invited to identify keywords and sketch drawings on white sheets. An illustrator followed the discussions making visualisations of what each group discussed and an overarching visual summary. In the following lines, I report highlights from the session that are informative for this article.

The first question was:

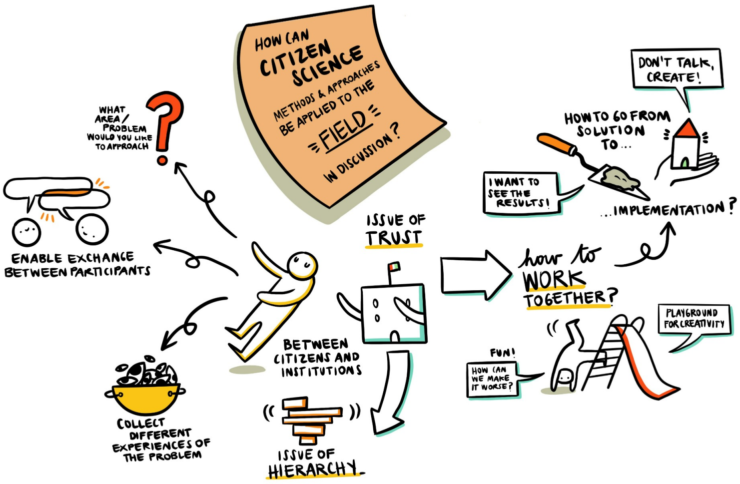
Figure 1. Live drawings during discussion on Question 1



- Question 1: Can decentralised data flows coming from spontaneous civic initiatives help innovate the public sector in particular in relation to offering services after crises? Highlights from the discussion include that the push to openly share data that are needed in crisis scenarios should be always paired with attention to the scientific quality of the data shared. Furthermore, participants viewed data as ‘boundary objects’ that may stimulate or rather hamper participation in crisis response. The discussion interestingly revolved around recipient institutional actors’ ability to embrace ‘explosive’ data (data that may generate conflicts) and ‘emotional’ data (data that express feelings around matters of concern). Lastly, it was suggested that connecting data to (individual and community’s) values could be particularly crucial for addressing crises. However, this can also expose the engaged citizens to the risk of sharing intimate information on what they care for, making them more vulnerable to e.g., attacks on social media from extremists or privacy infringements.

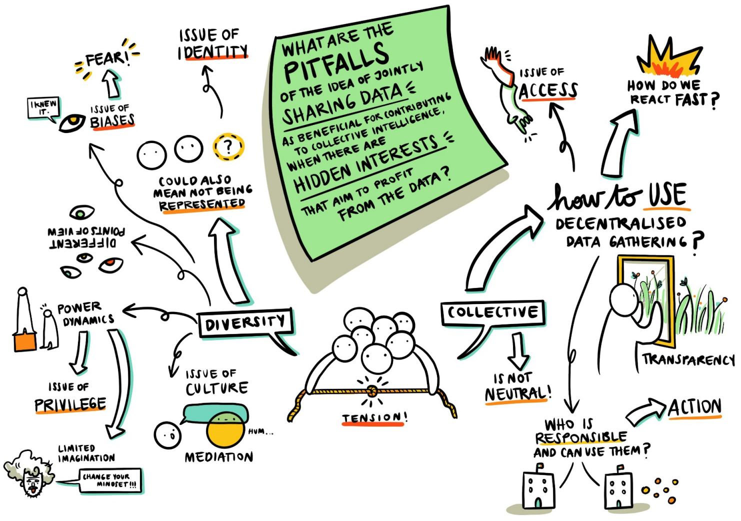
- Question 2: How to apply citizen science methods and approaches to the field in discussion to ensure that the data are shared with institutions and researchers in a legally compliant and ethical way? The participants engaged in the discussion noted that citizen science methods and approaches could teach us how to acknowledge the work of mobilised and active people as they put efforts in monitoring a certain matter by gathering first-hand data, and their efforts must be recognized. Giving due credits to the work that ordinary people put in the data they then share with official institutions and researchers for crisis response can increase trust in the institutional system and make the services provided more inclusive of what matters to people in a crisis. All this can promote a shared understanding of problems, while de-constructing (hierarchical) structures and innovating public governance. Borrowing from the experience of citizen science, we can imagine ways to rely on data gathered by civic actors in crises aftermaths in order to embrace the range of experiences surrounding a problem, and mobilise the best possible knowledge and expertise to find solutions.

Figure 2. Live drawings during discussion on Question 2



- Question 3: What are the pitfalls of jointly sharing data as beneficial for contributing to collective intelligence, especially when there are hidden (e.g., market) interests that aim to profit from the data? The discussion to address this question targeted the issue of potentially conflicting interests which could hinder a fruitful synergy between civic data and institutional actions. This may prevent uptake as people or institutions could be worried of hidden agendas (e.g., the citizens may fear surveillance from authorities and authorities may be worried of data fabrication). In addition, people may be concerned that the data shared with institutions could end in the hands of market actors that wish to access the data for profit purposes. In addition, although the concept of a dispersed collective taking action by sharing data for public benefit is fascinating, it is unclear who is represented by this collective (i.e., who are the people behind all this? How representative of wider society are these collectives?). The notion of 'collective' is certainly not neutral, and there are power relations and hierarchies as well in civic groups, implying risks of bias and exclusion. To address these challenges, it is crucial to identify roles and identities, disclose agendas and interests of the various participants around the table. Clarifying upfront collective stakes and goals can be helpful to overcome these issues.

Figure 3. Live drawings during discussion on Question 3



In the final part of the workshop, we summarized the key takeaways from our shared discussion. Relevant findings include the argument that we cannot assume that something like 'collective' and 'collective intelligence' exists as a homogenous and harmonious collaboration between civic actors. Due to (increasing) polarisation, 'infodemics', and social frictions, we must look at collective efforts in crisis responses in their nuances, stratifications and internal differences. We concluded that we need a stratified understanding of what a collective entails, to embrace the diversity and complexity of any collective. Consequently, we should imagine data governance and regulatory approaches that can grasp these differences. Another key takeaway is the argument that we need to broaden our scope of analysis from relying on citizen-gathered data in crisis scenarios to involve collective intelligence also in ordinary problem-solving, in order to offer inclusive, value-based and responsive public services to citizens.

A few methodological takeaways from the workshop discussion accompanied by live drawing are also relevant. First, the dynamism of the session stimulated participants from different backgrounds (e.g., civic, academic, policy) and with different attitudes to contribute to the discussion. The hands-on drawing and keywords identification proved to be useful to keep participants' attention high, stimulating creativity as well as assisting them with imagining complex and sometimes abstract concepts. A professional live illustrator accompanied the drawings from the participants. This helped us digest the inputs from the session in a comprehensible and catchy way. The image below shows the final digest.

Figure 4. Live drawings during discussion on Question 4

5. Discussion

5.1. APPROACHING THE EU LEGAL FRAMEWORK FOR DATA SHARING

n the preceding sections, I targeted the key questions that guided this study, namely how the reliance on decentralised citizen-gathered data flows coming from spontaneous civic initiatives can help innovate the public sector in crisis scenarios, borrowing from citizen science experience to inquire promises and perils of these practices. I adopted a lens built on the notion of citizen science as an agile response to crises, collective intelligence and symbiocratic (data) governance, technology appropriation and, lastly, on the still fuzzy notion of public benefit.

In this discussion section, I look at the matter from a legal perspective, to understand whether the current EU legal framework for data sharing can offer models to regulate the studied practices. Starting from the environmental field, environmental monitoring and reporting is a heavily regulated practice within the EU, as noted in the Staff Working Document (2017, n. 230, pp. 8-9).[15] The regulatory framework dates back to 1991 when the European Economic Community adopted the Standardised Reporting Directive (SRD-91/692/EEC).[16] The use of electronic means for transmission of environmental data and the possibility to report and publish data online, for example through open access maps, generated a move towards the definition and harmonization of electronic data standards. This need of data standards led to the adoption of the INSPIRE Directive[17] in 2007, creating an EU-wide spatial data infrastructure and setting technical standards for the interoperability of spatial data and for the online availability of data discovery and access services, therefore promoting comparability and data sharing (SWD(2017) n. 230, p. 9).

Efforts at the EU level to streamline environmental reporting (COM (2015)215)[18] notably included an institutional commitment "make better use of data [...] directly from the public (e.g. in the context of citizen science)" (EC COM(2017)312, p.4).[19] Especially this latter find-

---

[15] EC SWD(2017)230 'Fitness Check of Reporting and Monitoring of EU Environment Policy'.

[16] Council Directive 91/692/EEC of 23 December 1991 standardizing and rationalizing reports on the implementation of certain Directives relating to the environment (OJ L 377, 31/12/1991).

[17] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).

[18] EC COM(2015)215 on "Better regulation for better results - An EU agenda".

[19] EC COM(2017)312 on "Actions to Streamline Environmental Reporting".

ing deserves attention here. In the same COM(2017)312, at p. 11, a specific action in support of citizen science is outlined, i.e. “Action 8: Promote the wider use of citizen science to complement environmental reporting”. More recently, the European Commission published a SWD on best practices on using knowledge generated by citizen science initiatives across the EU (SWD(2020) 149).[20] The document argues that citizen science offers “a unique opportunity to help deliver on the European Green Deal (...) and to involve the public in EU policy-making.” The SWD summarises the opportunities for and benefits of using citizen science for environmental monitoring, highlights good practices and lessons learnt, and identifies the obstacles holding back its broader uptake. Currently, these discussions are feeding into the EU data strategy,[21] in particular with regards to the Environmental Data Spaces, defined as forms of “data exchange where trusted partners share data for processing without sacrificing data sovereignty”.[22] Such spaces are considered pivotal for the implementation of the European Green Deal,[23] as they ensure that frontier research can rely on the needed data availability, at the same time ensuring data protection and trusted access.

5.2. THE APPLICATION OF THE FRAMEWORK TO THE CIVIC INITIATIVES DISCUSSED

The initiatives and data flows discussed here, however, fit only partially in the legal framework just outlined in the previous sub-paragraph. Indeed, all the illustrated efforts mostly focus on environmental data sharing whereas most of the initiatives discussed here also include health data (e.g., on the health conditions of people affected by a disaster) and other forms of personal data, often belonging to special categories. Therefore, the framework just described cannot be applied, at least not ‘blindly’, to the matter at issue. A specific regulatory instrument that would respond specifically to environmental and health data ‘mashups’ (that is, combining data from different data sources into a single application) shared in crises aftermaths by civic actors for the public benefit (to be defined clearly) could be an avenue. In

---

[20] Available at https://circabc.europa.eu/ui/group/c1a5a4e9-7563-4d0e-9697-68d9cd24ed34/library/d08a6ffd-2a91-437e-a473-84c47bb74c7c/details?download=true.

[21] See https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.

[22] See https://environmentaldataspace.com/ and https://wetransform.to/green-deal-data-space-gaia-x/.

[23] See https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en.

the U.S., for example, a Crowdsourcing and Citizen Science Act regulates the practice since 2016.[24] A regulation could also ensure data protection and privacy of the data flows, balancing possibly conflicting interests (Berti Suman and Pierce 2018). Furthermore, regulating such data flows could help tackling the risk that the shared data is not trustworthy and becomes vehicle of false information.

As (Paseri, 2020, 59) notes, making the example of totalitarianism and the diffusion of erroneous scientific knowledge, this risk "in the long run, may imply an irreversible damage to democracy, causing forms of manipulation". Recognizing these data flows could also promote institutional reliance on the data ('policy uptake', Berti Suman, 2021). However, there may be a 'regulatory trap', which entails capturing in static regulations a dynamic practice, with the risk of hampering innovation from below (Berti Suman, 2019). The work of Scassa[25] on Canadian and U.S. legal and regulatory matters around citizen science could be of inspiration for an EU legal intervention on the matter.

5.3. A POSSIBLE WAY FORWARD

As often these data flows include health data, legal efforts should inspect how these initiatives may or may not fit within the European Health Data Space (EHDS). COM(2022)196/2 argues that the EHDS will "have a significantly positive impact on fundamental rights as regards personal data protection and free movement.[26] The right discourse seems particularly relevant for the purposes of this study as often the data shared could potentially harm privacy and other rights of the data subjects. COM(2022)196/2 links the EHDS with the European Open Science Cloud (EOSC) data space which will "enable researchers, innovators and policy-makers to more effectively use the data securely and in a way that safeguards privacy." The data flows studied here often are of value for and shared with researchers, thus it seems that they can be situated at the intersection of Environmental and Health Data Spaces and of the EOSC.

Particularly valuable for the discussed experiences is the EU's DGA of 2022, which took effect in September 2023, as the first legislation to enable third party data governance by regulating (for-profit 'data intermediation services' and) non-profit 'data altruism organizations' (see Chapters III & IV). Data intermediaries, which could belong to the civic

---

[24] 15 U.S. Code § 3724 (2016) - The Crowdsourcing and Citizen Science Act.

[25] Articles from the blog of Teresa Scassa available at `http://www.teresascassa.ca./index.php?option=com_k2&view=itemlist&task=tag&tag=citizen%20science`. See in particular Scassa 2018.

[26] EC COM(2022)196/2 on "A European Health Data Space: harnessing the power of health data for people, patients and innovation".

collectives discussed in the article, are third parties designated to act as mediators for data subjects and collectors. They will be permitted to collect, pool and share data about people to “enhance the agency of data subjects, and in particular individuals’ control over data relating to them” by helping them exercise their rights under the GDPR and negotiating terms of use on a collective basis (Recitals 30-31).

By linking the notion of altruism in the DGA with that of public benefit and with the rights of the data subjects under the GDPR, we could get the tools to make a wise and legally sound use of civic-gathered data for crisis response but also beyond such instances. The legal avenue should ensure the balancing the interests of openness with that of data protection both at an individual and at a collective level (Mantelero, 2017; van der Sloot & Graef, 2022). In addition, regulatory efforts of this kind could ensure that these approaches evolve towards forms of (more) “inclusive data governance” (Micheli, 2022).

At this stage, it will be also useful to identify the different roles of the actors processing personal data (Article 4 of the GDPR) within a certain ‘collective’ (see also Berti Suman and Pierce, 2018, on ‘Joint Controllers’). Lastly, as at the core of these initiatives there is the idea that data sharing can generate more value from the data, attention to the economic implications and market engagement in them should be closely scrutinized. Corporations could be interested in supporting these initiatives to scale up and to be more widespread, when they prove to be successful at the small scale. In these cases, it will be important to ensure that market access to the initiative and data does not become control and capture, at the detriment of the public benefit.

Last but not least, as the ECSA workshop discussion suggests, an important aspect to consider in envisaging a way forward is related to the potential ethical challenges raised by sharing data collected by civic initiatives in crisis scenarios. The ethical dimension is a fundamental aspect of defining how to govern and regulate the discussed data flows (Resnik et al., 2015). As such data flows stem from civic initiatives, the engaged people may (or may not) have a clear idea on how they wish the data to be shared in a manner that they consider ‘ethical’. Here, international standards such as the FAIR principles (which stands for ‘Findable, Accessible, Interoperable and Reusable’ data and are aimed at promoting the reuse of data)[27] should be complied with, but also be ‘shaped’ according to the wishes of the communities producing such data. In this context, drawing on Carroll et al. (2021), it could be worth exploring the integration of the FAIR principles with the CARE principles (originally intended for Indigenous data but with promising

---

[27] See https://www.go-fair.org/fair-principles/.

applications also beyond it), which encourage greater participation of and benefit-sharing with those that produce the data. CARE indeed stands for ‘Collective benefit, Authority to control, Responsibility and Ethics’.[28] Specifically drafted for citizen science projects, the Public Participation in Scientific Research - PPSR Core principles[29] could be valuable here as they offer standards for how to describe the data coming from civic initiatives (i.e., metadata) to facilitate data sharing and reuse that is mindful of the specific provenience of the data. These principles are increasingly being recognized as a benchmark for the European Open Science discourse and could help to preserve the interests of those civic actors producing and sharing the data.

6. Conclusion

The article offered a perspective on the opportunity of a greater reliance on decentralised citizen-gathered data flows coming from spontaneous civic initiatives to innovate interventions in crisis scenarios. I argued that citizen science methods and approaches could offer a useful guidance on how to structurally rely on these data flows in ways that are scientifically and legally sound. However, I also stressed the promises and perils of relying on and scaling up these practices. Theoretical notions such as technology appropriation, collective intelligence, and public benefit appeared key to navigate this complex matter.

From a legal review, it appeared clear that these data flows at present do not ‘fit’ existing legal frameworks, as there is none explicitly devoted to them. However, these data could find regulatory guidance in the DGA of 2022 (especially in terms of data intermediaries and the notion of data altruism), in the strategy for the European Environmental and Health Data Spaces, and in the EOSC, considering the environmental and health dimension of these data and the connection with research purposes. Ideally, these policy and legal instruments offer a ‘safe’ space where civic initiatives can develop and flourish, without being unduly restricted by regulatory burdens but aligning with data protection, privacy and data security principles. Lastly, it is important to consider the ethical dimension of data sharing for the public benefit. What is ‘ethical’ to share should be defined in a participatory manner, complementing international standards (such as the FAIR and CARE principles as well as citizen science-specific principles for metadata, i.e., PPSR Core) with inputs from communities directly engage in producing and sharing data.

---

[28] See https://www.gida-global.org/care.

[29] See https://core.citizenscience.org/.

This article addressed complex questions without the ambition and resources to reply exhaustively to them. For such answers, systematic literature review, exploratory case studies and analysis of secondary data from past and ongoing experiences would be needed. The context-dependency and cultural variations of notions such as ‘altruism’, ‘collective intelligence’ and ‘public benefit’ should also be considered and taken into account for designing inclusive regulatory interventions. As a concluding message that this contribution would like to stress is that the burst in crises (some of which unexpected) of the last decade, posing wicked problems to institutions, researchers and society at large, cannot be addressed only through top-down, tech-driven and centralized interventions. Rather, decentralized, socially supported and values-informed responses are needed.

If unaddressed or addressed only in a hierarchical manner, wicked problems risk to jeopardize the fundamentals of our societies and democracies, or at least increase public distrust and augment existing legitimacy gaps of institutional interventions. Hopefully, in the near future, initiatives such as those discussed in this article will become ‘visible’ and recognized by competent institutions so that they can strengthen our preparedness to crises.

Acknowledgements

The author would like to acknowledge the support of the ongoing the Marie Skłodowska-Curie grant n. 891513, under H2020-EU. In addition, a sincere thanks goes to the co-conveners and all participants of the workshop held at the European Citizen Science Association Conference, discussed in the article. In particular, the author would like to acknowledge the contribution from Marisa Ponti, University of Gothenburg; Nils Heyen, Fraunhofer Institute for Systems and Innovation Research ISI; Sven Schade, the European Commission Joint Research Centre; and Alice Toietta, freelance illustrator, as very valuable for shaping the article.

References

Albrecht, Glenn. “Exiting The Anthropocene and Entering The Symbiocene.” Humans and Nature 2016. Web. 8 February 2023. https://humansandnature.org/exiting-the-anthropocene-and-entering-the-symbiocene/

Anton, Victor et al. “An open-source, citizen science and machine learning approach to analyse subsea movies.” Biodiversity Data Journal 9, e60548 (2021): 1-14.

Balestrini, Mara, Kotsev, Alex, Ponti, Marisa, et al. “Collaboration matters: capacity building, up-scaling, spreading, and sustainability in citizen-generated data projects.” Humanities and Social Sciences Communications 8, 169 (2021): 1-15.

Ballantyne, Angela, and Schaefer, Owen. “Public interest in health data research: Laying out the conceptual groundwork.” Journal of Medical Ethics 46, 9 (2020): 610-616. Print.

Berti Suman, Anna, and Pierce, Robin. “Challenges for citizen science and the EU open science agenda under the GDPR.” European Data Protection Law Review 4 (2018): 284-295. Print.

Berti Suman, Anna, Heyen, Nils B., and Micheli, Marina. “Reimagining health services provision for neglected groups: The ‘personalization from below’ phenomenon.” Frontiers in Sociology 8 (2023):1052215, 1-6. Web. 20 February 2023.

Berti Suman, Anna. “Between Freedom and Regulation: Investigating Community Standards for Enhancing Scientific Robustness of Citizen Science.” Regulating New Technologies in Uncertain Times. Information Technology and Law Series, Vol 32. Ed. Leonie Reins. The Hague: T.M.C. Asser Press, 2019. 31-46. Print.

Berti Suman, Anna. The Policy Uptake of Citizen Sensing. Cheltenham: Edward Elgar, 2021. Print.

Carroll Stephanie R. et al. “Operationalizing the CARE and FAIR Principles for Indigenous data futures.” Scientific Data 8, 1 (2021): 1-6.

Cheung, Sarah. “Disambiguating the benefits and risks from public health data in the digital economy.” Big Data & Society 7(1) (2020): 1-15. Web. 5 February 2023.

Comandè, Giovanni, and Schneider, Giulia. “It’s time. Leveraging the GDPR to shift the balance towards research-friendly EU data spaces.” Common Market Law Review 59, 3 (2022): 739-776.

Den Broeder, Lea et al. “Public Health Citizen Science; Perceived Impacts on Citizen Scientists: A Case Study in a Low-Income Neighbourhood in the Netherlands.” Citizen Science: Theory and Practice 2, 1 (2017): 1-17. Web. 20 February 2023.

Franzen, Martina et al. “Machine Learning in Citizen Science: Promises and Implications”. The Science of Citizen Science. Eds. Katrin Vohland et al. Cham: Springer, 2021. 183-198. Print.

Gabrys, Jennifer, and Pritchard, Helen. “Just Good Enough Data and Environmental Sensing: Moving Beyond Regulatory Benchmarks toward Citizen Action.” International Journal of Spatial Data Infrastructures Research 13 (2018): 4-14..

Guclu, Hasan. “Collective intelligence in ant colonies.” Fountain Magazine Oct-Dec. 2004. Web. 24 Feburary 2023. https://fountainmagazine.com/2004/issue-48-october-december-2004/collective-intelligence-in-ant-colonies

Horst Rittel and Melvin Webber. “Dilemmas in a General Theory of Planning.” Policy Sciences 4, 2 (1973): 155-169. Print.

Hutchings, Elizabeth, Loomes, Max, Butow, Phyllis and Boyle, Frances M. “A systematic literature review of health consumer attitudes towards secondary use and sharing of health administrative and clinical trial data: A focus on privacy, trust, and transparency.” Systematic Reviews 9, 1 (2020): 1-41.

Johnson, Steven. “Emergence: The Connected Lives of Ants, Brains, Cities, and Software”. New York: Scribner, 2002.

Kullenberg, Christopher, and Kasperowski, Dick. “What Is Citizen Science? - A Scientometric Meta-Analysis.” PLoS ONE 11, 1 (2016): e0147152, 1-16.

Kwet, Michael. “Digital colonialism is threatening the Global South.” Aljazeera Mar. 2019. Web. 3 Feburary 2023.

https://www.aljazeera.com/opinions/2019/3/13/digital-colonialism-is-threatening-the-global-south

Levy, Pierre. Collective Intelligence. New York: Basic Books, 1999.

Mantelero, Alessandro. “From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era.” Group Privacy. Eds. Taylor, Linnet, Floridi, Luciano, and van der Sloot, Bart. Heidelberg: Springer International Publishing, 2016. 139-158. Print.

McClure, Eva et al. “Artificial Intelligence Meets Citizen Science to Supercharge Ecological Monitoring.” Patterns 1, 7 (2020): 1-9.

Mejias, Ulises Ali. “To fight data colonialism, we need a Non-Aligned Tech Movement.” Aljazeera Sept. 2021. Web. 3 Feburary 2023. https://www.aljazeera.com/opinions/2020/9/8/to-fight-data-colonialism-we-need-a-non-aligned-tech-movement/

Mészáros, János, and Ho, Chih-hsing. “Big Data and Scientific Research: The Secondary Use of Personal Data under the Research Exemption in the GDPR.” Hungarian Journal of Legal Studies 59, 4 (2018): 403-419.

Micheli, Marina, Ponti, Marisa, Craglia, Max, and Berti Suman, Anna. “Emerging models of data governance in the age of datafication.” Big Data & Society 7 (2020): 1-15. Web. 22 February 2023.

Micheli, Marina. “Public bodies’ access to private sector data.” First Monday (2022). Web. 20 January 2023. https://firstmonday.org/ojs/index.php/fm/article/view/11720

Paseri, Ludovica (2020), *Access to Scientific Information and Knowledge: A Matter of Democracy*, IFDaD International Forum on Digital and Democracy 2781, 2020, pp. 1-12.
https://iris.unito.it/handle/2318/1766470

Resnik, David B. et al. “A framework for addressing ethical issues in citizen science.” Environmental Science & Policy 54 (2015): 475-481.

Scassa, Teresa. “Data Ownership.” Centre for International Governance Innovation Paper Series n.187 Sept. 2018. Web. 8 February 2023. https://www.cigionline.org/publications/data-ownership/

Schade, Sven, (. . .) Berti Suman, Anna, et al. “Activity Report on Citizen Science – discoveries from a five year journey”. EUR 30551 EN. Luxembourg: Publications Office of the European Union, 2020.

Suter, Samantha, Barrett, Brian, and Welden, Natalie. ”Do biodiversity monitoring citizen science surveys meet the core principles of open science practices?” Environmental Monitoring and Assessment volume 195, 295 (2023): 1-14.

van der Sloot, Bart, and Graef, Inge. “Collective Data Harms at the Crossroads of Data Protection and Competition Law: Moving Beyond Individual Empowerment.” European Business Law Review, 33, 4 (2022): 513-536.

Verhulst, Stefaan. “Where and When AI and CI Meet: Exploring the Intersection of Artificial and Collective Intelligence Towards the Goal of Innovating How We Govern.” AI & Society 33 (2018): 293-297.