Can civic data be counterdata and open data? Exploring the limits of data, contestation and governance

Yuya Shibuya, Seyi Olojo, Andrea Hamm, Radhika Krishnan, and Teresa Cerratto Pargman

Abstract: The increasing surveillance by big tech companies or/and governments has raised concerns about the democratic and participatory structure of the datafied society. Meanwhile, over the course of the past decade, various bottom-up civic tech and digital civic initiatives have emerged to tackle pressing local issues, such as air pollution and disaster response, often via technology-mediated data collection, curation, analysis, design and visualisations, thus promoting democratic participation. In this article, we discuss how these data are understood in diverse contexts beyond the realm of civic tech and digital civics. In doing so, we explore the potential and limits of civic data by exploring the intersections of and differences between civic data and adjacent data-related concepts often used by civic tech communities themselves: counterdata and open data. Through our discursive exploration of these three data concepts, we conclude that understanding is limited when it comes to determining which data are 'civic', and that discussion of questions related to power structures, diversity and inclusion and infrastructuring of civic data has been minimal.

Keywords: civic data, civic tech, counterdata, open data, participation

1. Introduction

Civic tech initiatives (Harrell, 2020; Schrock, 2019) often use the Internet of Things (IoT), online survey tools and others to engage citizens to collect data with the aim of improving community services, civic engagement/participation and citizens' quality of life. Such initiatives have made various contributions to generating open data (Dunn, 2016; Shibuya et al., 2022), community building (Cerratto Pargman et al., 2018; Le Dantec & DiSalvo, 2013) and outreach (Le Dantec et al., 2011; Wehn & Evers, 2014). Such '**civic data**' – both captured and owned by the citizens for the citizens –

(Hamm et al., 2021; Maskell et al., 2018; Shibuya et al., 2022) empower people with the knowledge and resources to take action on pressing local issues. Compared with data that are passively collected by big tech and governments, civic data involve relatively small amounts of data that are inextricably linked to a place, a group of people or a purpose/issue (Maskell et al., 2018; Taylor et al., 2018). Civic data is connected to what Wells (2015) called 'civic information', which is conceptualised as 'the continuous flow of facts, opinions, and ideas that help citizens understand matters of potentially public concern and identify opportunities for action' (p. 7). Civic data is understood as not just a collection of data but rather a process of formulating what is important, for whom and why, and thus developing strong bases for action (Maskell et al., 2018; Williams, 2020). In civic tech activities, the adjacent data-related concepts, counterdata and open data, are often used to describe their data activities. **Counterdata** can be defined as the production of data that correct institutional data sources that often misrepresent communities. Such counterdata enable former data subjects to regain political power (Burrell et al., 2024; Dunn, 2016; Meng & DiSalvo, 2018). **Open data** refers to data that are complete, primary, timely, accessible, machine-processable, universally available, non-proprietary and license-free (Dawes, 2010, p. 379). It should be freely sharable and reusable (Gao & Janssen, 2022, p. 2). According to existing definitions, availability and access, reusability, redistribution and universal participation in production and distribution are the cornerstones for open data.

This article explores the intersections and differences between civic data, counterdata and open data to provide future research topics on the subject of citizen participation in the datafied society. In doing so, we ask whether these three data concepts are interchangeable (i.e. 'Can A be B?'). This format allows us to explore and stretch the boundaries of each concept and gain a deeper understanding of their limitations and potentials. Through this exploratory study, we identify a need to continuously investigate the question of which data can be considered to be 'civic' and to examine issues regarding power structures, diversity and inclusion and how to infrastructure civic data. Addressing these issues will help strengthen citizen participation in the datafied society.

2. Can counterdata be civic data?

Much of the literature documenting counterdata production initiatives reports on the efforts of activists who use data as a means of bringing attention to social issues that have been mishandled by legislative bodies. This includes the production of data to contest harmful policy decisions (Meng & DiSalvo, 2018) and to create awareness around deeply stigmatising issues, such as gender-based violence (D'Ignazio et al., 2022). The collection and sourcing of data serve as the first – and oftentimes the most important – means of effectively garnering the attention of the government when the necessary recognition was initially lacking. Therefore, the nature of counterdata production does not fit neatly into 'civic' data initiatives. Issues related to whose advocacy is readily received by governing bodies become strikingly apparent when reflecting on the conditions that lead to the production of counterdata. For example, during the COVID-19 pandemic, many communities of colour in the United States were unable to track their local mortality rates from the virus, as official statistics, collected at the city and state level, failed to provide comprehensive data on this. As many members of Black communities in the United States are already seen as second-class citizens, the invisibilisation of their experiences within official statistics reflects the lack of regard that state-funded institutions have for them as compared with other racial groups. This example asks a central question for civic data initiatives: what does participation look like for communities whose relationships with governing bodies are politically tenuous?

3. Can civic data be counterdata?

Various local civic tech initiatives can collect a wealth of data, but such data can only be beneficial if they are actually used (Alvarado Garcia et al., 2017; Kim et al., 2011). It is crucial for the data to be collected, shared and interpreted in a way that enables people to take action related to their concerns and issues. These actions may involve countering existing power structures and institutions, just as counterdata do. For example, in Stuttgart, Germany, the data collection activities of the local civic tech initiative Luftdaten (now renamed to Sensor.Community) began by monitoring air quality because the municipality had installed only two air quality data collection points in the entire city. The city's reluctance to collect air pollution information was

linked to the fact that the automotive industry is the city's mainstay. Thus, to overcome the lack of air pollution information, the civic tech initiative, on its own, began installing IoT sensor devices to monitor air pollution levels. This bottom-up data collection and Luftdaten's visualisation of the data on their website became the basis for counter-narratives related to the city's reluctance to address air pollution.

However, in most cases, civic data may not be used as counterdata. Rather, many civic data initiatives work in collaboration with established entities, including local governments. For example, the core members of Safecast, the civic tech initiative that collected radiation data after the Fukushima nuclear disaster, were approached by local politicians and governments who were also eager to find missing radiation information in the communities (Hamm et al., 2021). Because the civic tech initiative and public sectors shared both a common concern and the objective of filling in the missing data for the cities, they collaborated to collect more data and disseminate the radiation information to communities. Similarly, the relationships of other civic tech initiatives with the public sector are collaborative in nature (Harrell, 2020; Le Dantec & DiSalvo, 2013; Shibuya et al., 2021). These examples prompt underexplored questions, such as how these civic data production initiatives contest and collaborate with previously underrepresented community members, such as noncitizens, indigenous populations and marginalised communities.

4. Can civic data be open?

Transparency is highly critical in civic data; thus it may seem that civic data may be intuitively understood as open data, yet this has not been always the case. For example, publishing data can be complex in cases where a need exists for independent verification if that data and information has been produced and maintained by powerful actors, such as the state or private corporations. The processes of institutional capture and institutional bypass lead to the unfettered exercise of executive power, even in functional democracies such as India in the Global South (Anderson et al., 2020; Chatterjee et al., 2019; Hansen et al., 2020; Jaffrelot, 2021; Komireddy, 2019). In addition, in some cases, opening up data comes with risks of being attacked by those who are unhappy about the publication of data. In particular, civic tech projects in non-democratic countries or underrepresented communities may face such challenges. Furthermore, some instances have difficulty

in anonymizing data due to the size, quality or nature of the dataset, or people are simply unwilling to share the data. Depending on the local context and a project's focus, participants engaged in data collection can represent a narrower demographic. For example, people considered cyclist data collected in a small town in the UK to be non-representative, because the cyclist population in the town was limited and less diverse (Maskell et al., 2018). It should also be noted that infrastructure for open data is resource-consuming. There is a lack of sustainable infrastructure for creating and maintaining the data structure so that anyone can access and use it. How open civic data can exist is a complex social and technical matter that brings us to question general, technical and universal understandings of open data that too often forget about the social risks and implications of making data transparent/open.

5. Can open data be civic data?

The skewness implicit in the production of data and knowledge represents another issue (Zuiderwijk et al., 2014). For instance, in the context of the Global South, skewing can impair the production of reliable and widely recognised civic data. Therefore, the question that must be asked is not just whether civic data should be open, but also whether open data can and should be civic. In a situation where the capacities to produce civic data are mediated through socially constructed obstacles, it behoves us to recognise and acknowledge this skewing. It is also incumbent upon us to recognise the new ways civic data plays out. The production, distribution and consumption of 'civic' data in ways that could feed into and strengthen undemocratic online spaces force us to rethink simplistic relationships between civic and open data. The opening of civic data, even if it circumvents the problems highlighted earlier, must be analysed in this context. Moreover, this opening must be rooted in questions of power: what do different articulations and frameworks of data tell us about the exercise of power and sociopolitical dynamics underlying the production of data? If data are 'open', what are the conditions that could also allow it to be classified as 'civic'?

If the production of and access to civic data are seen as enablers of democratic processes and as integral components of the process of allowing hitherto marginalised voices into the decision-making sphere, then civic data will often not co-exist as open data. There are structural, technological,

ethical and legal constraints that must be addressed to make the use of open data more effective (Bigagli & Nativi, 2017; Wessels et al., 2014, 2017). Open data can often act in ways that further strengthen underlying hierarchical social structures and historical inequities (Bigagli & Nativi, 2017; Wessels et al., 2014, 2017). Because open data mandates that data should be universally available to use, reuse and redistribute, this opens up space for often problematic majoritarian narratives to propagate if the data that are being used and distributed allow for the spread of widely accepted biases and prejudices (Chêne & Vrushi, 2020; Kirdemir, 2020). The means of producing and maintaining open data can allow for the undermining of myriad voices in the margin; this can produce and reproduce hegemonic narratives and, in so doing, be counterproductive to the deeper democratization of society.

6. Discussion

a) Collaboration between diverse stakeholders, including big tech companies

Despite the rising availability of easy-to-implement technological solutions to collect, analyse and visualize data, most communities may struggle to contribute to civic data. Many local communities lack the skills and resources to work with/on data. Sometimes, they need to use solutions available from big tech companies. In such cases, for example, a data partnership or agreement between said community and the big corporation is crucial. The question that remains is how local communities negotiate with big tech/government. These questions should be carefully handled because of the political hierarchy that is implied within their relations. Because of the issue of access to viable datasets, the need to work together with big tech will also be continuous. If communities decide not to use big tech solutions, civic tech initiatives themselves need to establish and maintain alternative schemes to support their collection, analysis, distribution and maintenance of data.

Data partnerships involving the state must be equally cognisant of the sociopolitical dynamics involved and the varying nature of citizen–state relationships in different contexts across the globe. In the context of the lack of functional democratic systems and protective checks and balances, data partnerships that involve the state can come with several (possibly unintended) consequences. In the deeply networked and technology-driv-

en contemporary era, the state is clearly capable of generating, analysing and utilizing large amounts of data. However, it is equally true that civil society is increasingly discovering the potential it has to enter into this domain of information gathering and analysis. As we see the systematic weakening of legislation intended to allow for more transparent oversight of the state (e.g. the Right to Information [RTI] in India), the role of independent data gathering will assume even more significance. Although the importance of broader participation in the policy-making process has been widely acknowledged (Chatterji et al., 2019), it is also important to ensure more collaboration and more widespread availability of information. The importance of numbers and data is undeniable; good data can empower local communities and improve the process of governance (Rukmini, 2021). Data, however, hardly comprise a neutral entity and they are often the subject of intense contestations, as social and economic forces attempt to exert their hegemony (Rukmini, 2021). The terms and conditions of collaborations between local communities and big tech/government may play a crucial role in dictating the outcomes. Co-production and co-management of data regimes would require the presence of active protections for non-state actors and transparent, horizontal decision-making structures of management.

b) Maintaining ownership and understanding the context

A recent trend of passive data collection (e.g. purchasing GPS data, rather than installing data collection devices with local community participation) may run the risk of diminishing people's sense of ownership over these data (Shibuya et al., 2021). Although passive data collection offers its own set of advantages, including efficiency and scalability, it may inadvertently weaken individuals' participation levels and undermine transparency in data collection processes. Without sufficient transparency in collection, publication and maintenance procedures, the resultant data may lack the essential qualities necessary for catalysing civic action.

Moreover, it is essential to recognise that civic data collection extends beyond the mere aggregation of raw data points; it encompasses the generation of context-rich data sets (Maskell et al., 2018; Taylor et al., 2015). Such contextual information enriches the understanding and usability of the data, empowering stakeholders to derive meaningful insights and drive informed decision-making processes. Therefore, initiatives aimed at foster-

ing civic engagement through data collection must prioritise transparency, active participation and the integration of contextual elements to maximise the utility and impact of the generated datasets.

7. Conclusion

The intersections of and differences between civic data, counterdata and open data present a complex landscape wherein various concepts and initiatives coalesce, often with overlapping objectives, yet distinct methodologies and outcomes. Although each concept embodies distinct principles and approaches, their similarities underscore the complex dynamics inherent in data-driven governance and civic engagement. Civic data have played critical roles in various communities, yet the dynamics surrounding data ownership, usage and transparency raise critical questions about the nature and potential of these data-driven approaches. For realizing the transformative potential of data in advancing democratic ideals and social justice agendas, transparency, inclusivity and contextual understanding remain imperative. Further research is necessary to understand the intersections and differences of civic data, counterdata, and open data to achieve wider societal impacts and maintain and strengthen participatory structures in the datafied society.

Acknowledgements

References

Alvarado Garcia, A., Young, A. L., & Dombrowski, L. (2017). On making data actionable: How activists use imperfect data to foster social change for human rights violations in Mexico. *Proceedings of the ACM on Human-Computer Interaction, 1*(CSCW) (pp. 1–19). https://doi.org/10.1145/3134654

Anderson, E., & Longkumer, A. (Eds.). (2020). *Neo-Hindutva: Evolving forms, spaces, and expressions of Hindu nationalism*. Routledge.

Bigagli, L., & Nativi, S. (2017). Mobilising data: Environmental data, technical and governance issues. In B. Wessels, R. Finn, T. Sveinsdottir, & K. Wadhwa (Eds.), *Open data and the knowledge society* (pp. 121–140). Amsterdam University Press.

Burrell, J., Singh, R., & Davison, P. (2024). *Keywords of the datafied state*). http://dx.doi.org/10.2139/ssrn.4734250

Cerratto Pargman, T., Joshi, S., & Wehn, U. (2019). Experimenting with novel forms of computing: The case of the Swedish Citizen Observatory for Water Quality Conservation. *LIMITS '19: Proceedings of the Fifth Workshop on Computing within Limits)* (pp. 1–10). https://doi.org/10.1145/3338103.3338111

Chatterji, A., Hansen, T., & Jaffrelot, C. (2019). *Majoritarian state: How Hindu nationalism is changing India*. Oxford University Press.

Chêne, M., & Vrushi, J. (2020). *Getting ahead of the curve: Exploring post-Covid-19 trends and their impact on anticorruption, governance and development*. Transparency International.

Dawes, S., & Helbig, N. (2010). Information strategies for open government: Challenges and prospects for deriving public value from government transparency. In M. A. Wimmer, J. L. Chappelet, M. Janssen, & H. J. Scholl (Eds.), *Electronic government. EGOV 2010. Lecture notes in computer science* (p. 6228). Springer.

Dunn, P. (2016). Open data: How access to public data is growing in Detroit. Model D. https://www.modeldmedia.com/features/open-data-090516.aspx

D'Ignazio, C., Cruxên, I., Suárez Val, H., Martinez Cuba, A., García-Montes, M., Fumega, S., Suresh, H., & So, W. (2022). Feminicide and counterdata production: Activist efforts to monitor and challenge gender-related violence. *Patterns, 3*(7), Article 100530. https://doi.org/10.1016/j.patter.2022.100530

Gao, Y., & Janssen, M. (2022). The open data canvas–Analyzing value creation from open data. *Digital Government: Research and Practice, 3*(1), 1–15.

Hamm, A., Shibuya, Y., Ullrich, S., & Cerratto Pargman, T. C. (2021). What makes civic tech initiatives to last over time? Dissecting two global cases. *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). https://doi.org/10.1145/3411764.3445667

Hansen, N. B., Klerks, G., Menendez Blanco, M., Maye, L., Strohmayer, A., De Waal, M., & Schouten, B. (2020). Making civic initiatives last: Ecosystems, technologies, approaches and challenges. *DIS '20: Proceedings of Companion Publication of the 2020 ACM Designing Interactive Systems Conference* (pp. 433–436). http://dx.doi.org/10.1145/3393914.3395921

Harrell, C. (2020). *A civic technologist's practice guide*. Five Seven Five Books.

Jaffrelot, C., (2021). *Modi's India: Hindu nationalism and the rise of ethnic democracy*. Princeton University Press.

Kim, S., Robson, C., Zimmerman, T., Pierce, J., & Haber, E. M. (2011). Creek watch: Pairing usefulness and usability for successful citizen science. *CHI '15: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2125–2134). https://doi.org/10.1145/1978942.1979251

Kirdemir, B. (2020). *Exploring turkey's disinformation ecosystem*. Centre for Economics and Foreign Policy Studies.

Komireddy, K. S. (2019). *Malevolent republic: A short history of the new India*. Westland.

Le Dantec, C. A., & DiSalvo, C. (2013). Infrastructuring and the formation of publics in participatory design. *Social Studies of Science, 43*(2), 241–264. https://doi.org/10.1177/0306312712471581

Le Dantec, C. A., Farrell, R. G., Christensen, J. E., Bailey, M., Ellis, J. B., Kellogg, W. A., & Edwards, W. K. (2011). Publics in practice: Ubiquitous computing at a shelter for homeless mothers. *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1687–1696). https://doi.org/10.1145/1978942.1979189

Maskell, T., Crivellaro, C., Anderson, R., Nappey, T., Araújo-Soares, V., & Montague, K. (2018). Spokespeople: Exploring routes to action through citizen-generated Data. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). https://doi.org/10.1145/3173574.3173979

Meng, A., & DiSalvo, C. (2018). Grassroots resource mobilization through counter-data action. *Big Data & Society, 5*(2), Article 2053951718796862. https://doi.org/10.1177/2053951718796862

Rukmini, S. (2021). *Whole numbers and half-truths: What data can and cannot tell us about modern India*. Wetland Publications.

Schrock, A. R. (2019). What is civic tech? Defining a practice of technical pluralism. In P. Cardullo, C. Di Feliciantonio, & R. Kitchin (Eds.), *The right to the smart city* (pp. 125–133). Emerald. https://doi.org/10.1108/978-1-78769-139-120191009

Shibuya, Y., Hamm, A., & Raetzsch, C. (2021). From data to discourse: How communicating civic data can provide a participatory structure for sustainable cities and communities. *Proceedings of the 27nd ISDRS '21*. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-74871-6

Shibuya, Y., Lai, C. M., Hamm, A., Takagi, S., & Sekimoto, Y. (2022). Do open data impact citizens' behavior? Assessing face mask panic buying behaviors during the COVID-19 pandemic. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-22471-y

Taylor, A S., Lindley, S., Regan, T., Sweeney, D., Vlachokyriakos, V. Grainger, L., & Lingel, J. (2015). Data-in-place: Thinking through the relations between data and community. *Proceedings of the 2015 CHI Conference* (pp. 2863–2872). https://doi.org/10.1145/2702123.2702558

Taylor, N., Clarke, L., & Gorkovenko, K. (2018). Community inventor days: Discussing data and creating speculative internet of things concepts. *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 747–758). https://doi.org/10.1145/3196709.3196772

Wehn, U., & Evers, J. (2014). Citizen observatories of water: Social innovation via eParticipation. *In Proceedings of the 2014 conference ICT for Sustainability* (pp. 1–10). Atlantis Press.

Wells, C. (2015). *The civic organization and the digital citizen: Communicating engagement in a networked age.* Oxford University Press.

Wessels, B., Linde, P., Mazzetti, P., Nativi, S., Riley, S., Smallwood, R., Taylor, M. J., Tsoukala, V., Wadhwa, K., & Wyatt, S. (2014). Issues in the development of open access to research data. *Prometheus: Critical Studies in Innovation, 32*(1), 49–66. http://dx.doi.org/10.1080/08109028.2014.956505

Wessels, B., Finn, R. L., Wadhwa, K., & Sveinsdottir, T. (2017). Navigating legal and ethical frameworks. In B. Wessels, R. L. Finn, K. Wadhwa, & T. Sveinsdottir (Eds.), *Open data and the knowledge society* (pp. 141–158). Amsterdam University Press.

Williams, S. (2020). *Data action: Using data for public good.* The MIT Press.

Zuiderwijk, A., & Janssen, M., (2014). The negative effects of open government data-investigating the dark side of open data. *Proceedings of the 15th annual international conference on digital government research* (pp. 147–152). https://doi.org/10.1145/26127 33.2612761