Participatory to what end? Mapping motivations for participatory approaches in data-driven projects

Judith Fassbender
judith.fassbender@hiig.de
University of St Andrews
St Andrews, UK
Humboldt Institute for Internet and Society
Berlin, Germany

Tristan Henderson
tnhh@st-andrews.ac.uk
University of St Andrews
St Andrews, UK

ABSTRACT

Participation in data-driven projects is a popular approach and often connected to the idea of more equitable projects. The lack, however, of an agreed definition of what constitutes participation leads to fuzziness surrounding possible motivations for participation. This in turn diminishes the ability of facilitators to communicate what to expect from a participatory process to participants and the public. To better understand this, we conduct a systematic literature review and analyse the claimed motivations for implementing participation in data-driven projects. We find three overarching categories: value-, effectiveness-, and efficiency-focused motivations. We discuss overlaps and issues within these categories, such as the implications of project-internal demands (the realisation and working of a project) and project-external demands (codified demands in frameworks, policies and rights).

CCS CONCEPTS

• **Human-centered computing** → *Computer supported cooperative work*; • **Applied computing** → Digital libraries and archives.

KEYWORDS

participation, data-driven, participatory data governance

ACM Reference Format:
Judith Fassbender and Tristan Henderson. 2024. **Participatory to what end? Mapping motivations for participatory approaches in data-driven projects**. In *International Conference on Information Technology for Social Good (GoodIT '24), September 04–06, 2024, Bremen, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3677525.3678675

1 INTRODUCTION

Participation is a popular approach in many fields, and data-driven projects are no exception. Including external participants is often connected to the idea of better systems and producing fairer or more equitable technologies [7, 14, 16]. At the same time, using participatory approaches for public relations or exploitative reasons is cautioned against [28], and the potential of participatory approaches to disseminate power is critically examined [5]. Yet, it is often noted that there is no agreement on a definition of what constitutes participation [5, 9]. Better knowledge of the variety of reasons for participation can help facilitators to better design and evaluate *problem-oriented* participatory processes, transparently communicate towards participants and the public, and potentially avoid participatory processes for their own sake. To understand the motivations for participatory processes, we look at data-driven projects with participatory elements from the academic literature and analyse the motivations discussed. We understand participation for the purpose of this paper pragmatically—as the deliberate inclusion of external parties in data-related aspects of a project. Data-driven projects are understood here as projects that rely on data for their functioning. We find three categories of motivations, concentrated on *values*, *effectiveness*, and *efficiency*. Based on our findings, we discuss overlaps between the categories and implications of the project's internal and external interests and demands.

2 BACKGROUND

Participation in data-driven projects has been encouraged for some time and can be found in political forms of participation e.g. since the 1970s by workers regarding technologies used at the workplace [8] as well as in citizen science approaches, which have been around since the 1990s [22]. Sloane et al. [28] note that, for the field of AI, participation is not necessarily a means to reach greater *justice*, but also for simply *consulting* participants or using participation for *work*-related tasks. More recently, we see projects where participation itself is a central goal, e.g. in data-driven projects related to equity or self-determination [11, 32]. At the same time, Groves et al. [17] collate factors that hinder the implementation of public participation in commercial AI labs. Whilst AI and Machine Learning (ML) systems are a focal point of much current research, we believe that there is value in looking more widely at data-driven projects in general — to capture knowledge from a broader variety of cases and to derive findings that apply to a broader field of projects. Our main research question is: What does the academic literature claim about why participatory processes are implemented in data-driven projects? By analysing the motivation of facilitators to implement participation, we aim to contribute to a better understanding of the participatory processes involved and the problems they are intended to solve.

3 METHOD

To identify data-driven projects with a participatory element, we conducted a systematic literature review, following Kitchenham and Charters [3]. We constructed two search strings. One combined

GoodIT '24, September 04–06, 2024, Bremen, Germany

ACM ISBN 979-8-4007-1094-0/24/09
https://doi.org/10.1145/3677525.3678675

participation-related attributes (e.g. participatory, deliberative, collaborative, fair, community, democratic) with modes of data handling (data governance, data stewardship, data curation). The other searched for normative modes of data handling (data stewardship, data commons, data trust, data care), which potentially included participation. We searched the *ACM Digital Library*, *IEEE Xplore*, and *Scopus* on 20/09/2022 for matching titles and abstracts published between 2012 and 2022. Our initial sample of 1,642 papers was reduced to 21 papers, describing 27 cases. The sample reduction was conducted as follows: removal of duplicates, unrelated titles, as well as unrelated abstracts and finally the removal of unrelated items based on full papers. Included were complete, peer-reviewed papers, written in English, that described implemented cases of data-driven projects which included a participatory element. We identified the motivations for participation using a qualitative content analysis, coding passages in the papers indicating *why* a participatory approach was included. In some cases, motivations are named explicitly—in others implicitly. Furthermore, a project can have more than one motivation.

4 RESULTS

Our 27 cases include projects that are collections of health data for scientists [4, 13, 15, 18, 21, 25, 30, 31, 33], research projects [1, 2, 10, 23, 24, 26, 27], data-driven services and products (mapping and encyclopedia-connected projects) [6, 19, 20], and data activism initiatives [12, 34]. We find three types of motivation.

Motivations to encourage values. Motivations that encourage the implementation of *values*, such as trust and accountability, are concentrated on the participants' or the public's needs and wishes. Through their implementation, facilitators may increase the legitimacy of their projects.

The motivation to evoke *'trust' in the project* is frequently mentioned [13, 15, 25, 29, 31], which is anticipated to be higher if participation takes place. Increasing or understanding the level of trust towards a project is entangled with e.g. the expectation of greater success for the project objective: "These mechanisms are useful for understanding the public perceptions of using data to develop and test algorithms, and the level of trust felt towards algorithms in clinical care supporting the role of human readers." [13]. Participatory processes are thus a reaction to a known or anticipated mistrust; the reason for this mistrust is therefore not necessarily an issue in the perception of the facilitators, but in the perception of the participants or other externals. In terms of accountability, this can be difficult, if facilitators do not consciously address the danger of obscuring responsibilities. Participation for *increased accountability* is mentioned in connection to a jointly drafted Data Governance Agreement (DGA): "Well-designed DGAs [...] support equitable relationships by increasing both transparency and accountability toward Indigenous partners." [23]. Although the facilitators of a data activism case ask themselves how to "[. . .] make data and decision-making more accessible and accountable?" [34] the topic of accountability is mentioned seldom in the sample.

In several cases, the motivations can be understood as a means to *align aspects of a project with the interests and needs of a specific group* other than the project facilitators. Those groups are Indigenous communities [23, 27], patients [13, 15, 31], or—more general—stakeholders/experts[25, 29]; in one case, the aim is to align the project with the "public interest" [15]. With regards to Indigenous communities and patients, authors point at the *rights* of Indigenous people [27] and the control of patients' *own* health data:"[. . .] participants are cast as a community that has interests and entitlements in controlling its data." [31].

Motivations to improve effectiveness. Motivations to improve the *effectiveness* of projects via participation are focused on how a project works. This is connected to the data, data handling in terms of curation, and its implementation in terms of design — as well as the work of the community of users, participants, or contributors. This category is focused on the project in its realisation and functions.

The motivation for including participants in the *data collection or data curation* is connected to realising the project as such by the contribution of labour, knowledge, or existing data. The motivation for participation in this context is to strengthen the general purpose of the project, e.g. to make research in a specific domain possible in the first place: "DPUK [Dementias Platform UK] was established by the Medical Research Council (MRC) to accelerate the development of new treatments for dementia." [4]. Several cases intend to *foster community interaction and collaboration amongst participants* who shall ideally work collaboratively or become an integral part of the project [4, 20, 25, 30, 33]. This motivation occurs frequently in research environments and is reflected in passages like the following:"Finally, the Data Commons was designed as a place for community interaction [. . .]. Discussion forums enable a lively exchange of ideas and expertise, with users posting questions and answers on a variety of technical, policy, and other issues." [20]. A further motivation is facilitators aiming to increase the effectiveness of their project by gathering input from users *to alter the functionality/design of a project*, so that it becomes of better use:"These [four stages of development each] consist on performing a requirement analysis (e.g., ask the community what data needs to be shared), followed by a period of design and development of tools and policies, and a period of feedback (testing) by the users and the community." [30]

Motivations to improve efficiency. Implementations of participatory processes motivated by improving the *efficiency* of a project are focused mainly on realising a project with fewer burdens or resources. This category revolves around necessary but resource-intensive aspects of the project.

Participation, especially regarding *laborious tasks*, is depicted as a way to make a project run whilst *keeping costs lower* than e.g. working with contractors or to help realising an underfunded endeavour [6, 12, 20, 26]. Many of the projects in the sample do not seem to have a business objective but rather aimed at a collective interest: "Ideally, the Data Commons could productively harness some of the community's attention, directing a little effort from many participants toward collective tasks." [20]. Besides putting parts of the labour in many participants' hands, *prohibitive logistical burdens* play a role in data collection in dispersed areas. Overcoming these is a second motivation for participation. Particularly in environmental research projects, this challenge is solved through a participatory and distributed approach to the data collection [1, 10, 26]: "Collaborative science was the best way to overcome the prohibitive logistical field and laboratory requirements associated with answering this question." [1]. Finally, several projects centralise data collections to be more efficient and *share infrastructure as well as technical knowledge* [2, 18, 21, 26]. By doing so, the facilitators aim to generally save costs. This includes allowing organisational participants to partake that would not have the financial resources to set up infrastructures themselves: "Our key motivation for this approach was cost-effectiveness and efficiency of the network: many OneFlorida partners did not have existing research infrastructure on which to build an independent PCORnet data mart." [18]

5 DISCUSSION

In the following we discuss how these motivations are related to each other as well as to the projects' internal and external demands.

Intersections of value-, effectiveness-, and efficiency-driven motivations. Value-, effectiveness-, and efficiency-oriented motivations can be understood as high-level categories. They overlap, and a project may have several motivations for applying participatory approaches. This raises questions about the relations between the categories and their role when coming together in one project.

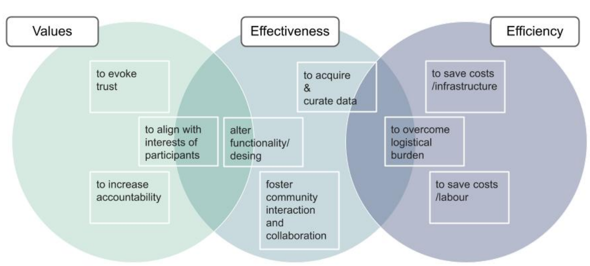

Figure 1: Visualisation of the three overlapping motivations to implement participatory approaches in data-driven projects. Each category entails examples from the cases.

Figure 1 shows that overlaps between the categories exist in at least two points. One lies between value- and effectiveness-focused motivations, e.g. in the federally-funded Pathways TB Project [23], which is focused on writing data governance agreements together with Indigenous communities for collecting, sharing, and storing health data on tuberculosis outbreaks in a co-ownership approach. One result of the process was improved usability of the data for the Indigenous partners. The data format, which was hard to access before, was now changed so the communities could make sense of local outbreaks and provide additional countermeasures as well as data [23]. The motivation was to realise Indigenous data sovereignty and align the project with communities' interests—a value-based motivation—which in the same vein affects the effectiveness of the project. Yet, aligning the participants and the project interests seems crucial for this overlap. Further, we do not see a value-efficiency overlap, which could point to a general conflict between both. This raises the question of *why a value-efficiency overlap does not occur.* A second area of overlap lies between effectiveness- and efficiency-driven motivations, e.g. overcoming logistical burdens can support both effectiveness and efficiency. An example is a bat monitoring program [26] which coordinated data collection at dispersed sampling sites. The collection of data by many professional and lay participants in different locations made the project possible in terms of the needed resources but also enabled the functioning of the project in the first place. The variety of the sampling sites strengthened its effectiveness. It is necessary to look further into the interplay between different participatory approaches and *how to communicate more complex combinations of motivations*—without omitting e.g. resource benefits for facilitators or value-related benefits of effectiveness-focused participation for participants.

Internal project demands and external participant interests. When contextualising the three high-level motivations, we see that effectiveness- and efficiency-focused motivations evolve out of project demands, which are challenges to realise a project in the first place. This relates to how the project functions and how facilitators deal with burdens to realise the project. Effectiveness- and efficiency-focused motivations for participation can thus be considered *project-internal.* Value-driven motivations, on the other hand, seem to be concerned with either the interests of participants and responsibilities towards the public or the external perception of the project—which are all rather *project-external* factors. Whilst the participant's/public's needs are in focus, a normative benefit for the project is supposedly expected: public legitimisation. In connection to value-driven participation, codified principles are mentioned, e.g. the UNDRIP, OCAP® for first nations [23] or the CARE principle [27]. Further, the legal ownership of data subjects of their data [31] is mentioned. Similarly, value-focused participation is used to counterbalance the lack of otherwise established modes of legitimisation. In one case, participation regarding a data access request protocol was implemented after health data was used without consent, which was possible due to specific COVID-19-related exceptions [13]. Participation was used here as a substitute for a recognised practice to gain legitimacy. This raises the *question of the role codified principles play in the decisions to implement participatory processes that are value-oriented.* In comparison to the other motivations, the implementation of value-focused participation seems to benefit from external, codified, and enforced demands. It is not as integral to the project as effectiveness and efficiency are.

Finally, it needs to be noted that we are talking about the implementation of participatory approaches on a predominantly voluntary basis. This raises the question of where the *limits of participation as a means to answer internal and external demands are and where other approaches are needed.*

6 CONCLUSION

In this paper, we show that motivations for participation in data-driven projects can be divided into three high-level, overlapping categories: value-, effectiveness-, and efficiency-oriented. Projects can have more than one motivation for implementing participation. We find such overlaps between value- and effectiveness-driven motivations and between effectiveness- and efficiency-driven motivations. We raise the question of overlaps that are not possible, potentially between efficiency- and value-oriented participation. The overlaps require a contextualised evaluation in implemented projects to enable a transparent communication of the motivations towards participants and the public. We point out that effectiveness- and efficiency-related motivations seem closely connected to project internal demands whilst value-oriented participation is often mentioned with references to external demands such as policies, rights, or frameworks. Further research regarding the role of such codified principles for facilitators to implement value-oriented participation is necessary to understand their potential to strengthen value-oriented participation. Lastly, as we pointed out that participation is not the only way to implement values in data-driven projects, it seems imperative to look into when participation is not the right approach and what an interplay with other approaches can look like. To deal with some of the questions raised, we intend to conduct in-depth case studies of participatory data-driven projects, by looking at the public communication, the interfaces for participation, and conducting interviews with facilitators. We will further examine the cases analysed in this paper by reviewing their openness to participation such as the used participatory processes regarding data governance and data handling.

ACKNOWLEDGMENTS

We thank Irina Kühnlein for her great assistance and the anonymous reviewers for their valuable comments. This work was partially made possible by funding of the German Federal Ministry of Education and Research. We have no conflicts of interest to declare.

REFERENCES

[1] Isabelle Aubin, Françoise Cardou, Laura Boisvert-Marsh, Eric Garnier, Manuella Strukelj, and Alison D. Munson. 2020. Managing data locally to answer questions globally: The role of collaborative science in ecology. *Journal of Vegetation Science* 31, 3 (May 2020), 509–517. https://doi.org/10.1111/jvs.12864

[2] Karen S. Baker and Helena Karasti. 2018. Data care and its politics: designing for local collective data management as a neglected thing. In *Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1*. ACM, Hasselt and Genk Belgium, 1–12. https://doi.org/10.1145/3210586.3210587

[3] Barbara Kitchenham and Stuart Charters. 2007. Guidelines for performing systematic literature reviews in software engineering. https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf

[4] Sarah Bauermeister, Christopher Orton, Simon Thompson, Roger A. Barker, Joshua R. Bauermeister, Yoav Ben-Shlomo, Carol Brayne, David Burn, Archie Campbell, Catherine Calvin, Siddharthan Chandran, Nishi Chaturvedi, Geneviève Chêne, Iain P. Chessell, Anne Corbett, Daniel H. J. Davis, Mike Denis, Carole Dufouil, Paul Elliott, Nick Fox, Derek Hill, Scott M. Hofer, Michele T. Hu, Christoph Jindra, Frank Kee, Chi-Hun Kim, Changsoo Kim, Mika Kivimaki, Ivan Koychev, Rachael A. Lawson, Gerry J. Linden, Ronan A. Lyons, Clare Mackay, Paul M. Matthews, Bernadette McGuiness, Lefkos Middleton, Catherine Moody, Katrina Moore, Duk L. Na, John T. O'Brien, Sebastien Ourselin, Shantini Paranjothy, Ki-Soo Park, David J. Porteous, Marcus Richards, Craig W. Ritchie, Jonathan D. Rohrer, Martin N. Rossor, James B. Rowe, Rachael Scahill, Christian Schnier, Jonathan M. Schott, Sang W. Seo, Matthew South, Matthew Steptoe, Sarah J. Tabrizi, Andrea Tales, Therese Tillin, Nicholas J. Timpson, Arthur W. Toga, Pieter-Jelle Visser, Richard Wade-Martins, Tim Wilkinson, Julie Williams, Andrew Wong, and John E. J. Gallacher. 2020. The Dementias Platform UK (DPUK) Data Portal. *European Journal of Epidemiology* 35, 6 (June 2020), 601–611. https://doi.org/10.1007/s10654-020-00633-4

[5] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the People? Opportunities and Challenges for Participatory AI. In *Equity and Access in Algorithms, Mechanisms, and Optimization*. ACM, Arlington VA USA, 1–8. https://doi.org/10.1145/3551624.3555290

[6] Nicholas Bolten and Anat Caspi. 2022. Towards operationalizing the communal production and management of public (open) data: a pedestrian network case study: A pedestrian network case study in operationalizing communal open data. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS)*. ACM, Seattle WA USA, 232–247. https://doi.org/10.1145/3530190.3534821

[7] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A. Killian. 2021. Envisioning Communities: A Participatory Approach Towards AI for Social Good. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (July 2021), 425–436. https://doi.org/10.1145/3461702.3462612 arXiv: 2105.01774.

[8] Susanne Bødker and li Pekkola. 2010. Introduction the debate section: A short review to the past and present of participatory design. *Scandinavian Journal of Information Systems* 22, 1 (2010), 45–48.

[9] Nico Carpentier. 2016. Beyond the Ladder of Participation: An Analytical Toolkit for the Critical Analysis of Participatory Media Processes. *Javnost - The Public* 23, 1 (Jan. 2016), 70–88. https://doi.org/10.1080/13183222.2016.1149760

[10] Cynthia M. Cohen, Liz Cheney, Khue Duong, Ben Lea, and Zoe Pettway Unno. 2015. Identifying Opportunities in Citizen Science for Academic Libraries. *Issues in Science and Technology Librarianship* 79 (March 2015). https://doi.org/10.29173/istl1629

[11] MIDATA Cooperative. [n. d.]. About us. https://www.midata.coop/en/cooperative/

[12] Morgan E. Currie, Britt S. Paris, and Joan M. Donovan. 2019. What difference do data make? Data management and social change. *Online Information Review* 43, 6 (Oct. 2019), 971–985. https://doi.org/10.1108/OIR-02-2018-0052

[13] Dominic Cushnan, Rosalind Berka, Ottavia Bertolli, Peter Williams, Daniel Schofield, Indra Joshi, Alberto Favaro, Mark Halling-Brown, Gergely Imreh, Emily Jefferson, Neil J Sebire, Gerry Reilly, Jonathan C L Rodrigues, Graham Robinson, Susan Copley, Rizwan Malik, Claire Bloomfield, Fergus Gleeson, Moira Crotty, Erika Denton, Jeanette Dickson, Gary Leeming, Hayley E Hardwick, Kenneth Baillie, Peter Jm Openshaw, Malcolm G Semple, Caroline Rubin, Andy Howlett, Andrea G Rockall, Ayub Bhayat, Daniel Fascia, Cathie Sudlow, NCCID Collaborative, and Joseph Jacob. 2021. Towards nationally curated data archives for clinical radiology image analysis at scale: Learnings from national data collection in response to a pandemic. *DIGITAL HEALTH* 7 (Jan. 2021), 205520762110486. https://doi.org/10.1177/20552076211048654

[14] Gregory Falco. 2019. Participatory AI: Reducing AI Bias and Developing Socially Responsible AI in Smart Cities. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. IEEE, New York, NY, USA, 154–158. https://doi.org/10.1109/CSE/EUC.2019.00038

[15] Suzy Gallier, Gary Price, Hina Pandya, Gillian McCarmack, Chris James, Bob Ruane, Laura Forty, Benjamin L Crosby, Catherine Atkin, Ralph Evans, Kevin W Dunn, Eliot Marston, Clark Crawford, Martin Levermore, Shekha Modhwadia, John Attwood, Stephen Perks, Rima Doal, Georgios Gkoutos, Richard Dormer, Andy Rosser, Hilary Fanning, and Elizabeth Sapey. 2021. Infrastructure and operating processes of PIONEER, the HDR-UK Data Hub in Acute Care and the workings of the Data Trust Committee: a protocol paper. *BMJ Health & Care Informatics* 28, 1 (April 2021), e100294. https://doi.org/10.1136/bmjhci-2020-100294

[16] Anne Gerdes. 2021. A participatory data-centric approach to AI Ethics by Design. *Applied Artificial Intelligence* (Dec. 2021), 1–19. https://doi.org/10.1080/08839514.2021.2009222

[17] Lara Groves, Aidan Peppin, Andrew Strait, and Jenny Brennan. 2023. Going public: the role of public participation approaches in commercial AI labs. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1162–1173. https://doi.org/10.1145/3593013.3594071

[18] William R Hogan, Elizabeth A Shenkman, Temple Robinson, Olveen Carasquillo, Patricia S Robinson, Rebecca Z Essner, Jiang Bian, Gigi Lipori, Christopher Harle, Tanja Magoc, Lizabeth Manini, Tona Mendoza, Sonya White, Alex Loiacono, Jackie Hall, and Dave Nelson. 2022. The OneFlorida Data Trust: a centralized, translational research data infrastructure of statewide scope. *Journal of the American Medical Informatics Association* 29, 4 (March 2022), 686–693. https://doi.org/10.1093/jamia/ocab221

[19] Timothy Kanke. 2020. Knowledge curation work in Wikidata WikiProject discussions. *Library Hi Tech* 39, 1 (Jan. 2020), 64–79. https://doi.org/10.1108/LHT-04-2019-0087

[20] Akos Kokai, Ann Blake, Michel Dedeo, and Tom Lent. 2020. Building shared information infrastructure for chemical alternatives assessment. *Elementa: Science of the Anthropocene* 8 (Jan. 2020), 26. https://doi.org/10.1525/elementa.422

[21] Emily M. Kraus, Kenneth A. Scott, Rachel Zucker, Dawn Heisey-Grove, Raymond J. King, Thomas W. Carton, Matthew F. Daley, Sara J. Deakyne Davies, Jason P. Block, Matthew Haemer, Alyson B. Goodman, Nedra Garrett, and Arthur J. Davidson. 2022. A Governance Framework to Integrate Longitudinal Clinical and Community Data in a Distributed Data Network: The Childhood Obesity Data Initiative. *Journal of Public Health Management and Practice* 28, 2 (March 2022), E421–E429. https://doi.org/10.1097/PHH.0000000000001408

[22] Christopher Kullenberg and Dick Kasperowski. 2016. What Is Citizen Science? – A Scientometric Meta-Analysis. *PLOS ONE* 11, 1 (Jan. 2016), e0147152. https://doi.org/10.1371/journal.pone.0147152

[23] Robin P. Love, Billie-Jo Hardy, Courtney Heffernan, Amber Heyd, Melissa Cardinal-Grant, Lori Sparling, Bonnie Healy, Janet Smylie, and Richard Long. 2022. Developing Data Governance Agreements with Indigenous Communities in Canada. *Health and Human Rights* 24, 1 (June 2022), 21–33. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9212824/

[24] Benjamin Murray, Eric Kerfoot, Liyuan Chen, Jie Deng, Mark S. Graham, Carole H. Sudre, Erika Molteni, Liane S. Canas, Michela Antonelli, Kerstin Klaser, Alessia Visconti, Alexander Hammers, Andrew T. Chan, Paul W. Franks, Richard Davies, Jonathan Wolf, Tim D. Spector, Claire J. Steves, Marc Modat, and Sebastien Ourselin. 2021. Accessible data curation and analytics for international-scale citizen science datasets. *Scientific Data* 8, 1 (Nov. 2021), 297. https://doi.org/10.1038/s41597-021-01071-x

[25] Alejandro Plana, Brian Furner, Monica Palese, Nicole Dussault, Suzi Birz, Luca Graglia, Maura Kush, James Nicholson, Stefanie Hecker-Nolting, Nathalie Gaspar, Mareike Rasche, Gianni Bisogno, Dirk Reinhardt, C. Michel Zwaan, Ewa Koscielniak, A. Lindsay Frazier, Katherine Janeway, Douglas S. Hawkins, E. Anders Kolb, Susan L. Cohn, Andrew D. J. Pearson, and Samuel L. Volchenboum. 2021. Pediatric Cancer Data Commons: Federating and Democratizing Data for Childhood Cancer Research. *JCO Clinical Cancer Informatics* 5 (Dec. 2021), 1034–1043. https://doi.org/10.1200/CCI.21.00075

[26] Brian E. Reichert, Mylea Bayless, Tina L. Cheng, Jeremy T. H. Coleman, Charles M. Francis, Winifred F. Frick, Benjamin S. Gotthold, Kathryn M. Irvine, Cori Lausen, Han Li, Susan C. Loeb, Jonathan D. Reichard, Thomas J. Rodhouse, Jordi L. Segers, Jeremy L. Siemers, Wayne E. Thogmartin, and Theodore J. Weller. 2021. NABat: A top-down, bottom-up solution to collaborative continental-scale monitoring. *Ambio* 50, 4 (April 2021), 901–913. https://doi.org/10.1007/s13280-020-01411-y

[27] Cathy J. Robinson, Taryn Kong, Rebecca Coates, Ian Watson, Chris Stokes, Petina Pert, Andrew McConnell, and Caron Chen. 2021. Caring for Indigenous Data to Evaluate the Benefits of Indigenous Environmental Programs. *Environmental Management* 68, 2 (Aug. 2021), 160–169. https://doi.org/10.1007/s00267-021-01485-8

[28] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation Is not a Design Fix for Machine Learning. In *Equity and Access in Algorithms, Mechanisms, and Optimization*. ACM, Arlington VA USA, 1–6. https://doi.org/10.1145/3551624.3555285

[29] C.G. Tai, J. Harris-Wai, C. Schaefer, P. Liljestrand, and C.P. Somkin. 2019. Multiple Stakeholder Views on Data Sharing in a Biobank in an Integrated Healthcare Delivery System: Implications for Biobank Governance. *Public Health Genomics* 21, 5-6 (2019), 207–216. https://doi.org/10.1159/000500442

[30] Abel Torres-Espín, Carlos A. Almeida, Austin Chou, J. Russell Huie, Michael Chiu, Romana Vavrek, Jeff Sacramento, Michael B. Orr, John C. Gensel, Jeffery S. Grethe, Maryann E. Martone, Karim Fouad, Adam R. Ferguson, the STREET-FAIR Workshop Participants, Warren Alilain, Mark Bacon, Nicholas Batty, Michael Beattie, Jacqueline Bresnahan, Emily Burnside, Sarah Busch, Randall Carpenter, Isaac Francos Quijorna, Xiaohui Guo, Agnes Haggerty, Sarah Haroon, Jack Harris, Lyn Jakeman, Linda Jones, Naomi Kleitman, Timothy Kopper, Michael Lane, Francisco Magana, David Magnuson, Ines Maldonado, Verena May, Katelyn McFarlane, Kazuhito Morioka, Martin Oudega, Philip Leo Pascual, Jean-Baptiste Poline, Ephron Rosenzweig, Emma Schmidt, Wolfram Tetzlaff, and Lana Zholudeva. 2022. Promoting FAIR Data Through Community-driven Agile Design: the Open Data Commons for Spinal Cord Injury (odc-sci.org). *Neuroinformatics* 20, 1 (Jan. 2022), 203–219. https://doi.org/10.1007/s12021-021-09533-8

[31] Effy Vayena and Alessandro Blasimme. 2017. Biomedical Big Data: New Models of Control Over Access, Use and Governance. *Journal of Bioethical Inquiry* 14, 4 (Dec. 2017), 501–513. https://doi.org/10.1007/s11673-017-9809-6

[32] Mozilla Common Voice. [n. d.]. Why Common Voice? https://commonvoice.mozilla.org/en/about

[33] Muhammad F Walji, Heiko Spallek, Krishna Kumar Kookal, Jane Barrow, Britta Magnuson, Tamanna Tiwari, Udochukwu Oyoyo, Michael Brandt, Brian J Howe, Gary C Anderson, Joel M White, and Elsbeth Kalenderian. 2022. BigMouth: development and maintenance of a successful dental data repository. *Journal of the American Medical Informatics Association* 29, 4 (March 2022), 701–706. https://doi.org/10.1093/jamia/ocac001

[34] Dawn Walker, Eric Nost, Aaron Lemelin, Rebecca Lave, and Lindsey Dillon. 2018. Practicing environmental data justice: From DataRescue to Data Together. *Geo: Geography and Environment* 5, 2 (July 2018), e00061. https://doi.org/10.1002/geo2.61

Received