Data & Policy (2024), 6: e12
doi:10.1017/dap.2023.43

RESEARCH ARTICLE

When data meets citizens: an investigation of citizen engagement in data-driven innovation programmes

Gefion Thuermer[1], Johanna Walker[1], Elena Simperl[1] and Les Carr[2]

[1]Department of Informatics, King's College London, London, UK
[2]School of Electronics and Computer Science, University of Southampton, Southampton, UK
Corresponding author: Johanna Walker; Email: Johanna.walker@kcl.ac.uk

Received: 11 February 2022; **Revised:** 16 October 2023; **Accepted:** 17 November 2023

Keywords: cascading funding; citizen engagement; data innovation; data justice; data justice plans

Abstract

Publicly funded data-driven innovation programmes frequently involve partnerships between small and medium enterprises (SMEs) and municipal authorities utilizing citizen data. The intention of these projects is to benefit citizens. However, few such projects achieve success or impact within the project timeframe. This may result in benefit accruing mainly to the SME partner, who gains both learning and data, engendering questions of data justice around whether citizen data are being exploited without sufficient benefit returning to citizens. Through case studies composed of interviews and document analysis, we examine how benefits for citizens are conceived and achieved in the publicly funded data-driven air quality projects Data Pitch and Smart Cities Innovation Framework Implementation. We find the differences between the programme funders' policies had a clear influence on the citizen engagement elements. There are also a number of ways in which the desired citizen engagement and benefit becomes diluted, including through misalignment of incentives and focus, a lack of prioritization and ownership, and power imbalances between citizens and the other actors in the quadruple helix model. To retain the focus on ensuring citizens benefit from data-driven innovation programmes using citizen data, we propose the use of data Justice plans. More work is required to specify the content and mechanisms of such plans for application in such programmes.

Policy significance statement

This work provides insight into how funding policies affect citizen engagement in funded data innovation projects. We find that generic goals for engagement and citizen benefit do not come to fruition, and argue that better specified responsibilities for grant receivers are required to ensure citizens consistently benefit from publicly funded data innovations.

1. Introduction

The European Commission's (EC) digital strategy describes data as, "the lifeblood of the economy and a driver of innovation" (European Commission, 2021). Substantial EC time and resources are invested into developing data-driven innovations across all industry sectors, which requires the availability of large amounts of data.

The EU has the strongest data protection laws in the world. Despite this, data gathering is a potential locus for "data colonialism" (Thatcher et al., 2016), with data flows disproportionately benefiting those who can make use of the data (Heeks and Shekhar, 2019), and the wide-spread perception that citizens

cannot control data collected about them (Turow et al., 2015). Sadowski (2019) suggests that common data gathering practices include "little regard for consent and compensation" in the extraction of data from every aspect of society. The city is an important locus for such activities, as data-driven innovation underpins much of the activities commonly understood under the "smart city" umbrella term (Coletta et al., 2018). How citizens are included in, or excluded from, the process of creating data-driven services or products that have the potential to affect them, is a key question for data justice.

"Data justice" identifies a research agenda that engages with the politics and impacts of data-driven processes and big data to establish how these affect issues of social justice. One analytic approach is to examine how social justice is changing in the context of datafication (e.g., Dencik et al., 2019). In this study, it is explored from the position of how data used in publicly funded programmes needs to engage with social justice in order to prevent unintended inequity. This provides a lens through which to understand the potential negative impacts of data beyond whether the use of the data is legally compliant or ethically acceptable. In this paper, we move beyond data gathering, to examine how citizens are included in, or excluded from, the process of creating data-driven services or products that have the potential to affect them. This is based on an extended case study of two air pollution projects. We further examine how public funders of innovation can create the structural conditions to facilitate greater citizen engagement and to ensure that due consideration is given to human rights in the data space, and these do not become eroded as technology evolves.

The resolution of air quality issues in part requires the cooperation of citizens, for instance, by reducing private vehicle journeys and using clean fuel sources. These individual actions require guidance at the national and local level, for example, through the development of policy, programmes, and services that enable citizens to make choices that contribute to better air quality. These approaches are driven by data. This data is essential to governments at all levels, but governmental organizations are rarely best placed to interpret, use, and in some cases collect it. Hence, they often seek partnerships with commercial or community organizations with the skills and knowledge to develop products, services, and insight for policy.

We examine the acquisition and use of data generated from one group of people (in this case citizens) by others (here commercial organizations) and ask whether the benefits (financial and otherwise) accrue equally, or whether the more powerful interest receives asymmetric benefit, sometimes to the detriment of the data sources. We focus on the broader social structures that lead to data (in)justice, examining the implementation of research and innovation (R&I) projects, including the institutions, resources, and relationships that affect how equity in data use is achieved in those projects. We are particularly interested in how data justice is, or should be, operationalized in the specific context of smart city public–private innovation processes. This paper makes three contributions to the literature:

1. It takes a nuanced look at data justice in publicly funded products and services in the smart city.
2. It analyses how their impact can be exacerbated or reduced by supra-national governmental funding processes.
3. It operationalizes data justice in these contexts by proposing a data justice plan that sits alongside other existing data governance requirements such as a data management plan or data protection impact assessment when utilizing citizen data to create new products and services.

In this paper, we explore two European-funded government-commercial R&I partnerships to develop data-driven innovations that address air pollution. We seek to locate citizens in these innovations, and understand how they are represented and affected by the use of data. Section 2 provides an overview of key concepts underpinning the use of citizen data by third parties to create new products and services. Section 3 is a detailed description of our methodology. Section 4 presents the results of our two case studies. In Section 5, we discuss our results. In our discussion, we propose the use of data justice plans, the development of which forms the backbone of future work.

2. Background

To set our theoretical context, we begin by exploring three concepts: data justice, smart cities, and citizen engagement. Section 2.1 outlines the possibility for data use to transgress certain rights. Section 2.2 sets

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

the context in which data are passed between city authorities and private enterprise for the purpose of creating new public services. Section 2.3 examines the routes by which citizens engage with governments in the fields of R&I.

2.1. Data justice

Taylor (2017) operationalizes data justice as three pillars of fairness in the way people are made visible in and by data; how they engage with the technology around data and how they are treated as a result of the production of digital data. Key to this is sharing in the benefits of data use. However, identifying the parts of this process in order to enact justice is challenging. Dencik and Sanchez-Mondero (2022) underline that the complexity of data flows has made it impossible to identify where any one piece of data came from or has moved to, let alone who is accountable for its proper and just use. The users of the data can be located anywhere, and the subjects can be unaware of the use of their data.

Mann (2018) provides a vivid example of this, noting that data extraction from African-based organizations is often justified on the basis that it is being used for humanitarian purposes, while it in fact enables foreign companies to stake a claim on entire data ecosystems for future exploitation. This appropriation of data is not limited to specific geographies, nor of particular activities—virtually, any part of the everyday life that can be captured and repackaged as data is a potential venue for "data colonialism" (Couldry and Mejias, 2019). On the flip side, those who are missing from data and thus invisible in datasets are equally disadvantaged (Milan and Treré, 2020).

Inequity in data processes therefore can perpetuate inequities in society, much as bias in datasets can enhance bias in society when amplified by algorithmic effects. There is a considerable body of work regarding "fairness" in data and algorithms, but this largely focuses on improving access to or use of data, preventing biases within the data, and other principles for more ethical algorithms (e.g., Garcia, 2016; Wilkinson et al., 2016). There is far less focus on the need for equity in data processes that prevent what Heeks and Shekhar (2019), term, "problematic distributive impacts," where the use of data further marginalizes already disadvantaged communities. As yet, it is not clear how these principles operationalize (Dencik, 2022) or whether they actually affect equity of access, representation, and participation (Dodds, 2020).

One approach to increased fairness in data use is that of data autonomy. Data autonomy refers to an individual's right to decide what happens to their data, as opposed to the right to simply keep it private (Fracassi and Magnusson, 2021). This includes the willing sharing of data. It is a core tenet of the "freedom and autonomy" human right theme of the proposed UN Habitat Digital Rights Governance Framework for Cities (Lähteenoja et al., 2022). Initiatives such as the MyData principles[1] or the City of Helsinki Data Strategy[2] have begun to acknowledge the human right of citizens to have autonomy over their data. However, assertion of such a right requires a workable framework for its enactment.

2.2. Smart cities

There are myriad definitions of what smart cities are and do (see Karvonen et al., 2018 for a brief overview) and what concepts and activities fall within its purview. At the most implemented end is the ubiquitous "smartcard" such as Oyster in the UK, OV-ChipKaart in the Netherlands or LisboaViva in Portugal. At the most conceptual end of the spectrum lies the (abandoned) Sidewalk Labs and Toronto City "Quayside" project, with autonomous cars, intelligent rubbish collection, smart air quality measurement, and heated streets. The common denominator for a smart city is that it uses information and communications technology (ICT) with data to increase efficiency and improve services for its citizens (Nam and Pardo, 2011).

[1] https://mydata.org/guiding-principles/.
[2] https://digi.hel.fi/english/helsinki-city-data-strategy/.

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

Essentially, "smart city technologies are enthusiastically seen as a solution to urban problems" (Borkowska and Osborne, 2018). Smart cities initiatives such as Open and Agile Smart Cities,[3] have catalyzed investment in infrastructure, enabling the monitoring of many aspects of city life: traffic, air quality, weather, parking, building repair, use of public space, energy consumption and more, creating tremendous amounts of data, which can then be used by private companies to create innovative solutions. Engagement with commercial organizations in this way often takes the form of digital innovation contests where a city or government agency presents an urban challenge and organizations compete to create pilot solutions (Mergei, 2017). In these contests, a municipal authority usually provides some financial and networking support, but the only core requirement of these solutions is access to data, which can be obtained by business in a number of ways. Almirall et al. (2014) note that such contests can sometimes be the venue of conflict between the partnership approach of the municipality, and the competitive approach of the business.

While there are many potential upsides of smart cities, there are conversely many criticisms. Whether smart cities are, in fact, drivers of social change is a matter of debate. Glasmeier and Christopherson (2015) find that smart city technologies are rarely capable of improving quality of life for citizens, addressing problems such as failing schools, or encouraging inclusivity. Hence, the smart city context itself leans toward the inequitable. The processes may lend themselves to numerous potential harms to citizens, including dataveillance, inference, obfuscation, identification, performative consent (Kitchin, 2016) and treating citizens as sensors (Dooley, 2021). Hollands (2015) argues that the focus on large corporations working with government institutions for "urban entrepreneurialism" leaves little room for citizens to engage in these processes. Borkowska and Osborne (2018) find that, in the context of the Glasgow Future City Demonstrator Initiative, funder priorities overruled citizen needs. More benefit for citizens could have accrued if there had been more communication with citizens about the planned smart solutions.

The "smarter" cities become, the more they rely on human involvement for successful implementations. If cities want to become smarter, they have to engage with the people who live in and are affected by their smart solutions (Gascó-Hernandez, 2018; Wachter, 2019). Engaging citizens in the development and improvement of new digital services increases their use, and reduces skepticism. This citizen engagement needs to involve a strong focus on data; but citizens are a notoriously challenging group to engage with data. This is for a number of reasons to do with skills, interest, awareness, incentives, and the sheer amorphous nature of the population (Worthy, 2012; Tinworth, 2016). Attempts to involve citizens in decisions about smart city implementations have shown that this often involves bringing citizens on board with already existing plans, rather than adapting the plans to citizens' needs (Breuer and Pierson, 2021). Even where bottom-up engagement happens, "smart" solutions are often trialed experimentally and at small scale. This subverts otherwise inclusive participation mechanisms, as only a select few, skilled citizens actually engage in processes where more formal engagement would lead to better representation (Webster and Lelux, 2018).

2.3. Citizen engagement

Citizen engagement is a traditional way to gain citizens' approval for projects in local politics. Many councils have established processes by which citizens can influence local development, especially, with regards to building and planning permissions, raise issues such as air pollution, or even distribute finances with participatory budgeting approaches. Citizens' rights to engagement in policy decisions are enshrined in the Lisbon Treaty of the EU and the Council of Europe, and strongly supported by international bodies such as the UN (European Commission, 2001; Rosenzweigova and Skoric, 2016). Similarly, the co-production of public services between municipal authorities and citizens has been a key international theme for over a decade (Bovaird, 2007; European Commission, 2018b. Lämmerhirt et al. (2018) describe how governments can actively involve citizens in campaigns to enable better policy-making.

[3] https://oascities.org/.

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

However, citizen engagement is often found lacking actual engagement with citizens' views and preferences, or power transfer to citizens. Instead, citizens are often "engaged" in processes that appear to give them influence, while actually manipulating them to rubber-stamp decisions by other stakeholders (Arnstein, 1969). Arnstein's *ladder of participation* is a useful tool to assess the validity of engagement based on the role citizens play, which are classified into the categories of nonparticipation, tokenism, and citizen power. Borkowska and Osborne (2018) also acknowledge that active citizen engagement is not always possible.

White (1996) also offers a classification of citizen engagement, based on the intentions of the engaged citizens and those who engage them, distinguishing nominal, instrumental, representative, and transformative engagement. She notes that not all engagement is meant to transfer power, and that there are different ways in which citizens can benefit from engagement, even if they do not make decisions. Thus, while some engagement may appear unjust in terms of direct benefit, there can be other, intangible benefits. If the benefits on both sides are sufficient to justify their respective activities, these can still be valid forms of engagement.

Citizen engagement in innovation can take many forms, such as bottom-up development of new ideas by citizens, citizens participating in market research, crowdsourcing, or co-production between citizens and companies. Involving citizens in innovation processes helps to ensure that those innovations actually meet citizens' needs, increases diversity and legitimacy, and supports "buy in" from citizens for the results (Simon and Davies, 2013). It can also reduce the cost of innovation (Nesti, 2018).

One form of engagement specifically suited to R&I is citizen science: The involvement of the general public in scientific research activities (European Commission, 2014). However, even in this focused engagement, citizens do not always benefit from projects in which they participate. In their study of the practical application of data justice in citizen science projects, Christine and Thinyane (2021) found that data processes in some projects were inherently unjust, as the benefits of the data were limited to the hosts of the projects. One such example is the use of citizen labour to create digital data for the UK Met Office's Old Weather project,[4] which is subsequently used for financial gain, without returning expected benefits to the citizens who generate that data.

3. Methodology

In order to understand data justice in data-driven innovation projects, we have carried out a qualitative case study analysis. The case study addresses two smart city-based data-driven air quality innovation projects, within two separate funding programmes. The same commercial partner was involved in both projects. Focusing on one company in two different funding contexts allows us to identify differences in project implementation that can provide insight into how benefits accrue to citizens. The case study approach enables a detailed description of a contemporary phenomenon within its context (Yin, 1994) and is particularly suitable for examining whether theory is applied in practice (Gerring, 2004). Yin (2003) argues that findings cannot be generalized to populations, but are useful for generalizing theories.

3.1. Case study selection

Air pollution is an area of data-driven innovation that is a major area of focus for cities and public innovation funders, including UK R&I, and European R&I funding. The problem of air pollution contributes not only to global warming, but also increases individual health risks, especially among vulnerable groups (European Environment Agency [EEA], 2020). In December 2020, air pollution was for the first time recorded as a cause of death in the UK (Laville, 2020).

Data-driven air quality projects make good subjects for the investigation of data justice for citizens. It is almost impossible for citizens to avoid being the subject or object of air quality smart city initiatives. By simply existing, citizens affect air quality. Citizens cannot opt out of data gathering if they live or work in the

[4] https://www.oldweather.org/.

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

studied area, and cannot avoid what Couldry and Mejias (2019) call, "the flow of everyday life" being "reconfigured and re-presented in a form that enables its capture as data." Therefore, they have no natural agency in how they engage with this data and its use for innovation (Taylor, 2017). Further, it is difficult to see how they can opt out of any air quality initiatives implemented as a result of data-driven innovation projects.

Air quality impacts many aspects of life. Primary amongst these is health. Poor air quality is a top cause of non-communicable disease. Improving air quality is a key global health challenge. It can be affected by dimensions such as time, weather, traffic, events, or street layout.

Air quality is also unequal. A study by the UK Department of Fisheries and Rural Affairs (Pye et al., 2006) found correlations between certain pollutant particulates and areas of urban deprivation. Within cities, often the most densely populated, deprived areas also suffer the most air pollution. This makes air quality a politically sensitive issue. Authorities can be reluctant to publish detailed air quality data openly, unless accompanied by a strategy for its improvement.

The projects we have chosen to review are both part of EU funded R&I strategies for smart specialization (RIS3) programmes. Consequently, the projects have clear terms of engagement and are part of a much wider picture of funding and innovation.

While citizen engagement should be active (Arnstein, 1969), the reality is that it is often passive, and benefits indirect, as described by White (1996) Therefore, analyzing cases with explicitly high levels of citizen engagement, although vital, is not fairly representative of what actually takes place in many of these programs and projects. By analyzing projects that are not attempting at the outset to attain high levels of citizen engagement, it is easier to assess where the issues of data justice may lie, while more accurately reflecting reality. Further, most projects of this type (in digital innovation contests) are not "successful," that is, they do not lead to full implementations. Therefore, it is important to look at how citizens may lose out where projects do not lead to the creation of implemented value. While case studies should not be over relied on as being representative of anything but their own situation, this context makes it productive to extrapolate learning and insight.

3.2. Analysis

The analysis was based on 11 relevant documents from both programmes, as well as four interviews with key stakeholders (two of these had more than one participant). A complete list of all documents and interviews, with links to public documents where these are available, is provided in Appendix A. Documents were collected from project files to which the authors had access. They were selected based on relevance to the funding distribution, and the specific cases. The coding was performed inductively, in that the codes arose from our analysis. All documents were thematically coded for the *data* in question, the *objectives* set out by the respective stakeholders, the *stakeholders* considered, and the *engagement* of citizens. This gave us an overview of the circumstances under which the projects were set up, and the priorities at the outset. We further supplemented this document data with four semi-structured interviews with relevant stakeholders from the projects. All interviews were conducted online or by phone between January and March 2020, and were between 25 and 40 min long. They focused on the project plans and implementation, especially on the role of citizens in them; stakeholders' views on what engagement was, or should have been, conducted; and where applicable, how and why they fell short of those expectations. All interviews were transcribed and subsequently analyzed in the same framework developed through the document analysis.

In the next section, we present the complete case studies of the two European funded programmes, the air quality challenges, the SME involved, and the activities in the respective programmes.

4. Case studies

Our case studies focus on two smart city projects implemented in the EU Horizon 2020 programme Data Pitch (DP) and the Interreg2Seas program Smart Cities Innovation Framework Implementation (SCIFI). These projects were partnerships between municipalities/governmental organizations and a small to medium company, HOP Ubiquitous (HOP).

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

4.1. Data Pitch

Data Pitch (DP)[5] was a Horizon 2020 programme that ran from 2017 to 2019. It was an open innovation program bringing together corporate and public sector data-holding organizations with start-ups and SMEs, specializing in innovating with data. It aimed specifically to, "create value from sharing data." (DP GA) From its outset, it aimed to develop the European innovation ecosystem with data-driven business-to-business collaborations. Part of the programme consisted of challenges that offered access to data by large organizations for the purpose of developing new solutions. Across 3 years, DP managed the sharing of data from 13 large organizations to 47 SMEs.

4.2. Smart Cities Innovation Framework Implementation

Smart Cities Innovation Framework Implementation (SCIFI)[6] was an Interreg2Seas funded programme, running from 2018 to 2021. It addressed RIS3, specifically the technological and social innovation program priority objective SO1.1: *to improve the framework conditions for the delivery of innovation.* The aim of the program was to increase the capacities of cities in the region for technology transfer, stimulate innovative sectors of the regional economy, and increase quality of life in the cities. It also aimed at increasing the capacities of innovative companies to engage in international activities. Like Data Pitch, SCIFI used the challenge format to engage businesses to use city and other data to address public sector problems.

4.3. HOP Ubiquitous

HOP is a Spanish company with around 25 employees specializing in Internet of Things (IoT) technologies, such as sensors and related analysis software. Their main market is environmental monitoring solutions.

4.4. EU research and innovation strategy and the quadruple helix

Research and innovation strategies for smart specialization (RIS3) has been a key concept in EU policy since the nineties. It requires regions across Europe to engage in an entrepreneurial discovery process that supports the implementation of R&I. The EC's guidance notes for RIS3 recognize the need for a participatory governance of science and technology, and argue the benefits of the quadruple helix approach (Deakin et al., 2018). Helices framework theory is a way of conceptualizing the knowledge flows and processes that result in and institutionalize economic growth via innovation and entrepreneurship (Etzkowitz and Zhou, 2017). The quadruple helix—where the actors are government, academia, business, and civil society—has been part of the EU's R&I strategy since 2012 (European Committee of the Regions et al., 2016).

The programmes that constitute our case studies are explicitly quadruple helix focused. The role of businesses and developers in EU open data policy is to exploit the data for economic value and job creation, via engaging in open innovation with data (Lassinantti, 2014). However, the positing of citizens as both users driving innovation and participants in the creative process (Yawson, 2021) opens the door to open innovation that involves a wider group of stakeholders. The idea of co-production or co-creation has been present over the last few decades and now appears in the documentation for Horizon Europe (European Commission, 2018a).

4.5. Cascading funding

EU cascading funding is a technique to redistribute funding to organizations who would be unable to apply for it at its source. Recipients of such grants are commonly start-ups, SMEs or non-governmental

[5] https://datapitch.eu/

[6] www.smartcityinnovation.eu

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

organisations (NGOs),completing a specific set of activities, who become subject to the same contractual framework and financial obligations as recipients of the original grant. Cascading funding is part of the implementation of the EU R&I strategy, and the awarding programmes need to ensure that funded projects adhere to those rules, and the wider objectives of the funders and their specific grant. Ideally, cascading funding processes translate the original goals and objectives of the source funders, through the intermediary programmes running the call, to the organizations implementing the innovation project with their grants. The cascading funding process in SCIFI and Data Pitch is outlined in Appendix B.

4.6. The data challenges

The specific data-providing partner relevant to the air quality challenge in DP was the UK's Met Office (MO), a government trading company and recognized global leader in weather and climate science, technology, and services. The MO challenge was quite wide, and sought solutions and opportunities in using weather data to drive innovative techniques and support communities, for example, in retail (supporting the value chain) or advisory services, for example, *When is a good time to cycle based on a route planner?* The data provided by the MO were aggregated historical pollen data collected by the MO and not as yet available openly. HOP was successful with an application to work on this data with the intention to improve local governments' and civil servants' ability to make decisions about air quality policy and urban planning. Their proposed solution was intended to increase the understanding of these factors, and enable more accurate predictions for mobility and tourism.

The air quality challenge in SCIFI was presented by the Flemish cities of Bruges and Mechelen, who, in addition to general transport and industry pollutants, have built environment issues where pollutants become trapped by air recirculation between narrow buildings (urban canyons). The cities sought data on the causes of pollution, and a way for them to formulate this evidence into impactful policies for improved air quality for citizens. HOP was selected by both Bruges and Mechelen to address their air quality challenge. They were the only SME working in Bruges, although Mechelen also selected another company, Nazka, to address the same challenge. The cities' goal was to use and publish open data. At the beginning of the project, the only data available was from the EU Copernicus data hub.[7] To collect local and specific data, HOP aimed to place air quality sensors in seven locations in different circumstances, to find differences in concentrations with a lot of traffic, green spaces in the city, and rural areas close to the city. These sensors were intended to track multiple gases, particulate matter, temperature, humidity and noise levels, alongside a weather station, to find correlations with (high) temperature, rainfall or wind. The intention was for those sensors to continuously send data to a platform every minute. The data would then be published as real time open data, and presented in dashboards for consultation by interested city departments and citizens, and also in a machine-readable format.

The DP challenge was completed, although the pilot did not then progress to full implementation, due in part to a change of personnel and goals within the MO. The SCIFI pilot was beset with multiple problems on the ground, including Spanish sensors that did not appreciate Belgian temperatures, and issues regarding compromised locations and poor connectivity. This is not uncommon; "real world messiness" such as dust, wind or even a fingerprint, can interfere with the digital perfection of data (Sambasivan et al., 2021). Mechelen now plans to fund citizen scientist-developed low-cost sensors, trading lower accuracy for reduced implementation cost. Bruges has started work with another commercial partner, and is exploring options to work with local citizen scientists in the future.

5. Results

5.1. Data Pitch

5.1.1. Objectives

The DP challenge for which the MO provided data did not explicitly address citizens' needs, rather stating that the challenge should be, *"meaningful and relevant to people, industry and government"* (DP GA).

[7] https://www.copernicus.eu/en

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

The project between the MO and HOP, however, focused on *"a real problem for citizens: Air pollen for people with allergies. How to create actions that will improve air quality via a citizen subscription system that will enable them to change their behaviour, educate them on whether to take public transport, close their windows, take actions to reduce their exposure when there is high pollen"* (HOP Interview). HOP also had a second market driven motive for participating, in that they wanted to use the pilots to enable a pivot from, *"a company focused on just selling IoT devices, to one that can also provide data-driven solutions."* Their hardware, then, would be enhanced not just by data but also by *"inclusion of algorithms for adding value to data (e.g., modelling, forecasting, transforming)"* (HOP Interview). The MO was also concerned about monetization of their data, as the institution is expected by the government to generate income. One way this was envisaged was through projects like HOP's, which could, if successful, form a long-term income stream. The objectives for both were product-focused with a goal of business-to-business and business-to-consumer commercialization. Thus, citizen benefit was not a primary objective.

5.1.2. Data

The focus of HOP's project in Data Pitch was on the creation of tools that could be used by local authorities, to make the data that the MO generated accessible, and enable better decision-making. However, instead of measurements, the MO data consisted of *forecasts* generated on the basis of measurements. The actual pollen count data were generated with the help of volunteers, as the DP advisor to the project explained: *"very old school with sticky tape that collects pollen, then has to be physically counted."* This pollen count data are used to generate the forecasts that were then shared with HOP in DP. The data HOP typically deals with are mostly measurements of air pollution such as pollen, ozone, and nitrogen oxide (NOx), which are all potential health threats. So, while NOx had a strong negative connotation as a pollutant (a common by-product of combustion), data about pollen were perceived as *"beautiful and natural*" (Interview HOP). It was therefore less politically sensitive because of this perception as natural, and theoretically had less potential to upset citizens than pollution data. However, the data shared with HOP was not raw, and therefore less usable. *"It was only when we got in front of the data that the engineers said they couldn't do much with the data,"* (Interview DP mentor). HOP worked with these forecasts and supplemented them with data generated by their own IoT sensors, to produce advice around air pollution and the intensity of specific pollen in specific areas. This combined data were then used to build dashboards for local authorities, to help them comply with legal requirements and reporting duties. HOP also planned to create subscription apps that could provide advice for citizens around the intensity of specific pollen in their area, albeit not within the DP program.

5.1.3. Stakeholders

The funding application for the DP program addressed the quadruple helix approach, outlining an intention to engage citizens as well as industry, research, and government. Citizens were identified as beneficiaries or interested parties, but not primary users of the products companies DP would develop. DP specified that it would benefit citizens through *"added value products and services to support them in their private, social, and professional life,"* and that *"citizens will become more aware of some of the most pressing societal and economical challenges"* (DP GA). In their work plan for DP, HOP identified both their customers (public authorities) and citizens as stakeholders of the project, who would "*be able to access and understand real aspects of their environment, […] and checking [the data] via easy-to-understand indicators"* (DP Work Plan).

5.1.4. Citizen engagement

Citizen engagement with open data or civic technologies was a voluntary activity within DP funded projects. Since it was not a focus of DP itself, it was not supported by its funding or framework. Information about the funding program and funded projects was available to citizens via social media, or on the media in general, but citizens were not engaged in the process itself. They would however be affected by its outcomes, if implemented.

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

While the data shared by the MO were collected by volunteers who collect pollen samples, the data were then published and valorized by the organization. To enable citizens to engage with this data, data products need to, *"provide a strong value proposition, because [citizens] are not forced to use the data for anything"* (HOP Interview). In practice, this meant that the project with the MO needed to make a strong case as to why citizens should engage—as customers. To this end, engaging urban planners first was perceived as crucial, with the intention to expand to citizens once the former were content.

More engagement might have been possible, and HOP did envisage creating a citizen-facing application, but "*because it was outside the scope of DP it wasn't prioritised or tested with citizens"* (Interview DP mentor). Citizens were never intended to engage with HOP's solution directly: *"it's beneficial for citizens. But [citizens] don't interact with the tool, they interact with the conclusions of the tool"* (HOP Interview).

5.2. SCIFI

5.2.1. Objectives

Bruges' objective in the SCIFI project was to *"have a view of air quality as of now and its use for directing policies."* The city wanted to develop their climate action plan on the basis of this data, and "*make a digital twin and have predictive models for how we can make low pollution zones or shut down streets if the pollution is too high, etc."* (Bruges Interview). This was partly driven by successful implementation of electrification of transport policies, which had been introduced in the wider Flanders area after the collection and analysis of air quality data.

Like Bruges, the objective for Mechelen was policy-related, in that the city already had a traffic reduced area, and *"wanted to see as well if we see an impact of that car free zoning during the day, and if we could see a difference between the air quality of a very rural area of the city and a place with high traffic density."* However, there was also at least a partial objective to make the data accessible to citizens: *"At first we didn't have the objective to really publish a lot of data with a lot of publicity towards the citizens. However we did specify within our pilot contract that the air quality data produced by the sensors should be available as open data."* However, ultimately the cost of sensors was a barrier to achieving the objectives (Mechelen Interview).

HOP's objective in SCIFI was to *"encourage urban planners etc. to use data in daily activities, like climate change mitigation action or other environmental impact activities,"* which would also improve *"the health of our citizens, who may suffer from lung, heart and other related problems"* (SCIFI HOP public profile). Despite this objective being aligned with that of SCIFI, HOP felt that *"there is not a strong market for urban planners at the moment, [but] we feel there is the gap*" (HOP Interview). *The objectives for both HOP and the municipalities were focused on policy and planning levels, although for HOP this was at a step removed, as they would be simply be supplying data to the planners rather than using it to make decisions. HOP's objective was to create a successful product aimed at planners as a purchasing base, who in turn had citizen benefit as a motivation.*

5.2.2. Data

The datasets utilized for the SCIFI project were obtained from several sources: data provided by Bruges, and directly generated from HOP monitoring activities through their IoT sensors. It was integrated in a shared platform, ensuring the data were standardized and interoperable.

While the cities in the SCIFI project were keen on obtaining this data, they were hesitant about sharing it with their citizens, especially so before they had a chance themselves to understand and process it, and predict what the reactions of citizens to pollution levels would be. In short, the goal was to present the problem *along with a solution*, rather than letting citizens explore the problem on their own: *"In theory all data should be available for citizens but we and the inhabitants of Mechelen don't have a lot of experience with real time open data"* (Interview Mechelen). There were concerns within the city administration that data, once released, could be interpreted by citizens in different ways, and thus cause conflict between citizens and the administration: *"We wanted to make sure we first get a feel of the data internally in the*

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

organisation before we involve citizens with it" (Interview Mechelen). This was also the perception of the project team at HOP: *"When you empower citizens [by publishing data] you do not want to offer them tools to fight you with,"* (Interview HOP).

5.2.3. Stakeholders

Like Data Pitch, the funding applications for SCIFI also addressed the quadruple helix approach. SCIFI specified cities, and by extension citizen needs, as the drivers behind the programs' challenges. Citizens would benefit through *"the actual solutions developed, which improve urban liveability,"* and *"the resulting innovative products and services, which aim to improve public services."* According to the quadruple helix, they also specified businesses and *"growth opportunities for the private sector,"* as well as research institutes as key stakeholders (SCIFI GA).

Bruges' challenge identified affected parties as *"all citizens"*; interested stakeholders were identified again, as, *"individual citizens,"* but also, "*'green' political parties."* This was picked up by HOP in their work plan, where they planned to *"know the specific needs of the city with the aim of developing useful tools for decision making […] to improve the lives of citizens,"* and *"confirm that the proposed solution will solve real needs of Bruges' population"* (SCIFI Work Plan). Despite being nominally focused on citizens, the SCIFI projects primarily targeted urban planners as users, with citizens as cascaded beneficiaries. The indirect route to citizen benefit, via their services provided to urban planners, was how HOP perceived how the products they developed would be used by the cities: *"They are not offering services to citizens, they are using them as tools for urban decision making"* (HOP Interview). This focus on authorities became a problem where that audience—the urban planners—were not properly engaged. One of the enterprises noted that implementation was negatively affected because of *"limited staff availability at the city environmental department"* (Nazka Public Profile).

5.2.4. Citizen engagement

Throughout all the SCIFI projects, policy makers defined citizen benefit generally as "everything we do is for citizens anyway," with very little active or planned and executed engagement activities. There was some inward citizen engagement, especially at the challenge definition stage. Bruges surveyed citizens to find out about their priorities, and *"air quality came out of that as important,"* several times. *"The result of polls we did with citizens […] showed that air quality was the most important issue. This is why we picked it as a pilot project"* (Interview Bruges). Mechelen, while not having the benefit of this specific engagement, also identified air quality and climate change as of central concern to citizens in the wider region, as well as a special interest for a local activist group and the city council. Even where citizen engagement was planned, such as in HOP's work plan for the Bruges project, which anticipated a number of workshops that citizens could attend, this element was abandoned due to time constraints and the importance of engaging with other groups, such as urban planners.

While the project was underway, citizen engagement by the policy makers was largely confined to communication, with the city council communicating updates to citizens. There was little to communicate about, and much of it the cities perceived as too premature to discuss: "*How is that policy making communicated to citizens? At the moment it is not. It is too sensitive to already communicate to citizens"* (Bruges Interview).

HOP pointed out that this is a behavior they encountered frequently in their work with policy makers. Cities are hesitant to publish data that may potentially cast the city, or locales therein, in a negative light.

6. Discussion

The combination of intended engagement of civil society in innovation with government and private enterprise through the quadruple helix, along with the ambitious goals of smart cities, and the possibilities and limitations of citizen engagement, form the context for our analysis of data justice.

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

Both air quality challenges were set by governmental organizations. The smart city composition of SCIFI, with city authorities as the data providing partner, meant that citizens were framed as more immediate beneficiaries of the results of the pilots. However, in reality, they would be beneficiaries of improved outcomes of good strategy created by a successful pilot project—if all of these worked as planned.

DP, in following the quadruple helix anticipated an eventual impact on citizens, but this was only ever envisaged as a cascade effect, not direct engagement. However, the pollen subscription service imagined by HOP for the Met Office would have had citizens as the direct beneficiaries—if it had been implemented. But since the program was focused on innovation and solution development, and citizen benefit did not form part of the success criteria the project was ultimately assessed against, this element was the first to go. This was to be expected, not least because the involved organizations, by their very nature, followed different goals: While public administrations or organizations act, in principle, with intentions to improve services and public value, private companies' goal is to enhance their competitive advantage (Kankanhalli et al., 2017). Either way, both programs would have benefitted from better integration between the collaborative approach of the hosts, and the competitive approach of the SME (Almirall et al., 2014).

The gap between these two approaches allowed data *appropriation* to take place, with the data of citizens being "repackaged and reused" elsewhere as described by Couldry and Mejias (2019). In DP, the MO and HOP benefited from knowledge gained, even without full implementation. In SCIFI, HOP gained access to cities' data and the learning from that acquisition, both the data itself and the infrastructure around it. This experience supported them to achieve their business goal of pivoting from a hardware only company to having a software focus. This scenario resembles Mann's (2018) conceptualization of data colonialism.

A key differentiator between SCIFI and DP was that DP deliberately sought to facilitate the sharing of non-open data, whereas SCIFI sought (not always successfully) to utilize open data. Open data play an important role in SCIFI in ensuring citizens are actively engaged as well as being passive beneficiaries. Were the SCIFI air quality challenge to have been more successful, this might have changed the outcome vis-a-vis citizens. For instance, if data have been successfully published, then it could have been used by or shared to MeetMeeMechelen, the citizen observatory platform. It is difficult to speculate accurately on how much use the citizens would have made of the proposed dashboards given issues of the skill and knowledge required to engage with data (Kassen, 2021). However, enabling citizen access to the data would have been one potential way to reduce the *imbalance of distributive impacts.* Citizens might still benefit from this data use in the future, but would have to wait for this learning to be implemented through another, more successful project. Even then, the international dimension of the SCIFI program means that the learning could have arisen from data from Belgian citizens, and the value would be derived by Spanish citizens, or indeed in any locale where HOP decides to exploit learning based on data acquired during SCIFI.

We found that there appeared to be a misalignment in the air quality focus between the stakeholders, including the funders. Essentially, citizens were less engaged than perhaps they should be in something that is framed as being specifically for them. In particular, there was no mechanism for contributing to the discussion of how data about them should be used. Throughout, the highest level of engagement was the survey preceding the challenge definition in Bruges, which would qualify as *Consultation*, along with different forms of *Information* (Arnstein, 1969) in both programmes. However, it is worth noting that even the workshops HOP had anticipated but did not implement would not have reached higher than this. Neither of the programmes had planned formal engagement in decision-making regarding the data, not least because there were no decisions to be made during the project implementations; citizens were meant to benefit through the *results* of the projects, not engage in their selection or implementation. Citizens were informed at best, and at worst, completely unaware of what happened with data that affected them, both due to their lack of agency in the projects, and lack of representation by the host organizations. Essentially, Taylor's second pillar of data justice, *engagement with technology*, was not available to citizens in these projects.

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

Since citizens were the most silent or passive of the stakeholders, the anticipated benefit for them became diluted over time, and across the stages of the funding; from funder, to programmes, to funded innovators, to project implementation. Furthermore, there was a distinct power imbalance between the citizens and cities and institutions who were privileged to have information that the citizens could not access. Processes to ensure *data autonomy* could potentially have helped redress this imbalance.

7. Implications for data justice

Throughout these projects, there were a number of formal governance mechanisms implemented to ensure financial and legal compliance and risk minimization. At the program level, these took the form of project deliverables which reported on the projects to the funders. At the project level, these took the form of data protection impact assessments, which aimed to ensure compliance with data protection law, and data inventories or registers, which listed all datasets used, whether personal or not, in order to ensure clarity of ownership and reduce the possibility of triangulation and reidentification. Although citizen benefit was a stated expected impact, ensuring fairness in citizen data use was not explicit in the projects. This meant that there was no requirement for a way for either the data sharing municipalities or HOP to assess or audit the expected benefits for citizens. Consequently, in a situation with many competing calls on the projects' time and money, there was no focus on this. Ensuring that citizen benefit is optimized and evidenced requires two things: firstly, mandating such benefit as a project output, and secondly, providing methods and guidance for measuring and demonstrating that benefit.

Far fewer initiatives of this type succeed than fail, so it is vital that how citizens can benefit throughout the process is taken into account, rather than just on completion of a successful project. While we focus on cities, as they are the context in which our case studies are based, the issues surrounding funding and citizen engagement occur at other levels. For example, while the HOP sensors were situated in one location, the MO data were collected nationwide. The data justice questions cannot be limited only to where the data are used—it must account for where it was collected as well. We propose that this rigorous monitoring take the form of a "*data justice plan.*" In the same way that data management plans have become part of the key deliverables of R&I projects, or that data protection impact assessments have become a key tool of managing data privacy within an organization, *data justice plans* would be a part of data governance structures. But rather than explaining how data are processed, protected, or consent is managed, it enshrines how data are used to generate value for its subjects or contributors, throughout its lifetime. It is envisaged that such data justice plans could be a valuable contribution to the tools available to city governments aiming to achieve the proposed Digital Rights Governance Framework for Cities, an assembly of foundations, structures, and tools.

Such a template would have allowed the municipalities to identify and specify how they wished value to accrue to their citizens—the data subjects—should the projects fail to achieve a full implementation. In terms of the open innovation context of these projects, this approach could also offer some upside to municipalities in this situation. Such municipalities often invest a great deal of time and effort in preparing the data for open innovation, and data justice not only considers the rights of the data subjects, but can also be of benefit to the public organization managing that data. Participatory data stewardship can support some of this work (Ada Lovelace Institute, 2021).

8. Conclusion, limitations, and future work

Fundamentally, many data-driven innovation programmes are urban technology projects that can come under the umbrella of smart cities. To an extent, the barriers to citizen engagement in these projects reflect those of the smart city as outlined by Coletta et al. (2018). R&I funders need to increase their focus on participation, to enable data justice for citizens in both the design and the implementation of these processes; particularly, where it is entirely reliant on the market.

In terms of city-led projects, the nature of the project is that the government must speak for the citizens. Therefore, ensuring engagement happens should be a key and non-negotiable part of project design, rather than additional or intermediable. Cascading funding can play a role here by preventing the "leakiness" of

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

the current model through (1) rigorous monitoring of not just project delivery, but also data justice, (2) ensuring funding is available for citizen engagement within these projects; and (3) making funds available to social entrepreneurs or advocacy groups (Calzada, 2020).

Citizen science is an even more promising approach to including citizens in innovation and ensuring they benefit from it. While data quality in citizen-led air pollution projects may not reach the same standard of data quality as professionally organized measurements (Mahajan et al, 2020), these citizen data are still sufficient to raise issues onto the political agenda, and thus enhance political engagement (Gabrys et al., 2016), contribute to sustainable development goals (SDGs; Fraisl et al., 2020), and ultimately achieve real benefits for citizens (Schaefer et al, 2020).

8.1. Limitations

A case study is by its very nature limited to individual cases. It is illustrative of only the example under exploration. We do not attempt to say every project would be like this, only to show how under certain circumstances, projects can be like this. We still believe that the insight generated from our comparison highlights some inherent weaknesses in data innovation projects, along with recommendations for their mitigation.

8.2. Future work

We plan to test and further investigate the engagement of citizens in program challenges; explore the use of citizen science data for innovation; and develop the idea of data justice plans, by developing a template, and a workable integration into data management best practice. These aspects will be implemented in one of the new Horizon Europe grants: *Impetus* builds on the cascading funding and innovation acceleration models of DP, SCIFI, and ACTION,[8] to support citizen science project implementations across Europe.[9]

Acknowledgments. We thank our candid interviewees at HOP, Stad Bruges, Stad Mechelen, and the Open Data Institute for their time and insights.

Data availability statement. All data sources are listed in the appendix, including links where they are publicly available. Due to the competitive nature of the programs, application data and interviews are confidential and cannot be made public.

Funding statement. This work was supported by Data Pitch (Horizon 2020, Grant No. 732506); Interreg2Seas; ACTION (Horizon 2020, Grant No. 824603); and Impetus (Horizon Widera, Grant No. 101058677). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interest. All authors declare that they have no competing interests.

Author contributions. G.T.: Conceptualization, data curation, formal analysis, investigation, methodology, project administration, visualization, and writing (draft, review, and editing); J.W.: Conceptualization, investigation, and writing (draft, review, and editing); E.S. and L.C.: Funding acquisition and supervision.

References

Ada Lovelace Institute (2021) Participatory Data Stewardship. Available at https://www.adalovelaceinstitute.org/wp-content/uploads/2021/09/Participatory-data-stewardship_Final-report.pdf

Almirall E, **Lee M and Majchrzak A** (2014) Open innovation requires integrated competition-community ecosystems: Lessons learned from civic open innovation. *Business Horizons 57*(3), 391–400. https://doi.org/10.1016/J.BUSHOR.2013.12.009

Arnstein SR . (1969) A ladder of citizen participation. *Journal of the American Institute of Planners 35*(4), 216–224. https://doi.org/10.1080/01944366908977225

Borkowska K and Osborne M (2018) Locating the fourth helix: Rethinking the role of civil society in developing smart learning cities. *International Review of Education 64*, 355–372. https://doi.org/10.1007/s11159-018-9723-0

Bovaird T (2007) Beyond engagement and participation – User and community co-production of public services. *Public Administration Review 67*(5), 846–860.

[8] actionproject.eu.

[9] impetus4cs.eu.

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

Breuer J and Pierson J (2021) The right to the city and data protection for developing citizen-centric digital cities. *Information, Communication & Society 24*(6), 797–812. https://doi.org/10.1080/1369118X.2021.1909095

Calzada I (2020) *Smart City Citizenship*. Amsterdam: Elsevier.

Christiane DI and Thinyane M (2021) Citizen science as a data-based practice: A consideration of data justice. *Patterns 2*(4), 100224. https://doi.org/10.1016/j.patter.2021.100224

Coletta C, **Heaphy L and Kitchin R** (2018) Actually-existing smart Dublin: Exploring Smart City development in history and context. In Karvonen A, Cugurullo F and Caprotti F (eds), *Inside Smart Cities: Place, Politics and Urban Innovation*. New York, NY: Routledge, pp. 85–101.

Couldry N and Mejias UA (2019) Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media 20*(4), 336–349. https://doi.org/10.1177/1527476418796632

Deakin M, **Mora L and Reid A** (2018) The research and innovation of smart specialisation strategies: The transition from the triple to quadruple helix. In *27th International Scientific Conference on Economic and Social Development*, Rome, 1–2 March 2018.

Dencik L, **Hintz A**, **Redden J and Treré E** (2019) Exploring data justice: Conceptions. *Applications and Directions, Information, Communication & Society 22*(7), 873–881. https://doi.org/10.1080/1369118X.2019.1606268

Dencik L and Sanchez-Monedero J (2022) Data justice. *Internet Policy Review 11*(1), 1615. https://doi.org/10.14763/2022.1.1615

Dodds L (2020) FAIR, fairer, fairest? *Lost Boy*, 30 July. Available at https://blog.ldodds.com/2020/07/30/fair-fairer-fairest/ (accessed 17 August 2022).

Dooley K (2021) Direct passive participation: Aiming for accuracy and citizen safety in the era of big data and the Smart City. *Smart Cities 4*(1), 336–348. https://doi.org/10.3390/smartcities4010020

Etzkowitz H and Zhou C (2017) *The Triple Helix University–Industry–Government Innovation and Entrepreneurship*. Oxford: Routledge.

European Commission (2001) *European Governance: A White Paper, COM(2001) 428*. Brussels: European Commission. Available at https://ec.europa.eu/commission/presscorner/api/files/document/print/en/doc_01_10/DOC_01_10_EN.pdf

European Commission (2014) *Green Paper on Citizen Science*. Citizen Science for Europe. Available at https://ec.europa.eu/digital-single-market/en/news/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research

European Commission (2018a) *Proposal for a Regulation of the European Parliament and of the Council Establishing Horizon Europe – The Framework Programme for Research and Innovation, Laying Down its Rules for Participation and Dissemination, COM(2018) 435 Final (7.6.2018)*. Brussels: European Commission. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018PC0435

European Commission (2018b) *Co-Production – Enhancing the Role of Citizens in Governance and Service Delivery*. Technical Dossier. Available at https://ec.europa.eu/esf/transnationality/content/co-production-dossier-explains-co-trends

European Commission (2021) Building a Data Economy. Available at https://digital-strategy.ec.europa.eu/en/library/building-data-economy-brochure

European Committee of the Regions, **Volpe M**, **Friedl J**, **Cavallini S and Soldi R** (2016) Using the Quadruple Helix Approach to Accelerate the Transfer of Research and Innovation Results to Regional Growth. Publications Office of the European Union. Available at https://data.europa.eu/doi/10.2863/40804

European Environment Agency (EEA) (2020) *Healthy Environment, Healthy Lives: How the Environment Influences Health and Well Being in Europe*. Publications Office. Available at https://data.europa.eu/doi/10.2800/53670

Fracassi C and Magnusson W (2021) Data autonomy. *Vanderbilt Law Review 74*(2), 327.

Fraisl D, **Campbell J**, **See L**, **Wehn U**, **Wardlaw J**, **Gold M**, **Moorthy I**, **Arias R**, **Piera J**, **Oliver JL**, **Masó J**, **Penker M and Fritz S** (2020) Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science 15*, 1735–1751. https://doi.org/10.1007/s11625-020-00833-7

Gabrys J, **Pritchard H and Barratt B** (2016) Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data & Society 3*(2), 2053951716679677. https://doi.org/10.1177/2053951716679677

Garcia M (2016) Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal 33*(4), 111–117.

Gascó-Hernandez M (2018) Building a smart city: Lessons from Barcelona. *Communications of the ACM 61*(4), 50–57. https://doi.org/10.1145/3117800.

Gerring J (2004) What is a case study and what is it good for? *The American Political Science Review 98*(2), 341–354.

Glasmeier A and Christopherson S (2015) Thinking about smart cities. *Cambridge Journal of Regions, Economy and Society 8* (1), 3–12. https://doi.org/10.1093/cjres/rsu034

Heeks R and Shekhar S (2019) Datafication, development and marginalised urban communities: An applied data justice framework. *Information, Communication & Society 22*(7), 992–1011. https://doi.org/10.1080/1369118X.2019.1599039

Hollands RG (2015) Critical interventions into the corporate smart city. *Cambridge Journal of Regions, Economy and Society 8*(1), 61–77. https://doi.org/10.1093/cjres/rsu011

Kankanhalli A, **Zuiderwijk A and Tayi GK** (2017) Open innovation in the public sector: A research agenda. *Government Information Quarterly 34*(1), 84–89. https://doi.org/10.1016/j.giq.2016.12.002

Karvonen A, **Cugurullo F and Caprotti F** (2018) Introduction: Situating smart cities. In Karvonen A, Cugurullo, F and Caprotti, F eds., *Inside smart cities: Place, politics and urban innovation*. New York, NY: Routledge, pp 1–12.

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

Kassen M (2021) Understanding motivations of citizens to reuse open data: Open government data as a philanthropic movement. *Innovations 23*(1), 44–70. https://doi.org/10.1080/14479338.2020.1738940

Kitchin R (2016) The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374*(2083), 20160115. https://doi.org/10.1098/rsta.2016.0115

Lähteenoja V, **Dominguez H**, **Facchina M**, **Westerberg P**, **Nkuidje L**, **Goodwin H**, **Shale Sagar A**, **Birchall C**, **Blok J**, **van Eemeren A**, **Ramírez Chico G**, **Pérez Batlle M**, **Boet Serrano P**, **Perrino M**, **Portier F and Manso J** (2022) *Digital Rights Governance Framework for Cities (Draft)*. Cities for Digital Rights Coalition. Available at https://citiesfordigitalrights.org/sites/default/files/DIGITAL%20RIGHTS%20FRAMEWORK_CONCEPT%20FOR%20FEEDBACK.pdf

Lämmerhirt D, **Gray J**, **Venturini T and Meunier A** (2018) Advancing sustainability together? Citizen-generated data and the sustainable development goals. SSRN Scholarly Paper ID 3320467. Social Science Research Network. https://doi.org/10.2139/ssrn.3320467

Lassinantti J (2014) Public Sector Open Data: Shaping an Arena for Innovation and Value Creation. Available at http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-18694

Laville S (2020) Air pollution a cause in girl's death, coroner rules in landmark case. *The Guardian*, December 16. Available at https://www.theguardian.com/environment/2020/dec/16/girls-death-contributed-to-by-air-pollution-coroner-rules-in-landmark-case

Mahajan S, **Kumar P**, **Pinto JA**, **Riccetti A**, **Schaaf K**, **Camprodon G**, **Smári V**, **Passani A and Forino G** (2020) A citizen science approach for enhancing public understanding of air pollution. *Sustainable Cities and Society 52*, 101800. https://doi.org/10.1016/j.scs.2019.101800

Mann L (2018) Left to other peoples' devices? A political economy perspective on the big data revolution in development: A political economy of big data in development. *Development and Change 49*(1), 3–36. https://doi.org/10.1111/dech.12347

Mergel I (2017) Open innovation in the public sector: Drivers and barriers for the adoption of Challenge.gov. *Public Management Review 20*(5), 726–745. https://doi.org/10.1080/14719037.2017.1320044

Milan S and Treré E (2020) The rise of the data poor: The COVID-19 pandemic seen from the margins. *Social Media + Society 6*(3), 2056305120948233. https://doi.org/10.1177/2056305120948233

Nam T and Pardo T (2011) Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, pp. 282–291. Dg.o '11. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2037556.2037602

Nesti G (2018) Co-production for innovation: The urban living lab experience. *Policy and Society 37*(3), 310–325. https://doi.org/10.1080/14494035.2017.1374692

Pye S, **King K and Sturman J** (2006) *Air Quality and Social Deprivation in the UK: An Environmental Inequalities Analysis* [Final Report to Department of Environment, Food and Rural Affairs]. Department of Environment, Food and Rural Affairs (DEFRA). Available at https://uk-air.defra.gov.uk/assets/documents/reports/cat09/0701110944_AQinequalitiesFNL_AEAT_0506.pdf

Rosenzweigova I and Skoric V (2016) *Civil Participation in Decision-Making Processes. An Overview of Standards and Practices in Council of Europe Member States* (p. 55). European Committee on Democracy and Governance. https://www.oecd.org/cfe/cities/OECD_Policy_Paper_Smart_Cities_and_Inclusive_Growth.pdf

Sadowski J (2019) When data is capital: Datafication, accumulation, and extraction. *Big Data & Society 6*. https://doi.org/10.1177/2053951718820549.

Sambasivan N, **Kapania S**, **Highfill H**, **Akrong D**, **Paritosh P and Aroyo LM** (2021) 'Everyone wants to do the model work, not the data work': Data cascades in high-stakes AI. In *Proceedings of the 2021 ACM Conference on Human Factors in Computing Systems*, Yokohama, Japan: CHI'21. pp. 1–15 https://doi.org/10.1145/3411764.3445518

Schaefer T, **Kieslinger B and Fabian CM** (2020) Citizen-based air quality monitoring: The impact on individual citizen scientists and how to leverage the benefits to affect whole regions. *Citizen Science: Theory and Practice 5*(1), 6. https://doi.org/10.5334/cstp.245

Simon J and Davies A (2013) People powered social innovation: The need for citizen engagement. *Social Space*, 38–43.

Taylor L (2017) What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society 4*(2). https://doi.org/10.1177/2053951717736335

Thatcher J, **O'Sullivan D and Mahmoudi D** (2016) Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space 34*(6), 990–1006. https://doi.org/10.1177/0263775816633195

Tinworth A (2016) *Hacking the Hack: Routes to Community Engagement with Open Data*. Open Data Camp UK. Available at https://www.odcamp.uk/hacking-the-hack-other-routes-to-community-engagement-with-open-data/ (accessed 26 February 2024).

Turow J, **Hennessy M and Draper N** (2015) *The Tradeoff Fallacy—How Marketers Are Misrepresenting American Consumers and Opening Them up to Exploitation*. Penn Collection Departmental Papers (ASC). Available at https://repository.upenn.edu/asc_papers/521 (accessed 26 February 2024).

Wachter SM (2019) *What's fueling the smart city backlash? Knowledge at Wharton (Blog)*. Available at https://knowledge.wharton.upenn.edu/article/whats-behind-backlash-smart-cities/ (accessed 22 August 2022).

Webster CWR and Leleux C (2018) Smart governance: Opportunities for technologically-mediated citizen co-production. *Information Polity 23*(1), 95–110. https://doi.org/10.3233/IP-170065

White SC (1996) Depoliticising development: The uses and abuses of participation. *Development in Practice 6*(1), 6–15. https://doi.org/10.1080/0961452961000157564

https://doi.org/10.1017/dap.2023.43 Published online by Cambridge University Press

Wilkinson M, **Dumontier M**, **Aalbersberg I**, **Appleton G**, **Axton M**, **Baak A**, **Blomberg N**, **Boiten J**, **da Silva Santos L**, **Bourne P and Bouwman J** (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 3*(1), 1–9.

Worthy B (2012) Where Are the Armchair Auditors. Available at https://opendatastudy.wordpress.com/2014/01/12/where-are-the-armchair-auditors/comment-page-1/

Yawson RM (2021) The ecological system of innovation: A new architectural framework for a functional evidence-based platform for science and innovation policy. arXiv preprint arXiv:2106.15479. (accessed 26 February 2024).

Yin RK (1994) Discovering the future of the case study. *Method in Evaluation Research. Evaluation Practice 15*(3), 283–290. https://doi.org/10.1177/109821409401500309

Yin RK (2003) *Case Study Research: Design and Methods*. Thousand Oaks, CA: SAGE.

Appendix A: List of sources

- Data Pitch
  - Grant agreement (#732506-2)
  - MET challenge: https://datapitch.eu/challenges/challenges-2018/dpc8-2018/
  - HOP public profile: https://datapitch.eu/HOPu/
  - HOP work plan (confidential)
  - Interview with HOP programme mentor
- SCIFI
  - Interreg application form
  - Business case for challenge III, air pollution in urban "street canyons," Mechelen
  - Business case for challenge III, air pollution in urban "street canyons," Bruges
  - Air quality challenge: http://smartcityinnovation.eu/open-calls/air-quality/
  - HOP public profile: http://smartcityinnovation.eu/hopu/
  - Nazka public profile: http://smartcityinnovation.eu/nazka-4/
  - HOP work plan (confidential)
  - Interview with stakeholders in Mechelen
  - Interview with stakeholders in Bruges
- Other
  - Interview with HOP CEO

Appendix B: Cascading funding process

In both programmes, funding calls were arranged following the same structure:

1) *Preparation*: Requirements were identified, call documentation (such as applicant guides and application forms) prepared, and an application submission system set up. The call outline and selection criteria were based on the goals described in the programmes' Grant Agreement, in order to select projects that met the overall objectives of the programmes and their funders.
2) *Call*: Applications were written and submitted, and applicants supported, for example, through webinars or email, which should also outline how the project addresses the overall goals.
3) *Evaluation*: Applications were screened for eligibility, and eligible applications reviewed by experts based on predefined criteria, including the degree to which the projects meet the selection criteria and alignment with programe objectives. They may select a shortlist for interview, directly select applicants for the programme, or a mixture of both.
4) *Negotiation*: Each selected applicant was brought into the legal framework of the funding programme, through a contract including a work plan and budget outline for the duration of their engagement, which also outlines their commitment, objectives, and Key Performance Indicators (KPIs)

Once contracts were signed, successful applicants joined a 6-month acceleration programme, with funds released in lumps at different stages, support provided throughout, and intermittent progress assessments against the defined KPIs.

Cite this article: Thuermer G, Walker J, Simperl E and Carr L (2024). When data meets citizens: an investigation of citizen engagement in data-driven innovation programmes. *Data & Policy*, 6: e12. doi:10.1017/dap.2023.43