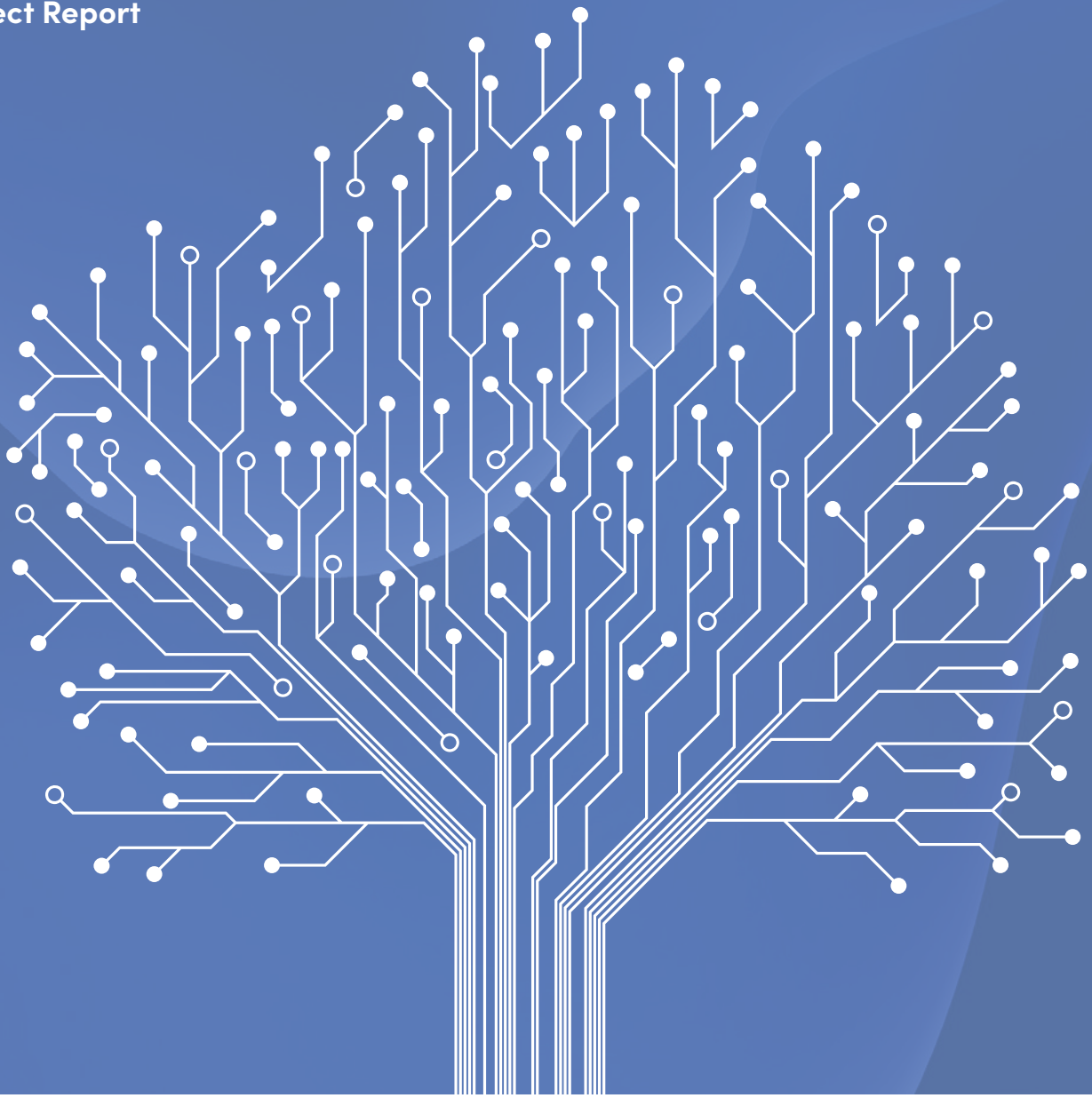
Leverhulme Research Centre for Forensic Science – Science, Policy and Law series #3

Valuing Public Sector Data in Scotland and Europe:

Data Governance for Economic, Environmental and Social Development

Project Report



University of Dundee

scottish universities insight institute
mobilising knowledge for a better Scotland

University of Strathclyde Glasgow

Authors

1 **Angela Daly**
University of Dundee

2 **Esperanza Miyake**
University of Strathclyde

3 **Andrew Porter**
University of Strathclyde

4 **Anna Berti Suman**
Libera Università Internazionale degli Studi Sociali

5 **Niamh Nic Daeid**
University of Dundee

6 **Stephanie Switzer**
University of Strathclyde

7 **Lucille Tetley-Brown**
University of Dundee

8 **Julián Valero Torrijos**
University of Murcia

Daly, Miyake and Porter led the writing of this report with input from the rest of the listed co-authors.

Acknowledgements

This is the final report from the Scottish Universities Insight Institute (SUII) programme 'Valuing Public Sector Data in Scotland and Europe: Data Governance for Economic, Environmental and Social Development' (Themed Call 2022/23: Scotland in Europe). We thank SUII for funding and supporting this work. We also thank all speakers, chairs and delegates who participated in our workshops in 2023 and our final hybrid workshop in January 2024, including Tim Murray-Browne for sharing his artwork during our public evening event in Glasgow. We thank the Glasgow Media Co-op for designing this report and Atlas Translations for copyediting the English version and translations into French and Spanish. We thank Strathclyde Print Services for the printed versions of this report.

Cite as: Daly, A., Miyake, E., Porter, A., Berti Suman, A., Nic Daeid, N., Switzer, S., Tetley-Brown, L., & Valero Torrijos, J. (2024). *Valuing Public Sector Data in Scotland and Europe: Data Governance for Economic, Environmental and Social Development*. (Science, Policy and Law Series; No. 3). University of Dundee.

Contents

Executive Summary

There is an abundance of data all around us in the digital age.

Data is collected by many different kinds of people and organisations, including the government and big businesses but also ordinary people. Data can give us many benefits and opportunities to improve society, the economy and the environment, and it can therefore meet the needs of diverse, plural publics. Doing this requires a good understanding of data and its uses in order to navigate this landscape, beginning with clarity about the range of issues and the definitional challenges around data and its rich complexities.

The 'Valuing Public Sector Data in Scotland and Europe: Data Governance for Economic, Environmental and Social Development' Programme, funded by the (now retired) Scottish Universities Insight Institute between October 2023 and January 2024, was a timely exploration into public (sector) data's benefits and challenges. During the programme, the nuances of ascribing value to data, and how that data may contribute in good – and bad – ways to the general public, were discussed and explored. We considered how 'public data' is gathered and used, and by whom, in Scotland, the European Union and beyond, identifying key insights to inform further research and policymaking in this area.

Here we give some context for our programme on public data, before providing an overview of the workshops we ran and key insights which emerged from them. Against this backdrop, we present a set of recommendations for policymakers under the following themes which emerged from our programme:

1 Definitions of 'public data'

2 Public understandings of data use

3 Value

4 Digital literacy

5 Health data, financial data and data related to the criminal justice system

6 Data sharing

7 AI considerations

01

Introduction and Context

Between 2022 and 2023, the Scottish Government worked with an Independent Expert Group (IEG) to drive discussions within its Unlocking the Value of Data (UVOD) programme, which considered private sector access to public sector personal data in Scotland.

As chair (Daly) and a member (Miyake) of this IEG, we wished to continue exploring some of the urgent questions that arose as a result of this work, beyond Scotland, beyond personal data and beyond private sector access – namely:

> **What is public (sector) data? How is it used? What value(s) is/are derived from it?**

These key questions now guide this report, the main outcome of a Scottish Universities Insight Institute (SUII)-funded programme, Valuing Public Sector Data in Scotland and Europe: Data Governance for Economic, Environmental and Social Development. A series of workshops in 2023 and 2024 brought together individuals from academia, policymaking and civil society, as well as practitioners and members of the general public. Through various activities, we shared research, policy insights and good practice around how public sector data can best be managed and its value realised in appropriate and ethical ways within the contexts of Scotland and Europe.

Our programme considered the issue of public sector data, its uses and governance to serve economic, environmental and social development purposes, from a comparative Scotland-Europe perspective. While the focus remained on examining the differences in terms of public data practices, policies, definitions and cultures between Scotland and Europe – especially within the post-Brexit environment – the programme also took into account cross-sector, cross-border and cross-disciplinary insights from the rest of the UK and the rest of the world beyond Europe (e.g. Algeria, Brazil, USA).

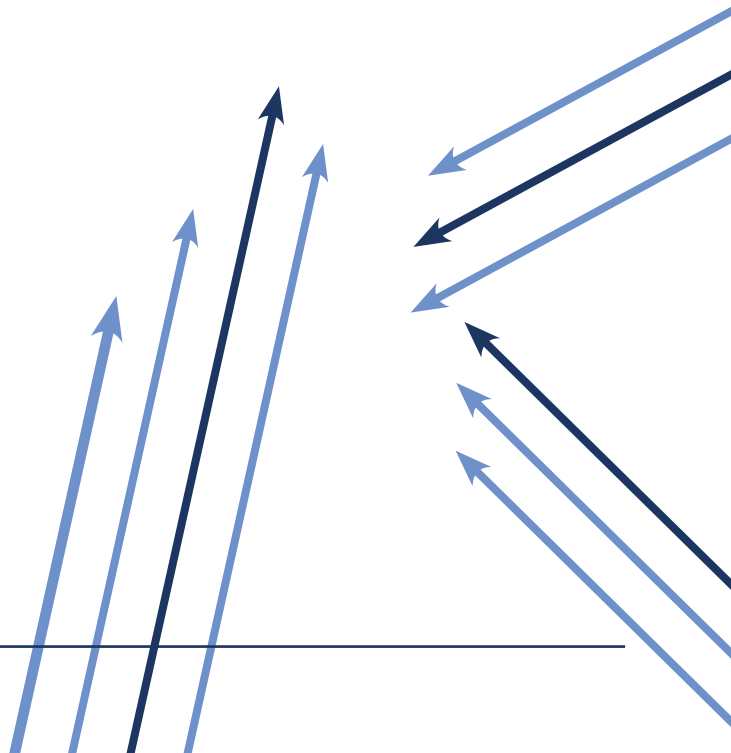
Public sector data gathering and data use for government objectives – in research, innovation and development, for example – are key issues internationally. This is due to the richness and comprehensiveness of data collected by the public sector, which is needed in a variety of administrative and democratic applications to support better government service provision. The significance of these issues also stems from data's potential value for onward and secondary uses in research, innovation and development, both within the public sector and by the private sector, the third sector and academia. In other words, the value of public data is one that has the 'holistic' potential to transform communities, economies, technologies and our environment.

The differing – and sometimes conflicting – perspectives, agendas and motivations between various groups around the value of public

In the Scottish context, the UK's departure from the European Union (Brexit) has had a significant impact on the legal and policy spheres, including as regards digital services and data.

data inherently create the need to open up discussions around 'value'. For example, there are complexities that arise from public authorities' access to data held by other actors, especially in the private sector. Similar complexities also apply to issues around government entities' ability to access (and rely on) other data sources, such as third sector sources, or even citizen science initiatives that encompass novel methodological approaches like (ethical and consented) crowd-sourced data. How can we reconcile such tensions? These issues are highly relevant to the policy agenda in Scotland, with the aforementioned Scottish Government UVOD programme, along with EU developments including the Data Act and the Data Governance Act, which create a framework for the development of sectoral data spaces pooling public, private and citizen-generated data.

At the outset, there are definitional issues around the notion of public (sector) data. The public sector may differ from country to country, and may encompass ideological positions on what services and activities should be provided by the state, and/or by the private sector. There are further complications brought about by public procurement and public-private partnerships in which 'public services' may be provided in full or in part by the private sector. Indeed, these intricacies are acknowledged by the EU Open Data Directive's definitions of 'public sector body', 'bodies governed by public law' and 'public undertakings'. There are further complications still when it comes to private services which are widely and intensively used by the public, such as proprietary social media networks or online platforms (e.g. TikTok or Amazon), which gather, monetise and surveil large amounts of data about users with little transparency or clarity of use. Other actors such as charities and non-governmental organisations (NGOs) may gather and analyse data about the public, whether alone or in collaboration with the public and private



Questions arise, though, if financial value is generated: what happens to it? Do the public – or, better, diverse publics – get a share of any financial gains?

sectors. In the digital age, with its associated data superabundance, there is a lack of certainty around how to determine what data is admissible or accessible, in what circumstances, and who is in a position to decide and arbitrate on this – an issue that may challenge established reference frameworks. Intellectual property (IP) and trade secrets can obscure what is being done with data from public view. Conventional notions of IP are subverted however by digital commons initiatives, which can also present new ways of organising data and digital public goods beyond conventional notions of public and private sector organisations. Rather than restricting our analysis to public sector data, to cover all these possibilities, we use the broader term of 'public data'.

In the Scottish context, the UK's departure from the European Union (Brexit) has had a significant impact on the legal and policy spheres, including as regards digital services and data. For the time being, the UK adheres to the EU's General Data Protection Regulation (GDPR) via implementation of these standards in its own laws. New EU legislation – such as the Data Act and the Data Governance Act which contribute to establishing the large public-private-citizen data spaces, and the AI Act which will regulate artificial intelligence – is not part of UK law and policy. As a result, the UK and EU approaches to governing data are diverging. It is also worth noting that the UK is taking a very proactive approach to data and the anticipated benefits, apparent in recent government policy outputs such as the National Data Strategy. While there is now a new Labour government in the UK, it remains to be seen whether this will usher in any fundamental changes to data policy.

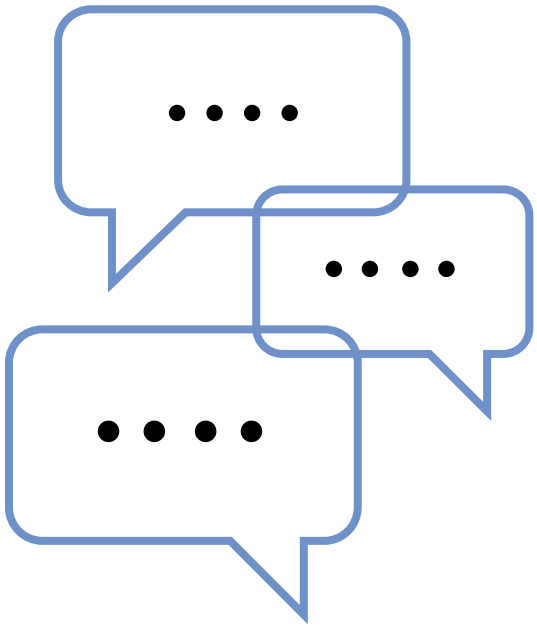
Furthermore, within the UK, devolution means that certain issues, notably health, are within the competence of devolved administrations such as the Scottish Government. This was thrown into sharp relief when the COVID-19 pandemic started and measures were imposed to contain it, which varied to some extent across different UK devolved nations. This in turn had a digital impact, with different contact-tracing apps being created by different devolved administrations and different data being collected. This continues to the present day, where although the pandemic continues, according to the World Health Organization (WHO), there has been a severe decline in the data governments collect, with only the Scottish Government within the UK continuing any kind of full-population monitoring in the form of wastewater data analysis. We also have further, local levels of government in Scotland, which provide their own digital services to the public and gather data about the public in a number of ways, adding to the layers of complexity in defining and understanding public data.

Such issues then lead to important questions in terms of why (and what) public data is collected, where the value of collecting it lies, and who is responsible for making these decisions. Public data is 'valuable' for both good and bad purposes – even if it is used in health ('good'), when collected unethically it can lead to detrimental effects ('bad'). While an objective, value-agnostic evaluation of what is truly, unarguably 'good' or 'bad' within pluralistic societies is out of reach, what this notion of 'value' means with regards to public data can be evaluated against prevalent societal norms. We looked to notions of economic, environmental and social development as benchmarks for public data's value. This acknowledges that value is broad and should not just be confined to the financial or economic. Questions arise, though, if financial value is generated: what happens to it? Do the public – or, better, diverse publics – get a share of any financial gains?

Furthermore, the public itself is also a complex notion. We have used 'publics' rather than 'the public' singular as a way of acknowledging the diversity of members of the public and publics when considering their interplay with public data in Scotland and the EU, and their differing views on and experiences of public data. This is also relevant to issues around inequalities in data, such as where, for instance, ethnic minority groups may be underrepresented or misrepresented in data because of the lack of data to begin with, or the problematic ways in which data was collected, as has happened in Scotland. However, there are also problems when data is collected about publics in exploitative and unnecessary ways, as we have seen with some areas of predictive policing. Involving publics in data is important, as well as considering public data beyond the public sector/government (e.g. citizen science initiatives) as these can even lead to new kinds of participatory governance and organisations, such as the MIDATA health data cooperative in Switzerland.

As we increasingly inhabit a world in polycrisis, with the ongoing pandemic, biodiversity loss, pollution, climate change and wars combined, data plays an important and complex role. For some, it will aid our green transition, although data also has its own environmental costs. For others, it will open up economic opportunities for social change, but equally create new challenges due to complex international financial infrastructures and/or constraints relating to a specific geopolitical and social contexts. These questions demonstrate the contingent and political nature of public data which can be used towards different and sometimes contradictory objectives. There is thus an urgent need to consider what constitutes 'good' and 'bad' uses of public data, who decides, and indeed who benefits from these uses.

Considering these issues will become even more urgent as we enter the age of AI, where algorithmic decision-making processes used in the public sector (e.g. police, health) will further complicate and even transform the practices, ethics, legalities and outcomes of using public data. Furthermore, with the wide and popular use of generative AI, one of the key challenges we will face is not just how we collect, access and use public data – and its value, whether 'good' or 'bad' – but also the need to question with greater rigour the modes of production of public data and the ends it is used for. It is crucial for future research, practice and policymaking to consider the role of AI in the generation, collection, analysis and use of public data.



02

Workshops and Insights

It is against this backdrop that we ran our programme, which consisted of three online workshops and one hybrid final workshop between October 2023 and January 2024.

Our workshops sought to understand and address the aforementioned issues around valuing public data by sharing research, policy insights and good practice on best management and use of public sector data to ensure its value is of public benefit from social, economic and environmental perspectives.

In partnership with other academics, policymakers, practitioners and public members from the EU, these workshops provided the opportunity to explore:

> **a) how Scotland is situated within broader post-Brexit European data governance frameworks;**
>
> **b) the diversity of different publics and their interplay with public sector data in Scotland and the EU;**
>
> **c) inequalities in data to ensure economic, social and environmental development is both ethical and equitable.**

The three online workshops addressed data for social, economic and environmental benefits, respectively, and the final hybrid workshop brought all of these themes together. Here we summarise some key topics and points which emerged and which, along with the context described above, inform our recommendations in the next section.

Workshop 1
Public Data for Social Development in Scotland and Europe
Thursday 5 October 2023 (online)

This workshop started with a series of invited talks on digital commons, human rights and health data, data science for good, AI in health, and data and inequalities.

We discussed challenges involved in anonymising data, particularly health data, needs for diversity in data, and public issues of trust, security and privacy. In the breakout rooms, discussions centred on what we can consider public data and who constitutes the 'public', how public data can be used for social development, and associated opportunities and challenges. Discussions also covered opportunities and challenges around using public data for social developments, including the thorny matter of demarcating development for social benefits. Participants recognised the challenges and nuances in defining public data, the dangers of large AI models and the erosion of the right to be forgotten, and public discourses around understanding of the use of technologies and data.

The group was unanimous in thinking it is inherently more useful to reflect on what is within and beyond our control when considering use of public data. There is a clear need to increase understanding on an individual level of what public data is, and, most importantly, how it is used/shared. Discussions also covered the changing value of data as a resource, the social implications of viewing data in this way (detachedly as an 'asset'), particularly health data, and how this may feed into issues facing humanity and the environment.

Workshop 2

Public Data for Economic Development in Scotland and Europe

Thursday 9 November 2023 (online)

This workshop also commenced with a series of talks – on open data financial architecture, data in economic development, collective intelligence and the impact of the forthcoming EU data spaces. With AI being a current global concern, some of these discussions also considered the use of AI for citizen science.

We discussed the benefits and challenges of integrating new data with existing data, current and upcoming EU legislation, and how the concept of 'digital altruism' – meaning how to use digital technologies and associated data to support the development of broad societal virtues, transcending traditional economic outlooks – translates into policy and decision making. Indeed, the EU's new Data Governance Act translates this notion through the concept of 'data altruism', which involves people and companies allowing their data to be used for the public interest for no cost or reward. Further topics of discussion centred on what we can consider public data and the 'public/s', how public data can be used for economic development, and associated opportunities and challenges. Discussions also covered differences in public data for economic development between the UK and EU since Brexit (and internationally), and lastly, who pays for and who benefits from public data for economic development.

In wider group discussions, participants recognised:

→ the challenges of dealing with data use within specific legal frameworks, and the costs and values of data maintenance and metadata;

→ difficulties in defining the quality of data under its specific terms of usage;

→ the question of thinking about data as a tangible asset, which is often central to the concept of economic development;

→ the inherent frameworks and modes of operation that result from this thinking, particularly while data has little inherent value outside of its modes of interpretation and usage.

Workshop 3

Public Data for Economic Development in Scotland and Europe

Thursday 7 December 2023 (online)

The third and final online workshop, on environmental development, featured talks on the right to access environmental information under the United Nations Economic Commission for Europe's Aarhus Convention, the Data Space for Smart and Sustainable Cities and Communities (DS4SSCC) initiative, and the EU's policy landscape, including the INSPIRE and Open Data Directives.

Discussions centred on what we can consider public data and the 'public/s', how public data can be used for environmental development, and the differences in public data for environmental development between the UK and EU since Brexit (and internationally). Discussions also covered opportunities and challenges when using public data for environmental development, with a focus on contextualising data across different sites including local, regional and national as well as urban and rural areas.

Participants recognised:

→ the complexities and tensions when building 'universal' frameworks across geographical areas with differing contexts and needs;

→ the need to empower both publics and public bodies with digital literacy in order to properly contextualise and make proper use of environmental data;

→ potential inequalities across data access and how data is used when comparing urban and rural areas across local, regional and national sites;

→ a need to re-think data sharing and how information flows, particularly in relation to what data is shared with publics from public bodies, what data is obtained from publics, and how data can be disseminated to increase public understanding of environmental needs and challenges.

Hybrid Workshop

Valuing Public Data for Social, Economic and Environmental Development

25-26 January 2024
Scottish Universities Insight Institute

Learnings from the online workshops on social, economic and environmental data fed into a two-day knowledge exchange hybrid workshop that brought together academics and practitioners from Scotland, the EU and beyond.

The hybrid workshop took place physically at the Scottish Universities Insight Institute, in the centre of Glasgow, and online via the Zoom platform, welcoming attendees from across the world. Throughout the two days, the workshop brought together a variety of talks, activities and perspectives. Both days began with a keynote talk, followed by a range of different sessions – from panels and a roundtable discussion to an interactive workshop – with academics and practitioners from different disciplines/sectors sharing their perspectives and experiences around public data.

The workshop also hosted a public event by artist Tim Murray-Browne at the end of the first day, 'Joy and Ethics in Making AI Art'. This free event, open to all and well-attended by the general public, explored the artist's audio and visual experimentations with artificial intelligence tools, and his use of his own data as an integral element in the creation of new AI-mediated artworks. The event brought to life some of the issues that had been discussed during the day, providing an opportunity for the public to also engage with these through Murray-Browne's work.

The whole workshop came to a close with a 'Research to Policy' roundtable talk, featuring members of data-focused organisations seeking to effect policy change. Speakers shared their experiences working with academic researchers and offered recommendations for academics seeking to make an impact on policy.

Cross-cutting Themes

Several discussion themes arose across the three workshops. Similar concerns and approaches to the concept of public data were often shared between these themes, and in some cases participants offered details of challenges and solutions.

Definitions of 'Public Data'

Participants recognised the nuances in defining 'public data', and the challenges in applying definitions in multiple contexts. Acknowledging these debates and complexities, and coming to a consensus on definitions, were considered productive in order to reach complementary outcomes across groups.

In addition, concepts of ownership and control are very difficult to parse when who owns the data has not been established. For example, in voluntary sector organisations, data exists in environments with no clear ownership. Might this be considered public data?

Public Understandings of Data Use

Participants recognised challenges in publics' understandings of nuances in data use. There is often an emphasis on data as intangible and factual matter, and a lack of awareness of both who is collecting data and whose data is being collected. Better public understanding of who is collecting data – and why – is needed.

There is often an assumption that public data is being shared as a standard mode of practice. Some parts of the public are hesitant to share data because of an assumption that it is being shared with someone else. This highlights a need to increase individuals' understanding of data sharing practices.

Examples were raised from localised data strategies, where those on the front line are trying to develop best outcomes. In some data strategies, data is deliberately manipulated to get the 'best' outcomes and does not accurately

represent reality. In turn, this can create confusion and a breakdown of understanding.

Key examples of gaps in public understandings of data use include data flows between the state and the public; data sharing practices from public health organisations to private companies which can create public distrust; and, under surveillance capitalism, the broader social problems beyond privacy arising from the extraction of the publics' data by digital platforms.

Digital Literacy

Participants across the workshops noted a need to increase the publics' capacities to understand and properly contextualise data relevant to their areas and needs. For instance, different geographical areas, across local, regional and national sites, may have different data definitions and uses due to cultural and historical parameters. It is important to remember that there are no answers inherent in data, and they are instead revealed through framing and the processes with which the data is used. This all applies to real-world use, especially when looking at expansive data use across organisations and governments. There is also a need to increase the public sector's own understanding of data use, meaning and contextualisation, bearing in mind that the public sector is ultimately composed of individuals who may not all share the same expertise and understanding. As individual public servants, those working in the service of publics need to have the competencies to enable them to maximise digital age opportunities while being vigilant about the risks of harm to those the public sector has a mandate to provide for (whether as the government or other public service entities).

Health Data

Across each of the workshops on social, economic and environmental data, health data was raised as a topic of interest. In particular, participants noted how variations in use and management of public health data can contribute towards health inequalities. When using health data, there are complexities around the balance of opening up data and keeping patient data private. Anonymisation of data and safeguarding of data are essential in some instances. Pseudonymised data presents its own issues, where individuals might still be identifiable if data is presented at a large scale and with enough identifiable attributes.

There are significant challenges concerning data on attributes that can be used to identify individuals, such as unusual or rare pathologies, or rural areas (i.e. smaller populations), where individuals may be more likely to be re-identified. Other challenges include instances of individuals being identified or held to a prior identity which may have changed. In the UK, extensive efforts are under way to create trusted research environments aiming to prevent re-identification.

There are benefits to joining up health data at scale (data linkage) and building interoperable systems around the same data. Examples were given regarding the need to avoid having to retell the same stories when consulting with multiple healthcare professionals. More efficient processing of data and interoperability are vital to help overcome these barriers.

Finally, it was noted that within healthcare, ethics processes and research standards are taken very seriously. This highlights a need for a wider code of ethics and standards to help govern acceptable and unacceptable uses of data. The forthcoming EU AI Act provides a complementary example of the kinds of standards and frameworks that could be implemented.

Data Sharing

Participants noted that concepts of data sharing, and subsequently the protection of personal data, are often a source of tension. Data is often seen as something that public service acts on, before being seen as a resource in service to publics. In particular, data flows between publics and the state are not fully balanced. The public's right to submit data is limited in many ways, through legalities and what the state is willing to accept, as well as lack of sufficient resources within the state to handle and process information and data submitted by publics. This is complicated by expectations of accountability and transparency, topics that do not always go easily together, especially in political settings where diverse publics, sometimes with opposing preferences or ideologies, are served. It is broadly acknowledged that governments and their public sector agencies have a dominant culture of caution and risk-aversion. This may make shifting to new modes of working in the face of digital developments, like the ascendance of vast magnitudes of data in the past decade, a slower and more awkward undertaking.

Participants noted that the common interpretation of data as a tangible 'asset' is often central to the economic development conversation and seen as integral to an information society. However, data is shaped and defined by the ways in which it is used.

There was also a repeated focus on the concept of 'digital altruism', and discussions centred on whether this is a correct approach to policy and decision making. In many cases, reciprocity in data sharing and use is desirable. Without these expectations, the exchange of data becomes a different mode of exchange. For example, there are some use cases of private data being used in public settings or for research. This in turn creates challenges around perceptions and transparency of use, as well as policy. One major challenge lies in persuading private companies to behave in an altruistic manner, especially around the use, sharing and management of data.

Finally, the point was raised that different forms of data do not have the same value, and the costs of maintaining different forms of data can vary. For instance, in some cases metadata may have a higher 'value' when used and shared.

Integration of New and Existing Data – Participants noted the difficulties, and potential benefits, of integrating new and historical data sets. While new projects can often reveal new sources of data, they can also reveal new ways of interpreting and working with existing data. These new methods of interpretation can increase the use value of existing data, and in turn contribute to the people and communities from which data is sourced. In addition, participants noted that new data can be used to check, verify and enhance existing data. Effective use of data often lies in bringing both new and existing forms of data into dialogue. This is envisaged in the forthcoming EU data spaces, an emerging model for sharing new and existing data, which may have a major impact on how public data sets are combined and used.

Perceptions of Data in Objective Terms

Participants noted that the common interpretation of data as a tangible 'asset' is often central to the economic development conversation and seen as integral to an information society. However, data is shaped and defined by the ways in which it is used. The conceptualisation of data as an asset leads to an imperative to always collect, and create, more data.

In addition, the non-fungibility of data within the digital commons is important to perceiving it not just as private property or as a traditional 'asset'. Data may not necessarily be of high quality under its specific terms of usage when the focus is placed on objective attributes. Effective use of data, and the reasons behind its use, are ultimately what matter.

03

Key Recommendations for Policymakers

Based on our programme, we have devised the following recommendations, in particular for policymakers working on public data:

1. Definitions of 'Public Data'

We recommend that policymakers recognise the challenges in defining what public sector data or 'public data' is. This includes consideration of who is part of that 'public' (there may be more than one kind of public or publics), and other contextual and structural issues (there may be regional or environmental differences). Furthermore, we recommend policymakers also consider these issues along with concepts of control over and ownership of data, and engage with public discourses and understandings of how data and technologies are used.

2. Public Understandings of Data Use

We recommend that policymakers seek to understand how publics themselves understand 'public data' and how they perceive their data is being collected, by whom, and why. These should inform how policymakers increase publics' understanding(s) of what public data is, who is collecting it and why, and how it is being used and shared.

3. Value

We recommend that policymakers identify and assess the meaning of the 'value' of public data, also taking into account how different publics value data. A first step in this process should be to identify what and whose norms and benchmarks are being used to evaluate the 'value' of public data. We strongly encourage assessing 'value' beyond financial or economic value, understanding these in relation to wider issues such as social or environmental value, and embracing complexity in handling these topics rather than seeking to oversimplify and missing nuances.

4. Digital Literacy

We recommend that policymakers seek to increase publics' capacity to understand and properly contextualise data relevant to their areas and needs. In addition, we recommend that policymakers seek to increase the public sector's own understanding of data use, meaning and contextualisation, bearing in mind that the public sector is ultimately composed of individuals who may not all share the same expertise and understanding.

5. Health Data, Financial Data and Data related to the Criminal Justice System

We recommend that specifically for health data and health, policymakers, practitioners and researchers acknowledge the complexities around the balance between opening up data and keeping patient data private, and the challenges concerning data on attributes that can be used to identify individuals. At the same time, ensuring joined-up and interoperable data where appropriate is necessary so individuals do not need to retell stories when consulting with multiple health professionals. Weighing the potential value against the harms of using private data for the benefit of the publics is, of course, a consideration that extends to other kinds of sensitive data, such as financial data and data in the criminal justice system. This sort of balancing is a highly skilled undertaking. We thus recommend that policymakers and practitioners in other public data sectors learn from the ethics processes and research standards already in place – such as in health – to devise wider codes of ethics and standards to help govern acceptable and unacceptable uses of data.

6. Data Sharing

Policymakers and practitioners should acknowledge the tensions within data sharing, including vis-a-vis power and resources imbalances between publics and the state, and between publics and companies. Policymakers should also consider ways in which greater 'data altruism' from the private sector can be achieved in terms of private sector actors sharing their data with the public sector and with the public.

7. AI Considerations

We recommend policymakers, practitioners and researchers acknowledge the increasingly widespread use of AI in the generation (making), analysis and use of (public) data and consider its implications. This consideration should involve a broad view of the costs and benefits of AI, the appropriate role of public data in AI ecosystems (taking account of what is useful and valuable for publics) and questions of ownership of and control over AI.

University of Dundee

scottish universities insight institute

mobilising knowledge for a better Scotland

University of Strathclyde Glasgow