Private sector access to public sector personal data: exploring data value and benefit-sharing

TABLE OF CONTENTS

| Table o | of Contents | 1 |
|-----------|---|----|
| Execut | ive Summary | 2 |
| Introdu | iction | 7 |
| Backg | ground to the study | 7 |
| Key d | efinitions | 9 |
| Method | dology | 11 |
| Inclusio | on and exclusion criteria for the literature review | 12 |
| Discus | sion | 14 |
| Struct | ture and approach | 14 |
| A. Un | derstandings informing this review | 14 |
| B. Hig | ghlights from the thematic analysis | 17 |
| I. | Public benefit | 17 |
| ii. | Conceptions of value | 21 |
| iii. | Benefit-sharing | 23 |
| iv. | 'Public Benefit' Duty to Share? | 25 |
| V. | Public data governance models | 26 |
| | tential overarching principles for the (benefit) sharing of publi | • |
| | global level | |
| - EU wide | | |
| - Cou | ntry-specific insights | |
| i. | The United Kingdom (UK) | |
| ii. | Australia | |
| iii. | The United States (U.S) | |
| iv. | The Netherlands | |
| D. Ca | se studies insights | 37 |
| Referer | nces | 47 |
| Rios of | the Authors | 53 |

EXECUTIVE SUMMARY

The aim of this review is to enable the Scottish Government to explore the issues relevant to the access of public sector personal data (as defined by the European Union General Data Protection Regulation, GDPR) with or by the private sector in publicly trusted ways, to unlock the public benefit of this data. This literature review will specifically enable the Scottish Government to establish whether there are

- (I) models/approaches of costs/benefits/data value/benefit-sharing, and
- (II) intellectual property rights or royalties schemes regarding the use of public sector personal data with or by the private sector both in the UK and internationally.

In conducting this literature review, we used an adapted systematic review, and undertook thematic analysis of the included literature to answer several questions central to the aim of this research. Such questions included:

- Are there any models of costs and/or benefits regarding the use of public sector personal data with or by the private sector?
- Are there any models of valuing data regarding the use of public sector personal data with or by the private sector?
- Are there any models for benefit-sharing in respect of the use of public sector personal data with or by the private sector?
- Are there any models in respect of the use of intellectual property rights or royalties regarding the use of public sector personal data with or by the private sector?

For all questions our inquiry focused on both the UK and internationally but was limited to literature available in English. Our literature review returned **7,715 items**. We carried out additional citation tracing and also took on board literature suggestions by the commission team. After applying agreed inclusion and exclusion criteria, **55 sources** were included in this study (Appendix I).

Our key findings can be summarised as follows:

• In the literature reviewed, notions are not understood in a straightforward way and their nuances should be embraced and disclaimed as appropriate.

- Public-sector bodies generally lag behind in developing and implementing data sharing regimes and data analytics as compared to the private sector; however, there are existing good practices in specific policy areas – such as in public health – where governments' sharing of personal data with the private sector is becoming a component of many governments' health and economic strategies.
- Studies demonstrate that ordinary people are supportive of health and social care data being used for public benefit but wish those public benefits to outweigh private profits and interests.
- When assessing costs and benefits of data sharing with the private sector, the studies assessed indicate that such costs and benefits should not be conceived of as solely financial but understood in broader, more social terms. In a related sense, while governments often conceptualise public benefit in economic terms, the public itself may not share this view and there is a need to distinguish between the framing of public benefit espoused by governments, with that understood by a plurality of public(s).
- The term public benefit is not well defined in the literature and is often used interchangeably with other terms such as the public interest and the common good. Where attempts are made to define public benefit, the literature indicates that this may include direct benefits – such as an income stream to the UK National Health Service (NHS) – and indirect benefits.
- There were prerequisites for achieving public benefit identified in the literature
 with values such as transparency that are very much connected to public
 benefit. By extension, public engagement in terms of data assessment is
 necessary as an input for public benefit, and to ensure a social license such
 that data is shared for the common good.
- There is some concern within the literature that a lack of a definition of public benefit may enable the concept to be exploited to facilitate access to, and commercialisation of, government-held personal data. Accordingly, the lack of a definition for public benefit could leave room for exploitation of the concept to promote government economic interests, which may not align with public conceptions of public benefit.
- There is a vast literature on data value in general. In relation to understandings of the 'value' of personal data, some of the literature surveyed indicated that this is jurisdiction specific and informed by the underlying context and socio-technical settlement. An additional theme in the literature surveyed was that the value of publicly held personal data cannot simply be defined in economic terms. Where the value of publicly held personal data is conceptualised in narrow economic terms, this is not sustainable in the long term, particularly in the healthcare domain.
- There were concerns that the so-called 'assetisation' of personal data may influence conceptions of value, thereby potentially resulting in a lack of public

scrutiny and inequity in terms of access to knowledge. To overcome issues with 'assetisation' - and taking account of the socio-technical underpinnings of conceptions of value - value co-creation and exchange beyond the market was suggested as a way to take such issues out of the economic arena and re-contextualizing them to embrace co-creative approaches to how value is conceived of.

- Benefit-sharing is a concept typically associated with international environmental law and in particular, international biodiversity law to deliver commutative and distributive justice. Benefit-sharing is thus linked to justice and emphasises the optimisation of benefits to society, together with the minimisation of harm, and the achievement of equity.
- If data has the potential to benefit the public, it should be shared. Public benefit cannot be obtained in the absence of such sharing. However, the absence of common principles for trusted government-to-businesses (G2B) personal data sharing may lead to undue restrictions on data flows resulting in detrimental economic impacts. The literature surveyed identifies this and constraints for promoting G2B data sharing. Legal certainty was hence identified as required for such sharing to take place.
- Creative and fruitful collaboration schemes are being established between civic organisations and private actors (at times also engaging the public sector) to share personal data, as occurs for certain citizen science activities, with creative commons licensing schemes, value co-creation, and the reality of 'data cooperatives'.
- There is a growing attention in the literature for the concept of 'data altruism' as also incorporated in the European Union Data Governance Act; this reflects a tendency to embrace a fair and open sharing of personal data for public benefit.
- In the literature surveyed, we identified several guiding principles that could be of help for designing an appropriate framework for public sector personal data (benefit) sharing, among the others key principles are of proportionality, transparency, public engagement, co-creation of the concept of value, legal certainty and respect for ethical values and norms.
- Cross-national initiatives are being set up to provide options for the operational framework and governance models for the exchange and secondary use of personal data between countries and actors, respecting the principles of transparency, trust, FAIRness, citizen empowerment and common good.

Based on our findings, we recommend the following steps:

• When using certain terms in a policy document or official recommendations, remember that notions are not understood in a straightforward way in the

literature and their nuances should be embraced and disclaimed as appropriate;

- Start with pilot sharing in those fields where studies demonstrate that ordinary people are supportive of health and social care data being used for public benefit but make sure that public benefit outweigh private profits and interests;
- Assess costs and benefits of data sharing with the private sector not only from a financial but also broader perspective, including broader social dimensions;
- Be aware of how framings from government as well as other dominant actors (such as market actors) can erode or undermine 'true' public benefit and engage the public via processes of public engagement on the assessment of public benefit and in the co-creation of the notion of value;
- Emphasise principles of commutative and distributive justice in considering benefit-sharing arising from the use of publicly held personal data;
- Implement effective strategies that tackle the identified constraints for promoting G2B data sharing, for example the absence of common principles on the matter:
- Turn to creative and fruitful collaboration schemes and initiatives existing between research centres, civic organisations and private actors (at times also engaging the public sector) to share personal data, taking for example certain citizen science activities and the reality of data cooperatives and of creative commons licensing schemes;
- Respect key principles such as that of proportionality, transparency, accountability, and respect for ethical values and norms in designing frameworks for public sector personal data (benefit) sharing. Value cocreation should also be promoted in the construction of benefits.
- Seek to build social licence as a fundamental resource to ensure that private data sharing with businesses operates as a trigger for public good.

In terms of further work and study which could be undertaken to enhance understanding and knowledge in this area of focus;

- While beyond the scope of this literature review, it may be useful for the Scottish Government to define key terms to facilitate discussions in this field with a view to ensuring shared understandings of such terms.
- While also largely beyond the scope of this literature review, there may be
 much to learn from surveying benefit-sharing arrangements in the biodiversity
 world where benefit-sharing is firmly embedded as both a principle and an
 outcome of access to certain forms of 'information. While such arrangements
 were included in this review to the extent they fell within our search focus, a

- broader review would arguably do greater justice to the potential commonalities across these areas.
- Our study, by necessity, focused on the sharing of personal data by public bodies and reviewed the literature within this context. There would be merit, however, to elicit further studies specifically focused on concepts such as value within data more generally. This would arguably yield a wider, and potentially more rounded, approach to the concept of value.

INTRODUCTION

BACKGROUND TO THE STUDY

The aim of this review is to enable the Scottish Government to explore the issues relevant to the access of public sector personal data (as defined by General Data Protection Regulation, GDPR) with or by the private sector in publicly trusted ways, to unlock the **public benefit** of this data.

This literature review will specifically enable the Scottish Government to establish whether there are models of costs/benefits/data value/approaches to benefit-sharing or the use of intellectual property rights or royalties regarding the use of public sector personal data with or by the private sector both in the UK and internationally.

This literature review is the final in a series of three. The literature review, 'Unlocking the Value of Scotland's Data: Public Engagement Around the Access of Public Sector Data with or by <u>Private Sector Organisations</u>, undertaken by Dr Sonja Erikainen and Professor Sarah Cunningham (University of Edinburgh) aimed to explore the question: (w)hat public engagement has been undertaken regarding the use of public sector data by or with the private sector? The review made a number of key findings;

- There is widespread conditional acceptance of private sector use of public sector data especially among informed publics. Public benefit is the primary driver of acceptability and commercial gain or private profit the primary driver of unacceptability. The definition and scope of 'public benefit' is open and contested, but publics want to see evidence that public benefit of some kind is the primary driver of public sector data access, that it can actually be achieved, and that it outweighs any possible private benefits.
- Publics want to see the development of equitable benefit-sharing models for collaborations or partnerships between private and public sector organisations, as they expect benefits – including profits – to be returned to publics and reinvested into the public sector.
- Public trust and distrust are key factors around private sector access to public sector data. Publics are more trusting of private sector uses of public sector data when public sector organisations retain control over the data during collaborations with the private sector.
- Publics expect to see stringent oversight, governance, and safeguard arrangements around private sector use of public sector data, especially concerning an oversight or governance body, transparency and accountability

processes, and arrangements for data security and safety, consent, and confidentiality.

- Publics want there to be public involvement or engagement processes and activities around private sector use of public sector data.
- There is no singular 'public perspective' on private sector use of public sector data, but rather, while overarching patterns can be identified, publics are plural, and individuals' views are shaped by a diverse range of intersecting demographic and attitudinal variables.

An additional review, 'Public Sector Personal Data Sharing', undertaken by Morgan Currie, Steven Earl, Victoria Gorton and Matjaz Vidmar, was tasked to review pathways for public sector organisations sharing a particular type of data, called personal data, to private organisations. It made the following recommendations;

- Focus on creating shared data standards and protocols across agencies and local and national contexts - a public agency could be dedicated to this role.
 These data standards should create confidence in the quality of the data as well as the consistency of the data sets.
- 2. Develop a clear way to define and demonstrate public interest and public value. This includes involving the public meaningfully from early stages in the designs of the data sharing infrastructures and models to generate public licence.
- 3. From earliest phases, develop ways to market the value and utility of the data sharing infrastructure to immediate stakeholders and users (i.e. researchers, private sector innovators) and be transparent about the risk and opportunities. Ways of doing this include involving stakeholders in the designs of the infrastructure or creating a typology of data and datasets that may be of value.
- 4. Develop a central resource or agency, such as a data permit authority, that helps aggregate, combine and link data and has the autonomy to decide which permissions to grant, as well as the resources needed to provide quality data. Make the process of this as transparent as possible.
- 5. Share ethical standards and best practices internationally. Develop and maintain an international community of practice that explores the "Futures of GDPR". Many countries and regions developing data-sharing approaches will have similar needs for support and learnings, which they could do from each other.

Our literature review follows on from the previous two reviews outlined above. In this adapted systematic literature review, we aimed to address a number of core questions, develop a thematic analysis when doing so, and provide recommendations to guide the (benefit) sharing of public sector personal data;

- Are there any models of costs and/or benefits regarding the use of public sector personal data with or by the private sector?
- Are there any models of valuing data regarding the use of public sector personal data with or by the private sector?
- Are there any models for benefit-sharing in respect of the use of public sector personal data with or by the private sector?
- Are there any models in respect of the use of intellectual property rights or royalties regarding the use of public sector personal data with or by the private sector?

Our literature review focused on both the UK and internationally, but was limited to literature available in English, published after 2007 and until mid-September 2022. In conducting our literature review, we relied upon the following key definitions;

KEY DEFINITIONS

Personal data

According to Article 4 of the GDPR, personal data 'means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.' S 3(2) of the DPA 2018 defines personal data as meaning, 'any information relating to an identified or identifiable living individual.'

Public sector private information

As defined by the EU General Data Protection Regulation (GDPR) and the UK Data Protection Act of 2018.

Business to Government data sharing (B2G)

Business to Government data sharing is the scenario in which a company or other private organisation makes available its data to the public sector. This *incidentally* is addressed in some of the resources reviewed.

Government to Business data sharing (G2B)

Government to Business data sharing is the scenario in which the public sector shares its data with a company or other private organisation. This is *the focus* of our review.

The two categories of B2G and G2B are informed by recent conceptualizations from academic and grey literature (e.g. <u>Aidinlis 2022</u> and <u>Martens and Duch-Brown 2020</u>).

METHODOLOGY

Methodological approach

A stepwise approach was followed to perform the review, as set out below:

- Desk research to devise the search strategy, relevant keywords, and inclusion/exclusion criteria (see Table 1);
- Data extraction to a pre-piloted form in Google sheets
- Thematic analysis of included literature;
- Production of a written report, synthesising the results of our thematic analysis and putting together recommendations based on the reviewed literature and thematic analysis to inform the Digital Directorate's programme and for sharing with partners and wider stakeholders.

We worked between 1st June 2022 and mid-September 2022 and considered the cut-off date for literature inclusion the year 2007. Limitations of the study include the short time frame, making the literature search not as comprehensive as it would have been had more time have been available for the project. Our extensive key chain had to be adapted to suit the databases we used; an issue discussed with the Digital Directorate. Other limitations included that the rapid nature of our literature review meant that each source was only reviewed by one team member, therefore introducing the possibility of error and/or bias. Had more time been available, all literature would have been reviewed by both team members. Our inquiry focused on literature produced both in the UK and internationally but was limited to literature available in English. Lastly, the literature on the targeted matter is scant, making this study inclusive also of adjacent literature, not always perfectly fitting to the criteria. Excluding the usual limitations of a rapid review and the language barrier, the proposed methodology is as robust as it can be, given the above noted constraints.

INCLUSION AND EXCLUSION CRITERIA FOR THE LITERATURE REVIEW

| Inclusion criteria | Exclusion criteria |
|---|--|
| All documents with a focus - either primary or secondary - on the use with or by the private sector of public sector personal data. | All documents where the use of public sector personal data with or by the private sector is either not included in the study or is incidental to the study. |
| All documents with a focus - either primary or secondary - on the use with or by the private sector of public sector personal data which discuss some form of benefit, benefit-sharing, value, royalties, models, intellectual property and/or costs. | All documents with a focus - either primary or secondary - on the use by the private sector of public sector personal data which do not discuss some form of benefit, benefit-sharing, value, royalties, models, intellectual property and/or costs. |

Table 1. Inclusion and exclusion criteria

Search strategy and inclusion criteria

We conducted a rapid systematic literature search of Scopus, JSTOR, Google, Google Scholar, ResearchGate, Academia Edu using the following key word chain, tailored and adapted as appropriate - and after discussion with the Digital Directorate¹ - for each database:

({public sector} OR government) AND ({personal data} OR {personal information} OR {private information} OR {personal details}) AND ({private sector} OR (private AND industry) OR {private compan*} OR commercial OR profit OR {profit share} OR financial*) AND ({data sharing} OR {data linkage} OR (sharing AND data) OR {information sharing} OR (data AND exchange)) AND (costs OR benefit* OR value

_

¹ This was the case for our search of JSTOR which we had to use an adapted, more limited keyword search for.

OR benefit-sharing OR {benefit-sharing} OR value OR {intellectual property} OR royalties OR licen* OR fees OR patents OR model OR governance)

The key word chain was reviewed by the Commissioning team from the Digital Directorate.

Our initial search of the above databases resulted in a total of **7,715 items returned**. We subsequently reviewed the titles, key words, abstracts and, where required, the full texts of the returned documents. After exclusion of duplicates and after further citation tracing, as well as taking on board additional suggestions from the Commissioning team, **55 sources** met the inclusion criteria. The outputs of the selection process are outlined in Appendix I - Comprehensive table with all relevant literature surveyed.²

The studies contained in the table were surveyed according to the following categories:

- URL/DOI
- Title
- Journal/other source name including whether peer reviewed
- Date
- Brief description & key message
- Main thematic tag(s) (see Appendix I)
- public benefit/societal benefit/ benefit-sharing/public good
- intellectual property/patents/license fees/royalties
- reference to models, real world case study/ies discussed

_

² The tags used can be found in Appendix I.

DISCUSSION

STRUCTURE AND APPROACH

We structure our discussion in the following way: in **section A**, we set out some intermediate findings which we came to - and discussed with the Commissioning team - when performing our review. These findings helped to shape our approach to both the analysis of the literature as well as our writing of the report. We include these findings, after discussion with the Commissioning team, to enable full transparency and ensure the reader is aware of the understandings which shaped our work.

In developing our thematic analysis, we were guided by the questions posed in this study. The results of the thematic analysis are set out below (**section B**). We build upon this thematic analysis in **section C**, in which we extract some potentially useful overarching principles of relevance to this study, again guided by the questions asked. In **section D**, we set out some potentially useful case studies before setting out our key findings and commendations.

A. Understandings informing this review

1. The notion of public sector personal data sharing

We were able to come to the preliminary finding that literature focusing strictly or exclusively on the sharing of public sector personal data (as defined by the GDPR) is scant. Indeed, as noted Aidinlis (2022), in general terms, 'G2B (government to business) data sharing is an underexplored – compared to business-to-government (B2G) data sharing – emergent practice,' a fact which is also reflected in the relatively scant literature in this area. More generally, and again as noted by Aidlinis (2022), 'public-sector bodies generally 'lag behind' in developing and implementing data analytics as compared to the private sector'³ though there is increasing recognition that countries' sharing of health (personal) data in particular with the private sector, 'has become a component of many governments' health and economic strategies' (Braunack-Mayer et al, 2021). More generally, it should be

³ The focus of Aidinlis' review was (G2B) data sharing for research and evaluation purposes, with such data sharing being an, 'emergent practice of governments which are interested in exploring the potential of data analytics to produce more accurate and cost-efficient policies in such contexts as improving access to justice, reducing poverty or supporting healthcare systems' (Aidinlis (2022).

noted that for the purposes of this literature review, public sector sharing of anonymous (not pseudonymized) private data - which earlier was personal data - was considered in scope, largely because it was almost impossible to anonymise fully. Indeed, a theme explored in the literature was the 'fuzziness' that exists between different types of data (Borgesius, Gray and van Eechoud, 2015). In this regard, it was difficult to make an exact comparison like for like comparison between G2B and B2B data sharing.

2. The notion of the private sector

During our discussions and exploration of the literature, we questioned how to frame the private sector and whether the third sector (e.g. a charity or a University) should be included. We acknowledge that the boundaries between these entities may be blurred, since non-commercial and academic research centres can evolve into commercial entities, as well of course due to the existence of public-private enterprises. We therefore conceived of the private sector in broad terms. We do, however, understand that public sector personal data sharing among public actors (e.g. government to municipality) is not within the scope of this review, unless it is a matter of a public-private actor or this sharing also involves a private actor.

3. The notion of data value

During our initial discussions with the Commissioning team, we realised that the requirement to undertake research into existing understandings of public benefit in relation to the use of public sector personal data with or by the private sector left open the question of whether costs and benefits were to be understood as solely financial or whether they could be conceived of in broader, more social terms. Following discussions with the Commissioning team, we embraced an open understanding including the social (and indeed other possible) dimensions. The use with or by the private sector is generally for profit but can also be non-profit. In this case, personal data would be used for common good (that is, benefit). Our review tried to grasp the circularity of this as private benefit may also create utility/value for society. This understanding is reflected in the literature, as discussed below (see 'conceptions of value').

-

⁴ For discussion, see Comandè, G., & Schneider, G. 'It's time. Leveraging the GDPR to shift the balance towards research-friendly EU data spaces' (2022) 59 (3) Common Market Law Review, 739-776.

4. The notion of models

We do not only include only resources looking at 'real' models but also those that offer potential hypothetical/theoretical models. We also dedicated a specific case study section on models that are spontaneous, initiated by citizens that could be embraced by the government and use the data produced with the private sector.

5. Benefits vs. benefit-sharing

When we refer to benefits, we mean benefits from the sharing of public sector personal data with a third actor, and we differentiated this term from benefit-sharing. These are distinct but overlapping terms. Benefit is mostly linked to (data) value; benefit-sharing is also embedded in the mandate of the government, conceptions of justice and legitimacy. Both terms imply a question on equity: when we talk about benefit(s), we should wonder 'benefit(s) for whom?' if we really want to achieve an equitable benefit and data sharing. Benefits and benefit-sharing are notions often discussed within biodiversity and genetic resources governance. We tried to apply some of the relevant literature from these fields to personal data sharing, to the extent that such literature appeared in our literature search. However, it should be noted that these fields are often kept very separate in the literature and indeed in legal discussions.

6. Risk sharing

We questioned whether to include in our search not only benefit-sharing but also risk sharing, e.g. societal, financial and legal risks. Some of the resources surveyed touch upon the fact that with benefits also come risks, and where relevant, we stressed these findings in our thematic analysis.

7. A legal basis for sharing

We assumed that all the sharing of personal data by the public sector here discussed has a legal basis under the GDPR. Literature embracing ethical issues, however, pushed us to explore the fringes of what is permissible and what is ethical.

8. A glossary of key terms

We think it could be useful to jointly define an 'official understanding' of key terms for any institution wanting to navigate the field. Indeed, and as will be explored in our thematic analysis, many key terms lack definitions, and this could potentially cause issues for clarity in the debate. We also return to this point in our recommendations.

B. HIGHLIGHTS FROM THE THEMATIC ANALYSIS

In the below section, we set out the results of our thematic analysis of the reviewed literature. At the forefront of our development of the thematic analysis were the questions informing this study. Our thematic analysis is therefore structured under five central headings; 1) public benefit; 2) conceptions of value; 3) benefit-sharing; 4) a public-benefit duty to share; 5) public data governance models. We use this analysis as the basis for **section C**, which puts forward several potential principles for the (benefit) sharing of public sector personal information.

I. PUBLIC BENEFIT

The literature on the notion of **public benefit**, at least within the health and social care sector, notes that public benefit is very much a subjective concept, with a myriad of different terms attached to the concept. Indeed, numerous terms are often used interchangeably with that of public benefit (see e.g. Waind, 2020), reflecting its subjective quality, as well as the uncertainty that exists surrounding its usage. According to <u>Ballantyne and Schaefer (2020)</u>, common terms found in the bioethics literature and relevant research regulations include, 'public interest, public benefit, public good and social value. In addition, the following less common terms are also used: community benefit, common good, common interest, collective interests, collective good and social benefit.' Ballantyne and Schaefer (2020), argue that within the context of consent waivers and the sharing of data, '(t)he number of different terms gives greater scope for divergent interpretations of their meaning. Second, for the most part these regulatory and guidance documents do not define the terms. This lack of clarity creates uncertainty about what the public interests criteria require.' Cheung (2020), has highlighted the lack of a definition of public benefit, raising concerns over the 'exploitation' of the term in respect of the use of patient health data. She argues that the, 'ill-defined concept of public benefit enables the concept to be harnessed to facilitate widespread commercial access to patient health data, despite public concerns towards such involvement. As a result, this source of data is potentially subject to logics of data accumulation similar to those seen elsewhere in the digital economy.'

Attempts have been made in the literature to define public benefit as related to the public interest. Writing in the sphere of healthcare data, <u>Ballantyne and Schaefer (2020)</u> posit that public benefit is in fact a narrow subset of public interest; whereas public interest, 'requires consideration of the trade-offs between competing common goods', public benefit is the, 'additional benefit produced by research that would enhance the current knowledge or health status of a community.' They argue that, '(n)et public benefit is typically assessed by weighing anticipated research harms against expected benefit.' While their focus is admittedly narrow and focused on the use of data in the healthcare context, the notion of public benefit as requiring some

sort of additional benefit of use to the community is useful for this study. The common good, for <u>Ballantyne and Schaefer (2020)</u>, links to the economic concept of non-rivalrous public goods, while the social good is generally conceptualised as a 'necessary basic requirement of all human research. They argue that a lack of conceptual clarity over the use of such terms, which as noted above, are often used interchangeably, has the potential to lead to debate over the, 'normative principles of good data sharing.' Their preference is for use of the term 'public interest', at least in respect of consent waivers (see also <u>Schaefer et al, 2020</u> who define 'public interest' - in the context of consent waivers for health data and tissue research- as the '(s)ubstantial expected advancement of the health-related interests of members of a group whose interests are, or should be, of particular concern to the society in question').

In a related vein regarding the need for clarity, albeit from a slightly different context, Mészáros and Ho (2018) point out that there are, 'different levels of public interest in the GDPR e.g., general, substantial, important' and 'clarification of these different levels is crucial, since they have a significant effect on the data subjects' rights.' The need for definitional clarity has also been (implicitly) echoed by others in the literature with Waind, 2020, for example, opining that, 'no widely understood definition of the concept (of public interest) appears to have been identified, and perhaps what matters more than defining the term is that the public perceives benefits of some sort... (outlining) the potential benefits of the work therefore remains an important goal of public engagement.'

Where the literature has attempted to clarify or indeed quantify the meaning of public benefit, a theme identified in the literature was that it could include indirect benefits such as enhanced knowledge of a particular area, even if just to confirm that existing systems are working (Putting Good into Practice, 2021) but might also include more direct benefits such as furthering research on particular diseases or even providing an income stream for the National Health Service (Putting Good into Practice, 2021). In respect of the possibility of a direct income stream for the likes of the NHS, the notion of profit-share agreements or charges for access should private companies have access to publicly held personal (health) data was discussed in the literature as one potential way acceptable to citizens to secure public benefit (Putting Good into Practice, 2021). Hence, while public benefit would likely exclude 'pure' profit making motives (as discussed above), the involvement of the private sector was not in itself necessarily at odds with securing 'public benefit.' There were also certain prerequisites for achieving public benefit identified in the literature with the "Putting Good into Practice" public dialogue (2021) stressing that, 'transparency cannot be separated from public benefit... Public benefit is undermined if authentic public engagement is not integrated into data assessment. This requires engaging people from a cross-section of society in data assessment processes.' We might think of this as the procedural elements necessary to achieve public benefit. At the heart of these elements is a focus on trust and transparency in terms of how data is used (Putting

Good into Practice, 2021). In the same vein, a lack of trust in the private sector was traced in the literature to a reluctance by at least some citizenry to have their information shared with commercial actors (<u>Braunack-Mayer et al, 2021</u>).

Generally, in terms of Open Data by the governments, <u>Munné</u> (2016) expounds that indirect benefits might include;

- "The free flow of information from organizations to citizens promotes greater trust and transparency between citizens and government, in line with open data initiatives.
- Information from both traditional and new social media (websites, blogs, twitter feeds, etc.) can help policy makers to prioritize services and be aware of citizens' interests and opinions.
- Tailoring government services to individuals can increase effectiveness, efficiency, and citizen satisfaction.
- Correlation of multiple sources of data will help government economists with more accurate financial forecasts.
- Automated algorithms to analyse large datasets and integration of structured and unstructured data from social media and other sources will help them validate information or flag potential frauds.
- The public sector is increasingly characterized by applications that rely on sensor measurements of physical phenomena such as traffic volumes, environmental pollution, usage levels of waste containers, location of municipal vehicles, or detection of abnormal behaviour. The integrated analysis of these high volume and high velocity IoT data sources has the potential to significantly improve urban management and positively impact the safety and quality of life of its citizens.
- Collect, organize, and analyse vast amounts of data from government computer networks with sensitive data or critical services, to give cyber defenders greater ability to detect and counter malicious attacks."

A dichotomy was also noted in the literature between the framing of public benefit by, for example, governments, and how publics conceive of public benefit. In a study by <u>Aitken et al. (2018)</u>, which reports the result of a series of workshops in Scotland engaging with members of the public on how the public might benefit from 'data intensive' health research, what is notable is that, 'no one spoke of societal benefits in terms of economic benefit, although this is often portrayed as a form of public or societal benefit by governments and funding agencies.' Accordingly, there is a need to separate how public benefit is framed by government and related entities, and how the public conceives of public benefit (see also Public Engagement study).

At least some of the literature surveyed noted a link between privacy and the idea of public benefit. In this regard, a literature review undertaken by Hutchings et. al,

(2020) found something of a conflict between changing perceptions of the value of privacy with public benefit but also found that, 'respondents were positive about sharing their data for research (and) in some circumstances, societal benefit may outweigh concerns regarding privacy.'5 In this vein, the same literature review identified that, '(p)ublic benefit was seen as a justification for access to health data and an individual's right to privacy should not prevent research that could benefit the general public.' However, such a view was not universal and Hutchings et. al 2020 noted some contestation in the literature on this issue with some unwilling to 'trade' privacy for the public good. In a related sense, Ballantyne and Schaefer (2020), who prefer the term 'public interest,' contend that it refers to the 'overall assessment of the potential impact of the research of common goods such as privacy, trust, justice in addition to population health.' By extension, at least some of the literature focused on the circulatory of benefits; with the notion that a perception of 'benefits' from the sharing of personal data could bolster trust, which in itself is an important component of any decision to share particularly sensitive data, at least in the health sector (Jenkner et al. 2022). Accordingly, benefits could have an instrumental function in bolstering data sharing in the first instance. In a similar vein, for Ormondroyd et al., (2022), 'trustworthiness was considered to encompass ideas of public (including future public) benefit, evidence-based standards, governance, ⁶ transparency, consistency and communication.'

On the theme of trust, and while not directly within the questions posed in the background to this literature review, it is nonetheless relevant to set down the ambivalence - expressed within the literature - on the part of citizenry to the sharing of their personal information with the private sector. Much of the literature in this regard focused on the governmental sharing of health data, and reflected an 'ambivalence about the roles, motives, and actions of the private sector with respect to health data' and concerns about the sharing of such data by commercial 'for profit' entities (see e.g. Braunack-Mayer et al, 2021 and Hutchings et. al, 2020). However, ordinary people are supportive of health and social care data being used for public benefit (Putting Good into Practice, 2021) but that public benefit must outweigh private profits and interests. This was a point explored in detail in the 'Public Engagement' literature review.

A related theme identified in the literature is that of 'data altruism' and its link to conceptions of public interest and societal interest (Comandè and Scheider, 2022). The concept of data altruism, encapsulated within the <u>Data Governance Act</u>, reflects the open sharing of data for the purposes of the public interest, such as for research purposes or to improve public services. The term altruism is central here and the use

⁵ References in original omitted.

⁶ Reference in original omitted

of such data under the principle of data altruism must be for *non-profit purposes*. Clearly, while the principle of data altruism provides an interesting insight into concepts such as public interest, its focus is narrow, based as it is on non-profit use.

II. CONCEPTIONS OF VALUE

There is a vast literature on data value which sits outside the scope of this review. However, for the purposes of this review, a theme in the literature surveyed was that value cannot simply be defined in economic terms. This aligns to the discussion above regarding how public benefit cannot simply be measured in direct terms but can also include indirect benefits. Saheli Datta Burto et al in Rethinking value construction in biomedicine and healthcare, contend that, 'taking the meaning of value as given, or reverting to technocratic or economic dimensions of value, obscures the non-technical and societal dimensions of value construction and operationalisation.' They argue for an, 'understanding of the value construction processes that makes a thing valuable (or not) for society as the first step towards' what value 'ought to be.' While their focus is on value/valuation within the healthcare and biomedicine, the themes of value as something context specific (see also Liddell, Simon and Lucassen, 2021) and more than merely the outcome of a cost-benefit analysis is also one explored in other literature included within this study (see e.g. "Service-dominant logic: continuing the evolution").

The difficulties posed by conceptions of value as economic are also illustrated by the collapse of the care.data project, which throws light on the threat data sharing may pose to the social licence under which research within the NHS and other related organisations functions. Vezyridis and Timmons (2017), for example, note that the 'new context of information sharing that care.data ...initially implied was that patients should also contribute data for the country's economic prosperity, in the widest and most ambiguous sense of the term. This new, economically narrowed, value caused care.data to be perceived as a threat to a long established social licence for research within the NHS and healthcare.' The trend illustrated by care.data illustrates, for Vezyridis and Timmons (2017) a direction of travel, 'that aims at constituting sustainability, governance and economic growth as the de facto social values, while reducing privacy (as a reassurance of trust, exercise of choice and protection from exploitation) to an individualistic preference.'7 For the authors, this is not sustainable in the long term, and hence conceptions of social value that are largely equated with economic growth and 'assetisation' are to be resisted since it will threaten the, 'integrity of public healthcare..., nurturing more healthcare and informational challenges for it.' Leading from this, the literature notes the importance of public

_

⁷ Reference in original omitted.

engagement to ensure a social licence for research and indeed data sharing (<u>Laurie and Stevens</u>, 2016).

Even where a narrow 'economic' conception of value is accepted, and particularly in the case of personal data, the so-called 'real market value' of such data is usually derived from its use in large data sets, rather than that personal data having value (narrowly conceived) in its own right (see e.g. <u>Fischli, 2022</u>). In a similar respect, it has been noted in the literature that economic conceptions of data value often neglect to factor in the cost, including the infrastructural cost, of developing such data in the first place (<u>Vezyridis and Timmons, 2021</u>).

Framing was also seen as central within some of the literature to conceptions of value. Vezyridis and Timmons, 2021 describe 'the role of and expectations by state actors to drive and facilitate, via investment and regulatory frameworks, the 'assetisation' and capitalisation of NHS patient data for the benefit of the national economy.' This 'assetisation' of NHS patient data may foster innovation, but also play their own important role in how GPs, NHS patients and the data they coproduce are imagined and valued to perform for (electronic health record) datadriven research within a high stakes biomedical knowledge economy.' Furthermore, by controlling the, 'flows of data and assets for the extraction of rents may, in the end, benefit only those networked actors that have the sociomaterial resources (including financial capital) and knowledge expertise to capitalise on these (public) datasets for narrowly specified purposes' (Vezyridis and Timmons, 2021). In essence, the very 'assetisation' of personal data may in itself frame conceptions of value and result in, 'unequal configurations of access to knowledge production and public scrutiny' (Vezyridis and Timmons, 2021). Clearly, (data) value is not a neutral concept but instead is socially mediated and co-produced, with the process of 'assetisation' potentially shaping understandings of value. At the same time, and in a related sense, how data is regulated is also a function of a context specific 'sociotechnical settlement' with, 'data regulation (...) co-produced with a particular understanding of personal data as a political-economic asset' (Guay and Birch, 2022) These understandings of the 'value' of personal data are jurisdiction specific, and informed by the underlying context and socio-technical settlement.

In the literature on services and service provision, one aspect that can be relevant for the present focus is that of service science embracing value co-creation. The article "Service-dominant logic: continuing the evolution" (Journal of the Academy of Marketing Science, 2008) advances the argument that service science should take the perspective of value co-creation and exchange beyond the market by providing a systems orientation that takes the issues out of the economic arena and recontextualizing them. The study offers an interesting take on 'the role of exchange between and among service systems at different levels of analysis (e.g., individuals, organizations, social units, nations, etc.), thus enriching marketing in ways that are

difficult from its usual tighter, enterprise, economic, and normative focus, even when enhanced through service-dominant logic.'

The article <u>Service Innovation:</u> A <u>Service-Dominant Logic Perspective</u> (MIS Quarterly, 2015) offers a conceptualization of service innovation grounded in service-dominant logic that transcends the producer-consumer divide. It emphasizes "(1) innovation as a collaborative process occurring in an actor-to-actor network, (2) service as the application of specialized competences for the benefit of another actor or the self and as the basis of all exchange, (..) and (4) resource integration as the fundamental way to innovate." Building on these core themes, the authors develop a tripartite framework of service innovation based on "(1) service ecosystems, that actors create and recreate through their effectual actions and which offer an organizing logic for the actors to exchange service and cocreate value; (2) service platforms, which enhance the efficiency and effectiveness of service exchange; and (3) value cocreation, which views value as cocreated by the service offer(er) and the service beneficiary (e.g., customer) through resource integration and indicate the need for mechanisms to support the underlying roles and processes." Especially point 3 can be valuable for the present review.

On the topic of co-creating value(s) with the market and civil society from a policy-maker perspective, the JRC report "Values and Identities - a policymaker's guide" can be particularly valuable as it presents current Europe-wise scientific knowledge on values and identities, and co-creation thereof at an interdisciplinary level. The report offers important insights (complemented by findings from a dedicated Eurobarometer) for policymakers to embrace values in addressing challenges of our time, including a dedicated toolbox section on co-creative approaches.

III. BENEFIT-SHARING

As we note above, benefit-sharing is a concept typically associated with international environmental law and in particular, international biodiversity law via the United Nations Convention on Biological Diversity (CBD) and its associated Nagoya Protocol. Regarding benefit-sharing under the CBD, the 'justification for benefit-sharing according to the CBD relies on a mutually beneficial instrumental approach. In Aristotelian terms, we are dealing with "commutative justice", where each party gives one thing and receives another, with a focus on the equivalence of the exchange' (Schroeder, 2007).

In the slightly different context of the sharing of benefits derived from human genetic resources, for Schroeder (2007), such 'sharing is the action of giving a portion of advantages/profits derived from the use of human genetic resources to the resource providers in order to achieve justice in exchange with particular emphasis on the clear provision of benefits to those who may.' Schroeder's focus in creating this

definition is on the imbalance that may often exist between data providers, and those undertaking research into human genetic information. While Schroeder's focus is on human genetic information, the notion of benefit-sharing as a form of justice necessary to address an (im)balance between those whose private information is used, and those who may subsequently benefit from it, is an implicit theme within the literature, linked as it is to related concepts such as the social licence to undertake research (see discussion above under 'conceptions of value). In a literature review of data sharing (in respect of African biobanks), the theme of justice was again made explicit in respect of benefit-sharing with Igumbor et al, 2021 noting the importance of the 'principle of distributive justice by optimising benefits to society, minimising harm and equitable beneficence related to accessing data and emergent health innovation.'

Related to the above, concerns have been raised in the literature as to the impact of the sharing of personal data upon the more vulnerable in society (see e.g. <u>Cheung (2020)</u>. As expounded by <u>Tzanou (2022)</u>, '(the) poorer and the socioeconomically disadvantaged in the society face pervasive forms of modern servitude—amounting to domination—as the design of the digital welfare state and public—private entanglements leave them to the mercy of private tech companies and technological tools that are not developed with their interests and needs in mind.' Accordingly, considering the impacts upon the disadvantaged in society, and the use to which personal information will be put, is an essential element of thinking upon the justice dimensions of such sharing.'

The notion of addressing an imbalance and its relation to benefit-sharing also finds expression in the literature on the application of property rights over personal data. At least some of the analysis in the literature is structured in general terms, and in part relates to arguments that the application of property rights to (personal) data may enhance economic efficiency. Liddell, Simon and Lucassen, 2021 discussed above, note the work of Purtova 2015 who argues that, as 'long as personal data bears high economic value, the real question is not whether there should be property rights in personal data but whose rights they should be.' In essence, ownership by patients of their personal healthcare data might ward off the 'hungry' corporation (Liddell, Simon and Lucassen, 2021 drawing on Purtova 2015). As noted by Liddell, Simon and Lucassen, 2021, the argument for property rights over personal information links in part to the 'demand' for benefits and moreover benefit-sharing regarding the use by the private sector of personal data, particularly in the health sphere. However, they reject the notion that property rights would provide any greater protection than the current GDPR based system offers.

IV. 'PUBLIC BENEFIT' DUTY TO SHARE?

In the <u>Putting Good into Practice</u> study discussed above, participants noted that if data was truly important, there was by extension a duty on governments to make use of such data sources. In essence, no public benefit can ever exist in the absence of the sharing of data, and so by extension, some participants in the <u>Putting Good into Practice</u> study felt that there was a duty on governments to use the data at issue (in this case, health and social care data). However, it has been noted that in the UK at least, there is an aversion to share on the part of public authorities, in part due to confusion around legal requirements (<u>Laurie and Stevens, 2016</u>).

The article "Reaping the benefits of Open Data in public health" (Can Commun Dis Rep, 2019) makes the related point that "Open Data represents a fundamental and massive shift in how we conduct research, make decisions, develop policy and evaluate our interventions. There is increasing pressure and expectation by the public for researchers and governments to show and share the data and information that public funds have generated. The potential benefits of making data open and accessible are very exciting; however, the challenges in making this happen are substantial and should not be underestimated." The article explores where we are in terms of addressing the challenges and reaping the benefits of Open Data in public health. The study argues that "With respect to the technological shift, there has been a lot of progress, but appropriate technology and infrastructure is still being developed at all levels of government. Some areas of public health science, such as bioinformatics, are well ahead in current activities and in future planning for technologies and infrastructures. Other areas are less well developed. In addition, a socio-cultural shift is still underway and there remain those who are still hesitant to share their data."

In the book chapter <u>Big Data in the Public Sector</u> (2016), the author notes that "The public sector is becoming increasingly aware of the potential value to be gained from big data. Governments generate and collect vast quantities of data through their everyday activities, such as managing pensions and allowance payments, tax collection, national health systems, recording traffic data, and issuing official documents." The contribution "takes into account current socio-economic and technological trends, including boosting productivity in an environment with significant budgetary constraints, the increasing demand for medical and social services, and standardization and interoperability as important requirements for public sector technologies and applications."

V. PUBLIC DATA GOVERNANCE MODELS

Another theme identified within the literature is that of governance. The article "Emerging models of data governance in the age of datafication" (Big Data & Society, 2020) examines four models of data governance emerging in today's societal dynamics. It argues that "while major attention is currently given to the dominant model of corporate platforms collecting and economically exploiting massive amounts of personal data, other actors, such as small businesses, public bodies and civic society, take also part in data governance." The article - despite not primarily discussing G2B data sharing - offers useful insights on four models emerging from the practices of these actors: data sharing pools, data cooperatives, public data trusts (on data trusts more generally, see Rinkin, 20) and personal data sovereignty. The study overall proposes a social science-informed conceptualisation of data governance. Drawing from the notion of data infrastructure, the authors identify the models as a function of the stakeholders' roles, their interrelationships, articulations of value, and governance principles. This conceptualisation brings to the forefront the power relations and multifaceted economic and social interactions within data governance models emerging in an environment mainly dominated by corporate actors. The said models "highlight that civic society and public bodies are key actors for democratising data governance and redistributing value produced through data."

A PhD thesis of 2017 addressing the legal problems coming from mashups of Open Government Data (OGD), drafted an informal ontology to help technical 'reusers' of Public Sector Information to utilise datasets according to their intended purpose and in compliance with the legal obligations that govern the rights to reuse the data. A survey of national OGD portals found that the majority of OGD are released under inappropriate licenses, not fully complying with the legal rules that apply to the reuse of the data. Open Government Data can be released and covered by multiple licensing regimes, up to 33 in a single country. The thesis analysed the European Union (EU) legal framework of reuse of Public Sector Information (PSI), the EU Database Directive and copyright framework and other legal sources (e.g., licenses, legal notices, and terms of use) that can apply to open government Datasets. The thesis can be useful for the drafted Informal Ontology of Open Government Data Licenses Framework for a Mash-up Model (iOGDL4M). The iOGDL4M aims to connect each applicable legal rule to official legal texts in order to direct legal experts and actors interested in data reuse to primary sources.

C. POTENTIAL OVERARCHING PRINCIPLES FOR THE (BENEFIT) SHARING OF PUBLIC SECTOR PERSONAL DATA

In this section, we build upon the above thematic analysis of the identified literature by setting out potential overarching principles for the (benefit) sharing of public sector private information. From the review performed, we identified several guiding principles that could be of help for designing an appropriate framework for public sector personal data (benefit) sharing. In doing so, we aim to address the key questions which informed this study.

We present the overarching principles for the sharing of public sector private information via a multi-level approach, exploring guiding principles developed by international organisations, the EU and a range of national jurisdictions which featured in our literature review.

- AT A GLOBAL LEVEL

Proportionality and respect for ethical values and norms

The Organisation for Economic Cooperation and Development (OECD) contributed to the debate with Recommendation on Enhancing Access to and Sharing of Data (2022). The Recommendation - that includes G2B - "sets out general principles and policy guidance on how governments can maximise the benefits of enhancing data access and sharing arrangements while protecting individuals' and organisations' rights and taking into account other legitimate interests and objectives. They explicitly encourage data access and sharing arrangements that ensure that data (including public sector personal data, duly anonymized) are as open as possible to maximise their benefits and as closed as necessary to protect legitimate public and private interests, including interests related to national security, law enforcement, privacy and personal data protection, and intellectual property rights as well as ethical values and norms such as fairness, human dignity, autonomy, self-determination, and the protection against undue bias and discrimination between individuals or social groups."

Comprehensive regulatory frameworks

The OECD "Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies" (2019) can also be considered a key

guidance as it "explores the possibilities available to policy makers and business leaders when aiming at establishing data-governance frameworks that do enough justice to important specificities but are comprehensive enough to be coherently applicable across application areas."

Remarkably, the OECD Committee on Digital Economy Policy <u>issued a statement</u> (2020) here it "raised concerns that the absence of common principles for trusted government access to personal data may lead to undue restrictions on data flows resulting in detrimental economic impacts. The Committee concluded that working toward trusted government access to personal data held by the private sector is an urgent priority requiring further international collaboration. The Committee noted the OECD's strength as a forum to foster discussion and collaboration among likeminded countries."

Whereas the focus is again on B2G, useful lessons for the reverse data flow can also be extracted from the OECD advanced need to develop "an instrument setting out high-level principles or policy guidance for trusted government access to personal data held by the private sector. (...) These may include safeguards relating to: the legal bases upon which governments may compel access to personal data; requirements that access meet legitimate aims and be carried out in a necessary and proportionate manner; transparency; approvals for and constraints placed on government access; limitations on handling of personal data acquired, including confidentiality, integrity and availability safeguards; independent oversight; and effective redress."

Relevant principles also come from the Research Data Alliance (RDA), launched as a community-driven initiative in 2013 by the European Commission, the United States Government's National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation with the goal of building the social and technical infrastructure to enable open sharing and re-use of data. We highlight the RDA Principles on the Legal Interoperability of Research Data (2016). The Principles focus on all types of data (including public sector personal data when they are shared with the private sector - G2B focus) that are used primarily in publicly funded research in government and academia. These principles include:

- "Facilitate the lawful access to and reuse of research data.
- Determine the rights to and responsibilities for the data.
- Balance the legal interests.
- State the rights transparently and clearly.
- Promote the harmonization of rights in research data.

Provide proper attribution and credit for research data."

- EU WIDE

Emphasis upon increasing high-quality data now available... though not always G2B

We found useful guidance in the EU-wide process leading to a new Directive on Open Data and Public Sector Information (PSI) (2019) - which can be for example "anything from anonymised personal data on household energy use to general information about national education or literacy levels" to be shared also with private actors (G2B). The process - in full compliance with the EU General Data Protection Regulation (GDPR) and the European Data Governance Act - "updates the framework setting out the conditions under which public sector data should be made available for reuse, with a particular focus on the increasing amounts of high-value data that is now available."

Recent updates of the <u>European Strategy for Data</u>, including the recent <u>Data Act</u> (2022), foresees "means for public sector bodies to access and use data held by the private sector that is necessary for specific public interest purposes. For instance, to develop insights to respond quickly and securely to a public emergency, while minimising the burden on businesses." However, here the focus seems to be shifting to B2G again (a trend more prominent in the literature reviewed than G2B) and not vice-versa as the focus of our review.

Legal certainty

In the <u>Science Europe response</u> (2021) to the European Commission's consultation on the future Data Act, other relevant guidelines - but also gaps - are set out: "Access to and use of privately held data by the public sector should benefit society as a whole. Clear rules and transparent information on reliable data sharing are crucial for all data holders and users to operate in full legal certainty. In addition to clear rules, it is also important to emphasise the soft positive effects that private-to-public data sharing can have for private entities, such as enhancing the entity's public reputation. Further incentives for the private sector to share their data are needed. These need to be carefully considered and supported by evidence and related analytical studies." Here the focus is mostly on B2G data sharing.

However, the response continues, "The information on the actual scope of the envisaged initiatives to the benefit of the public sector, as currently provided by the

European Commission, is still confusing. In its considerations, the Commission alternates between referring to the public sector in general and to business-to-government (B2G) data sharing in particular. It needs to be clarified whether the Commission intends to limit its proposal to governments only, or whether the public sector at large – including public research funders and research performing organisations – should benefit from it." A lack of clarity in G2B and B2G specific differentiations therefore arises.

Moving to the prominent health data-focus, the <u>Communication from the Commission COM(2022) 196/2</u> - titled "A European Health Data Space: harnessing the power of health data for people, patients and innovation" - argues that the European Health Data Space will "have a significantly positive impact on fundamental rights as regards personal data protection and free movement. Properly articulated with the European Open Science Cloud (EOSC) data space and the relevant European life sciences data infrastructures, it will enable researchers, innovators and policy-makers to more effectively use the data securely and in a way that safeguards privacy."

Along these lines, TEHDAS (The Joint Action Towards the European Health Data Space) helps EU Member States, associated countries, and the European Commission "to develop and promote concepts for the secondary use of health data to benefit public health and health research and innovation in Europe. TEHDAS consists of eight work packages including a specific work package on 'Sharing Health Data'. The overall aim of the Sharing Health Data work package is to provide options for the operational framework and governance models for the exchange and secondary use of health data between European countries, respecting the principles of transparency, trust, FAIRness, citizen empowerment and common good." A TEHDAS study identifies "as a barrier the non standardised data sharing agreements for products developed by private sector providers using public sector health data in order to (a) facilitate safe data sharing and (b) protect taxpayers' investment. Under this barrier, stakeholders highlight the importance of establishing rules and guidelines for equal access to health data for the public and private sectors."

The Analytical Report: Business-to-Government Data Sharing (2020) prepared for the European Data Portal posits (again with a more prominent B2G focus) that "reusing relevant privately held data increases the public sector's ability to understand, assess and predict different situations and phenomena that affect the citizens. It enables more logical and fact-based decisions, and at a higher pace. The benefits from data re-use are not only reserved to the private sector. In fact, to become more cost-efficient and to provide effective services for citizens, public sector bodies can benefit greatly from data sharing and need to exploit the potential of new data sources." The document also provides a relevant discussion of models and examples from real world case studies.

- COUNTRY-SPECIFIC INSIGHTS

I. THE UNITED KINGDOM (UK)

Ensuring privacy, clarity, and consistency

The UK Information Communication Office's "<u>Data sharing across the public sector:</u> the <u>Digital Economy Act codes</u>" discusses a framework for sharing personal data, for defined purposes across specific parts of the public sector, under the <u>Digital Economy Act 2017 (DEA)</u>. The aim is to improve public services through the better use of data, while ensuring privacy, clarity, and consistency in how the public sector shares data (this includes our focus, G2B).

The <u>UK Government Licensing Framework</u> (UKGLF) of 2016 provides a policy and legal overview of the arrangements for licensing the use and re-use of public sector information both in central government and the wider public sector. It sets out best practice and standardises licensing principles. These are:

- "Simplicity of expression the terms should be expressed in such a way that everyone can understand them easily;
- Non-exclusivity so that access can be provided to a range of users on fair and equal terms; fairness of terms;
- Non-discrimination terms are extended fairly to all for similar uses;
- The need for acknowledgment and attribution;
- The need for transparency by publishing standard licence terms."

Focusing on the very dominant field of public sector private data sharing in the health, care and medical sector, the <u>"Putting Good into Practice"</u> public dialogue (2021), run by the National Data Guardian for Health and Social Care, and UK Research and Innovation's Sciencewise programme, demonstrated that ordinary people are supportive of health and social care data being used for public benefit. The findings of the study will inform policy advice or guidance by the National Data Guardian (NDG), and can be summarised as follows:

• "Prerequisites for public benefit

Transparency cannot be separated from public benefit. It is not an add-on or nice to have. Health and social care data use requests only demonstrate public benefit if

they have integrated communications within their application including activity which demonstrates the value of data use to society. To demonstrate public benefit, transparency is required throughout the whole data life cycle (collection, storage, assessment and use), not just at the point of application. Public benefit is undermined if authentic public engagement is not integrated into data assessment. This requires engaging people from a cross-section of society in data assessment processes.

• Areas that matter most to dialogue participants

Equitable distribution of benefits of data use in health and social care with safeguards to protect against discrimination and geographic disparities. Identifiable and sensitive data should be treated with the utmost care, if it is, it has the potential to bring public benefit. Data was perceived as being particularly sensitive if it is of a personal nature, such as genomics or mental health data, or because greater care is needed in its interpretation, such as qualitative data. Safeguards and provisions in place to protect society from data manipulation, where the outputs from the data use could be interpreted in different ways, for example, to achieve political or financial ends. This includes publication of statements of data users' credentials and sources of funding.

• Public benefit must outweigh profit with profitable uses of data rigorously scrutinised for demonstrations of public benefit before access is granted

There is a recognition that data use in this context can enable health and social care improvements and innovations. Being ambitious for health and care data use - to realise public benefit from global collaboration; exploratory research driving breakthroughs; and using profit for new developments, such as drugs, treatments and services."

A 2018 report by the organisation 'Involve', UK's leading public participation charity, the 'Carnegie UK Trust' and 'Understanding Patient Data' offers other relevant findings. They organised a series of workshops in diverse local authority areas across England during the summer of 2017. These workshops brought together professionals from the public and voluntary sectors to explore "how they collectively understood, defined and valued the public benefits that may be delivered by the use of personal data about service users and the wider public. The purpose was to begin to make sense of where an acceptable balance may lie between the risks and benefits of data sharing and use in the context of public service provision." The report is informed by the findings from these workshops, alongside information drawn from a review of relevant literature on public attitudes to data sharing. The report "establishes a proposed framework for service providers, across the public and voluntary sectors, to both evaluate the public benefits that the better use of data

may support; and assess this potential benefit against the risks that sharing data may entail."

II. AUSTRALIA

The Australian Government through the <u>Data Availability and Transparency Act</u> of 2022 (DAT Act) shows to be committed to modernising how public sector data is used. It is working to "unlock its potential safely and in line with community expectations. As a national resource, public sector data can benefit all citizens through better and more targeted government policies, programs and service delivery, and improved research to address real problems. The Law, among the others, regulates authorisation regimes for data custodians to share public sector data and for accredited users to collect and use such data."

An <u>article by Australian independent law firm CORRS</u> discusses the DAT Act of 2022. The DAT Act allows data created, collected or held by a Commonwealth government body (known as 'public sector data') to be shared with other Australian government departments and Australian universities. The scope of the DAT Act has been reduced in that it no longer allows Commonwealth bodies to share data with private sector organisations. CORSS argues "While the private sector cannot currently receive data through this Scheme, the Act's Revised Explanatory Memorandum states that the reason for their exclusion is to allow the DAT Scheme to 'establish and mature'. The Act provides that the DAT laws will be reviewed in three years. It also has a five-year sunset clause. It may be that following further review, the DAT Act will eventually be expanded to allow private sector organisations to receive public sector data."

The article also points to the fact that "The requirements and obligations placed on public sector recipients under the DAT Scheme are helpful indicators as to what responsible data sharing in Australia will look like going forward. Already, data ethics are becoming a fundamental business consideration when organisations decide how to collect, use and disclose information. Customer and individuals' expectations about how businesses use and protect personal information are also increasing. CORSS anticipates that businesses' responsibility to deal with data ethically and transparently will become even more important and that lessons can be learnt from the DAT Act." The article outlines that - although the Act currently operates for public sector bodies - it includes certain data governance practices for the private sector to consider in the event that the data sharing scheme in the DAT Act is extended to the private sector.

According to <u>an earlier review by CORRS</u> (2021), the Australian DAT Bill would have invoked the 'required or authorised by law' exception to Australian Privacy Principles 3 and 6 to permit the collection, use and disclosure of personal information by the public sector. Nonetheless, the DAT Bill will continue to work in conjunction with the

Privacy Act 1988 (Privacy Act) to protect individuals' personal information. The text argues: "The privacy risks introduced by the DAT Bill may be heightened by the fact that the data-sharing scheme relates exclusively to government-held personal information, which is often collected on a compulsory basis to enable individuals to receive a public service or benefit. This data is often, or becomes, sensitive when it is linked with other government data sets. However, stakeholders who stand to benefit from the proposed data sharing laws, including those in the health and medical research sector, consider that the DAT Bill strikes an appropriate balance between providing an effective mechanism for utilising data for research purposes and mitigating privacy risks."

In a 2022 commentary on the Australian and New Zealand Journal of Public Health titled "Health and public sector data sharing requires social licence negotiations", the authors argue that, over the past two decades, Australian governments and researchers have invested in building the infrastructures and legal frameworks to enable public sector data linkage. For public health researchers, "securely and safely shared anonymised population-based data – as enabled by data assets such as the Australian Bureau of Statistics' MADIP – is akin to a collective resource managed as a public good." The commentary makes the point arguing that "Among its diverse applications (e.g. health systems efficiencies, treatment innovations and medicine safety monitoring), the analysis of linked public sector data (such as health, disability, welfare, taxation, education, census) offers a powerful new lens into the social and political drivers of health. Although the public good that can ensue from public sector data seems clear, it is not inevitable."

The authors also stress the economic value of public sector data which "are also envisaged as an economic asset, as illustrated by the 2021 National Data Strategy and the Productivity Commission's Data Availability and Use report. Beyond advancing healthcare and medical research, public sector data integration can drive social transformations, such as economic growth". However, this can happen "with or without redistribution of benefits arising from a public asset, over-surveillance of or poor policy pertaining to already disadvantaged or socially excluded groups."

The commentary interestingly touches upon Indigenous Data Sovereignty and broader civic agency in the process, suggesting that "genuine Indigenous ownership of data governance matters for shaping the outcomes of data integration. As emphasised in the Australian Medical Association's 2019 submission to the Office of the National Data Commissioner (ONDC), the specific details of both legislation and professional practices that regulate data infrastructures matter. So too does public engagement. (...) without broad, sustained public dialogue, it is difficult to gauge public awareness, concerns or hopes regarding data integration."

The authors of the commentary argue that "building a social license will be fundamental to ensuring that data integration operates as a public good". They conceptualise it as follows "While a social license has been used and conceptualised

in different ways, there is agreement that it is distinct from legal or economic legitimacy and that it foregrounds the – often contested – values that publics bring to social changes which may ensue from new developments. Moreover, there is clear evidence that social licences granted by publics are effective in supporting and shaping technological developments when they are co-produced through an ongoing process of public engagement, dialogue and negotiation. Neither passive public acceptance nor closed discussions with select community representatives indicate the existence or form of a social licence. (...) advancing public sector data integration without establishing a process for gauging and calibrating data integration development with its social licence [may] undermin[e] public trust in the government agencies and health services involved in collecting, curating and sharing data."

III. THE UNITED STATES (U.S)

A 2017 study titled "Accelerating the Sharing of Data Across Sectors to Advance the Common Good" by the Beeck Center for Social Impact + Innovation at Georgetown University (U.S.) offers useful insights for G2B data sharing. The study encourages governments "to lead and show itself willing to share far more data with others, enticing them to share their own data in return." It argues that "even when government does share microdata, it often only shares with academics and not the rest of the private sector. It argues that governments must re-evaluate laws that allow no or excessively narrow sharing of their data, and must embrace new technologies that could allow the private sector to get at least some value from many sensitive data sets."

IV. THE NETHERLANDS

A study by the Dutch Institute for Information Law (IVIR) of 2008 argues that "Creative Commons model seems an attractive instrument for public sector bodies that seek to enhance transparent access to their information, be it for purposes of democratic accountability or re-use for economic or other uses." The study examines that hypothesis and highlights the major opportunities and pitfalls of the Creative Commons model for public sector information. It covers the status of government information under copyright law (a needed analysis because the use of the creative commons model presupposes that there is copyright in the licensed information); the relationship between freedom of information principles as enshrined notably in the Dutch Freedom of information Act or Government Information Act – the Wet Openbaarheid van Bestuur- and the copyright prerogatives as exercised in the various Creative Commons licenses (the authors examined whether the use of the Creative Commons model, down to the individual license terms, is compatible with statutory rights to access public sector information); the relationship between the legal framework for the (commercial) re-use of public sector information, also as regards potential unfair competition by the public sector in information markets (the

authors reviewed the compatibility of Creative Commons licensing with the primary law in this area: the Public Sector Information Directive).

D. CASE STUDIES INSIGHTS

In this section, we elaborate upon a number of key case studies of B2G, G2B, C2B as well as multilateral benefit-sharing which may be useful and relevant for the purposes of this study.

B2G and G2B relevant experiences

In a Comment titled "Social license for the use of big data in the COVID-19 era" (Digital Medicine, 2020), the authors propose trust-enhancing governance practices that can increase trustworthiness in data gathering and sharing for public health purposes, drawing on experience in Canada, the United Kingdom and the United States. The authors argue: "Strategies to enable the reopening of businesses and schools in countries emerging from social-distancing measures revolve around knowledge of who has COVID-19 or is displaying recognized symptoms, the people with whom they have had physical contact, and which groups are most likely to experience adverse outcomes. Efforts to clarify these issues are drawing on the collection and use of large datasets about peoples' movements and their health." Based on the case studies, the authors stress the importance of "earning social license for public approval of big data initiatives, and specify principles of data law and data governance practices that can promote social license."

In a 2020 Commentary "How to use data for the public interest, even - or especially - in a pandemic" by the Heinrich Böll Foundation, it is noted that "many associate the 'common good' only with non-commercial activity and demand that research and civil society be given more access to data and that the state make better use of its data. While these actors explicitly serve the public interest, it does not mean that common good cannot also be produced elsewhere. Commercial players, in particular, should not be rashly excluded, as they often generate common good, even if it is just a by-product – the benefit of a service or good beyond its commercial price tag. Such a rather open notion of what constitutes common good offers the advantage that, as a starting point, all interests are considered, for the debate about what benefits society and what furthers the common good is often guided by different, sometimes conflicting values." The commentary offers examples demonstrating how social and economic interests are not mutually exclusive:

"Platforms for the exchange of medical research data, such as YODA also include corporate partners whose interest is to bring profitable new drugs to market. A faster (while equally safe) development of therapies is certainly also in the public interest. YODA has already helped calculate the effectiveness of individual drugs or improve models for calculating healing prospects.

(...) Many pharmaceutical companies drive current efforts to find and produce a vaccine for the coronavirus that causes Covid-19, as they anticipate huge global

demand and receive government financial support such as advance purchase guarantees to offset some of the costs."

It concludes: "While weighing the non-monetary and economic impacts of data-driven services is a complex task, it is necessary so that policymakers and regulators can make sound decisions on whether or how to impose stricter regulations or even bans on certain forms of data collection and use. Thus, a broader notion of societal benefits supports allowing a multitude of actors to promote data-driven common good, including civil society and researchers. Companies also often create considerable value via data that they do not (and cannot) fully exploit themselves. Therefore, commercial actors should not be prematurely excluded."

In a similar vein to the above warning regarding not prematurely excluding commercial actors, in a 2022 chapter by <u>Aidinlis</u>, the author expresses concern regarding the possible narrow construction of the public interest under Article 6 GDPR. They argue that instead of 'starting an enquiry about the 'public interest' from the organisational identity of the data sharing participants, one should start from the potential contribution of a data sharing partnership to the collective economic or societal welfare.' The author argues that 'public interest' might be synonymous with a contribution to a 'substantial improvement of social welfare; and this can include the furtherance of 'private interests'. In essence, '(in lieu of the organisational interests of particular public bodies, the 'public interest' is to be construed by reference to such proxies as established indicators of collective societal and economic welfare.'

In the article "Towards a Paradigm Shift in Governing Data Access and Related Intellectual Property Rights in Big Data and Health-Related Research" (International Review of Intellectual Property and Competition Law, 2019) offers an interesting (failed) case in terms of G2B data sharing. First the authors argue that "Due to claims of ownership over data, sharing and re-use of data may be restricted entirely or privileged access may be granted for a fee, or small data sets may be offered to university-based researchers. Such practices deepen inequalities based on privileged access, mostly because data-owning companies have total control over data and no responsibility to make their data available, nor accountability to data subjects to ensure that their data are used in a manner that does not lead to harm." They then introduce the case study "The ethical and governance challenges that beset Iceland in 1998 are very instructive in this regard. Serious issues arose from the declaration of health records, which included health, genetic and genealogical data, as a national resource that was owned by the Icelandic government and could be made available to private industry without the consent of the individuals. As a result of national and international opposition to the inappropriate manner in which the Icelandic government handled the issue of ownership of data, the project collapsed in 2003". In the 2013 book "Benefit-sharing", the chapter on "Donating Human Samples: Who Benefits? Cases from Iceland, Kenya and Indonesia"

discusses three cases, including the above mentioned Icelandic example (the Icelandic deCODE biobank for genetic research); the sex workers from Nairobi, Kenya, whose samples are used for ongoing HIV/AIDS research; and the Indonesian government's decision to withhold virus samples from the World Health Organization in order to achieve fairer benefit-sharing. The authors note that "A framework for equitable access to human genetic resources is urgently needed, but in order to ensure justice, this needs to be accompanied by sustained attention to benefit-sharing."

In a 2020 OECD study titled "Enhancing data access, sharing and re-use" it is provided - among the others - an overview of government initiatives to facilitate data access and sharing, including across borders. It is noted that many of these initiatives also aim to address the challenges associated with protection of privacy, intellectual property rights and data control. All surveyed countries had "initiatives that foster and enhance access to and sharing of public sector data in 2018. However, significantly fewer countries targeted private-sector data. Even fewer governments had initiatives to improve the capacity to analyse data in their countries. Of 205 policy initiatives across 37 countries, 61% aimed at enhancing access to public sector data, while 21% aimed to help share private-sector data."

The cited 2015 OECD report presents the results of the review of the OECD Council Recommendation for Enhanced Access and More effective Use of Public Sector Information (PSI). The review is based on the analyses of the information gathered through a survey of PSI strategies in 20 countries as well as the European Commission. In doing so, the report illustrates different strategic approaches to PSI policies. It includes the case of Estonia where all information in the public sector is public and accessible free of charge by default according to the Public Information Act of 2001. There are generally no distinctions or limitations between different uses. Formal actions with legal justifications are necessary to limit access to information. Due to public interest, some personal data is public, e.g. salaries of civil servants. Although data is generally free, there are a few exceptions where charges apply, e.g., detailed queries from the business register and real estate register, where basic data is free. There is also a fee for full copies of detailed maps; this is almost the only case of a difference in price and conditions for commercial reuse. PSI and open data issues are part of the Estonian Information Society Strategy 2020. The strategy currently concentrates on making public data available in better machine-readable formats and encouraging the use of PSI by businesses and civil society.

- C2G potentially relevant cases

Although not falling directly within the focus of G2B personal data sharing, we consider it worth to take stock of multiplying examples of grassroots-led civic responses that specifically respond to targeted needs for public (health) services producing data from 'below' which can then be shared with competent authorities (C2G). We consider these experiences relevant as (1) some of these initiatives entail

the handling of personal data and (2) the data received by governments through C2G data sharing approaches could in turn be shared by the governments with the private sector (G2B, that is our focus).

The identified initiatives from a rapid landscape analysis (not to be considered representative of the multifaceted initiatives in the field) have different levels of:

- Civic engagement: they can be more or less controlled by organized and unorganized civil society, e.g. often coupled with a role of universities and research centres.
- Needs for services: they may engage with risks that are unexpected, for example the pandemic or a natural disaster (sudden crises), or with persistent risks such as cancer or Alzheimer (protracted crises, although the first type of crises can evolve in the second type as we see with Covid-19).

Among the relevant cases, the recent Covid-19 pandemic gave new momentum to the already existing Folding@Home initiative, "a distributed computing project for simulating protein dynamics, including the process of protein folding and the movements of proteins implicated in a variety of diseases". The initiative "brings together citizen scientists who volunteer to run simulations of protein dynamics on their personal computers and scientific researchers. Insights from these data are helping scientists to better understand biology and providing new opportunities for developing therapeutics". In the past, the initiative also tackled other crises generated by infectious diseases like Ebola, but also addressed persistent crises, working for example on kidney cancer and epigenetic cancer targets, and on neurological diseases such as the Parkinson's disease. We can posit that, while providing an almost technical support to the initiative, participants also expect a certain degree of benefit-sharing associated with the data they contribute.

Another crises-driven civic initiative for providing (also) public health services is **MapSwipe**, an open-source project that aims at helping first responders working with communities affected by disasters, disease and conflicts with data that are needed to provide relief. In particular, as first responded have to "cover large areas, but lack the data necessary for an efficient, effective response", the initiative offers an app to volunteers to "pinpoint where critical infrastructure and populations are located, allowing mappers to focus only on areas where they know features need to be mapped". The initiative claims to help organisations coordinate humanitarian efforts, offer services and even save lives. We can see in this data contribution a precious resource for the public sector.

On the persistent crises' side of the spectrum, and as a highly distributed network, the project **Stall Catchers** was initiated as an online game open to everyone, involving thousands of registered players worldwide, with the aim of speeding up Alzheimer's disease research ongoing at Cornell University, U.S. The initiative argues to "reduce the time to find an Alzheimer's treatment from decades to just a few years". The argument for engaging such distributed intelligence to tackle the

disease is the following: in order to identify an Alzheimer's treatment, "there is still a lot of data to be analysed. But the data analysis is extremely time-consuming and beyond the reach of today's best computer algorithms. It could take decades to find a viable drug target if the researchers work on it on their own. With the help of citizen scientists playing Stall Catchers, we could reduce this analysis time to just a couple of years!".

Lastly, still responding to slow-onset crises, the **Genigma** project aims to study the genomic alterations in cancer cells from five different types of laboratory cell cultures, i.e., breast cancer, ovarian cancer, bone cancer, leukaemia and cervical cancer. The project page reads "To analyse the differences between healthy and cancerous cells means to scan every inch of the DNA in these cells and compare it with the healthy ones. But a machine can do this, can't it? Yes, but it is not completely reliable. It has been shown that our eyes perform much better than machines in identifying visual patterns. For this very reason, scientists decided to tackle this scientific problem with the help of the citizens". The initiative offers participants an application for smartphones that allow them "to participate in the common goal to build them a reference genome for the cancer cell line (...). The reference genome will be obtained joining together these small fragments that collectively the players have analysed." In the end, the experiment lasted 20 weeks and citizen scientists identified 181 genome regions of scientific interest in cancer, under the guidance of a scientific team from CNAG-CRG in Barcelona. The app was downloaded by more than 40.000 people from 154 countries and obtained 600,287 solutions. Here again we imagine that this data source can be a key asset to drive public sector interventions in terms of ensuring better care.

- Multilateral benefit-sharing – the World Health Organisations' Pandemic Influenza Preparedness Framework

The World Health Organisation's Pandemic Influenza Preparedness (PIP) Framework Partnership Contribution is another potentially useful and instructive case study. The WHO PIP Framework, adopted by the World Health Assembly in 2011, aims to, "bring(s) together Member States, industry, other stakeholders and WHO to implement a global approach to pandemic influenza preparedness and response." Under it, biological materials of influenza viruses with pandemic potential are shared. Two benefit-sharing arrangements attach to the PIP.

The first benefit-sharing arrangement is SMTA 2, a "legally-binding contract between WHO and an influenza product manufacturer, research institution, or other entity that

_

⁸ 'Pandemic Influenza Preparedness (PIP) Framework' (WHO, n.d.)

receives PIP Biological Materials (PIP BM), such as influenza viruses with pandemic potential (IVPP), from a laboratory which is part of the Global Influenza Surveillance and Response System (GISRS). In exchange for access to PIP BM, the entity commits to provide to WHO, benefits that can be used to prepare for (e.g. training, technology license) or respond to (e.g. vaccines, antivirals, diagnostic kits) pandemic influenza."⁹

The second PIP benefit-sharing arrangement is the partnership contribution, "an annual cash contribution of US\$ 28 million given to WHO by influenza vaccine, diagnostic and pharmaceutical manufacturers that use the WHO Global Influenza Surveillance and Response System (GISRS). Funds are allocated for pandemic influenza preparedness capacity building, response activities at the time of a pandemic, and the PIP Secretariat for implementation of the Framework." A questionnaire is sent by the WHO every year to potential contributors with the funds then allocated to capacity building, pandemic response and maintenance of the PIP Framework. 11

The PIP Framework could be a useful model in respect of benefit-sharing obligations tied to commercial entities receiving public sector personal information. As a case study, the PIP Framework is instructive in pointing towards a model of benefit-sharing consisting of payments from users to support the data infrastructure of the information they receive – and presumably gain a benefit from.

Key Findings, Recommendations and Areas For Further Study

Our key findings can be summarised as follows:

- In the literature reviewed, notions are not understood in a straightforward way and their nuances should be embraced and disclaimed as appropriate.
- Public-sector bodies generally lag behind in developing and implementing data sharing regimes and data analytics as compared to the private sector;

⁹ 'Standard Material Transfer Agreements (SMTA2)' (WHO, n.d.)

¹⁰ <u>'PIP Framework Partnership Cont</u>ribution' (WHO, n.d.)

¹¹ 'Influenza – PIP Partnership Contribution' (WHO, n.d.)

however, there are existing good practices in specific policy areas – such as in public health – where governments' sharing of personal data with the private sector is becoming a component of many governments' health and economic strategies.

- Studies demonstrate that ordinary people are supportive of health and social care data being used for public benefit but wish those public benefits to outweigh private profits and interests.
- When assessing costs and benefits of data sharing with the private sector, the studies assessed indicate that such costs and benefits should not be conceived of as solely financial but understood in broader, more social terms. In a related sense, while governments often conceptualise public benefit in economic terms, the public itself may not share this view and there is a need to distinguish between the framing of public benefit espoused by governments, with that understood by a plurality of public(s).
- The term public benefit is not well defined in the literature and is often used interchangeably with other terms such as the public interest and the common good. Where attempts are made to define public benefit, the literature indicates that this may include direct benefits – such as an income stream to the UK National Health Service (NHS) – and indirect benefits.
- There were prerequisites for achieving public benefit identified in the literature
 with values such as transparency that are very much connected to public
 benefit. By extension, public engagement in terms of data assessment is
 necessary as an input for public benefit, and to ensure a social license such
 that data is shared for the common good.
- There is some concern within the literature that a lack of a definition of public benefit may enable the concept to be exploited to facilitate access to, and commercialisation of, government-held personal data. Accordingly, the lack of a definition for public benefit could leave room for exploitation of the concept to promote government economic interests, which may not align with public conceptions of public benefit.
- There is a vast literature on data value in general. In relation to understandings of the 'value' of personal data, some of the literature surveyed indicated that this is jurisdiction specific and informed by the underlying context and socio-technical settlement. An additional theme in the literature surveyed was that the value of publicly held personal data cannot simply be defined in economic terms. Where the value of publicly held personal data is conceptualised in narrow economic terms, this is not sustainable in the long term, particularly in the healthcare domain.
- There were concerns that the very 'assetisation' of personal data may influence conceptions of value, thereby potentially resulting in a lack of public scrutiny and inequity in terms of access to knowledge. As a way to overcome issues with 'assetisation' and taking account of the socio-technical

underpinnings of conceptions of value - value co-creation and exchange beyond the market was suggested as a way to take such issues out of the economic arena and re-contextualizing them to embrace co-creative approaches to how value is conceived of.

- Benefit-sharing is a concept typically associated with international environmental law and in particular, international biodiversity law to deliver commutative and distributive justice. Benefit-sharing is thus linked to justice and emphasises the optimisation of benefits to society, together with the minimisation of harm, and the achievement of equity.
- If data has the potential to benefit the public, it should be shared. Public benefit cannot be obtained in the absence of such sharing. However, the absence of common principles for trusted government-to-businesses (G2B) personal data sharing may lead to undue restrictions on data flows resulting in detrimental economic impacts. The literature surveyed identifies this and constraints for promoting G2B data sharing. Legal certainty was hence identified as required for such sharing to take place.
- Creative and fruitful collaboration schemes are being established between civic organisations and private actors (at times also engaging the public sector) to share personal data, as occurs for certain citizen science activities, with creative commons licensing schemes, value co-creation, and the reality of 'data cooperatives'.
- There is a growing attention in the literature for the concept of 'data altruism' as also incorporated in the European Union Data Governance Act; this reflects a tendency to embrace a fair and open sharing of personal data for public benefit.
- In the literature surveyed, we identified a number of guiding principles that could be of help for designing an appropriate framework for public sector personal data (benefit) sharing, among the others key principles are of proportionality, transparency, public engagement, co-creation of the concept of value, legal certainty and respect for ethical values and norms.
- Cross-national initiatives are being set up to provide options for the operational framework and governance models for the exchange and secondary use of personal data between countries and actors, respecting the principles of transparency, trust, FAIRness, citizen empowerment and common good.

Based on our findings, we can recommend the following steps:

 When using certain terms in a policy document or official recommendations, remember that notions are not understood in a straightforward way in the literature and their nuances should be embraced and disclaimed as appropriate;

- Start with pilot sharing in those fields where studies demonstrate that ordinary people are supportive of health and social care data being used for public benefit but make sure that public benefit outweigh private profits and interests;
- Assess costs and benefits of data sharing with the private sector not only from a financial but also broader perspective, including broader social dimensions;
- Be aware of how framings from government as well as other dominant actors (such as the market) can erode or undermine 'true' public benefit and engage the public via processes of public engagement on the assessment of public benefit and in the co-creation of the notion of value;
- Emphasise principles of commutative and distributive justice in considering benefit-sharing arising from the use of publicly held personal data;
- Implement effective strategies that tackle the identified constraints for promoting G2B data sharing, for example the absence of common principles on the matter;
- Turn to creative and fruitful collaboration schemes and initiatives existing between research centres, civic organisations and private actors (at times also engaging the public sector) to share personal data, taking for example certain citizen science activities and the reality of data cooperatives and of creative commons licensing schemes;
- Respect key principles such as that of proportionality, transparency, accountability, and respect for ethical values and norms in designing frameworks for public sector personal data (benefit) sharing. Value cocreation should also be promoted in the construction of benefits.
- Seek to build social licence as a fundamental resource to ensure that private data sharing with businesses operates as a trigger for public good.

In terms of further work and study which could be undertaken to enhance understanding and knowledge in this area of focus;

- While beyond the scope of this literature review, it may be useful for the Scottish Government to define key terms to facilitate discussions in this field with a view to ensuring shared understandings of such terms.
- While also largely beyond the scope of this literature review, there may be
 much to learn from surveying benefit-sharing arrangements in the biodiversity
 world where benefit-sharing is firmly embedded as both a principle and an
 outcome of access to certain forms of information. While such arrangements
 were included in this review to the extent they fell within our search focus, a
 broader review would arguably do greater justice to the potential
 commonalities across these areas.

 Our study, by necessity, focused on the sharing of personal data by public bodies and reviewed the literature within this context. There would be merit, however, to elicit further studies specifically focused on concepts such as value within data more generally. This would arguably yield a wider, and potentially more rounded, approach to the concept of value.

REFERENCES

- 1. Abboud, L. et al. (2021) Summary of results: case studies on barriers to cross border sharing of health data for secondary use. TEHDAS Milestone. Retrieved 8 July, 2022, from https://tehdas.eu/app/uploads/2021/09/tehdas-summary-of-results-case-studies-on-barriers-to-sharing-health-data-2021-09-28.pdf
- Aidinlis, S. (2022) Chapter 5: Government-to-Business (G2B) research data sharing and the GDPR: Reconciling the public with the private? In E. Kosta, R. Leenes & I. Kamara (Eds.) Research Handbook on EU Data Protection Law (pp. 115-142). Edward Elgar. https://www.elgaronline.com/view/edcoll/9781800371675/9781800371675.00014.xml
- 3. Aitken, M., Porteous, C., Creamer, E., & Cunningham-Burley, S. (2018). Who benefits and how? Public expectations of public benefits from data-intensive health research. *Big Data & Society*, *5*(2). https://doi.org/10.1177/2053951718816724
- 4. Andanda, P. (2019) Towards a Paradigm Shift in Governing Data Access and Related Intellectual Property Rights in Big Data and Health-Related Research. *IIC* 50, 1052-1081. https://doi.org/10.1007/s40319-019-00873-2
- 5. Ballantyne, A., & Schaefer, G. O. (2020) Public interest in health data research: Laying out the conceptual groundwork. *Journal of Medical Ethics*, *46*(9), 610-616. doi:10.1136/medethics-2020-106152
- Beeck Center for Social Impact + Innovation (2017) Accelerating the Sharing of Data Across Sectors to Advance the Common Good" by the (U.S.). Georgetown University. Retrieved 5 August, 2022, from https://beeckcenter.georgetown.edu/wp-content/uploads/2019/07/Accelerating-Sharing-of-Data.pdf
- 7. Blankertz A. (2020) How to use data for the public interest, even or especially in a pandemic. Heinrich Böll Foundation Commentary. Retrieved 2 July, 2022, from https://hk.boell.org/en/2020/11/24/how-use-data-public-interest-even-or-especially-pandemic
- 8. Borgesius, F. Z., Gray, J., & van Eechoud, M. (2015) Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. *Berkeley Technology Law Journal*, *30*(3), 2073–2131. https://www.jstor.org/stable/26377585
- 9. Braunack-Mayer, A. et al. (2021) Sharing Government Health Data With the Private Sector: Community Attitudes Survey. *J Med Internet Res*, 23(10): e24200. https://doi.org/10.2196/24200

- 10. Cheung, S. (2020) Disambiguating the benefits and risks from public health data in the digital economy. *Big Data & Society*, 7(1). https://doi.org/10.1177/2053951720933924
- 11. Comandè, G., & Schneider, G. (2022) it's time. Leveraging the gdpr to shift the balance towards research-friendly eu data spaces. Common Market Law Review, 59(3), 739-776. doi:10.54648/COLA2022051
- 12. CORSS (2021) Streamlining the responsible sharing of public sector data. Retrieved 10 August, 2022, from https://www.corrs.com.au/insights/streamlining-the-responsible-sharing-of-public-sector-data
- 13. CORSS (2022) New public sector data sharing laws: key considerations for the private sector. Retrieved 10 August, 2022, from https://www.corrs.com.au/insights/new-public-sector-data-sharing-laws-key-considerations-for-the-private-sector
- 14. European Commission (2019) Digital Single Market: EU negotiators agree on new rules for sharing of public sector data. Press release. Retrieved 2 August, 2022, from https://ec.europa.eu/commission/presscorner/detail/en/IP 19 525
- 15. Guay, R., & Birch, K. (2022) A comparative analysis of data governance: Sociotechnical imaginaries of digital personal data in the USA and EU (2008–2016). *Big Data & Society*, 9(2). https://doi.org/10.1177/20539517221112925
- 16. Hopkins Van Mil (2021) Putting Good into practice A Findings Report. PGiP report. Retrieved 2 June, 2022, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/977737/PGiP_Report_FINAL_1304.pdf
- 17. Huston, P., Edge, V.L., & Bernier, E. (2019) Reaping the benefits of Open Data in public health. *Can Commun Dis Rep.* 45(11). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6781855/
- 18. Hutchings, E., Loomes, M., Butow, P., & Boyle, F. M. (2020) A systematic literature review of health consumer attitudes towards secondary use and sharing of health administrative and clinical trial data: A focus on privacy, trust, and transparency. *Systematic Reviews*, *9*(1) doi:10.1186/s13643-020-01481-9
- 19. Igumbor JO, Bosire EN, Vicente-Crespo M et al. (2021) Considerations for an integrated population health databank in Africa: lessons from global best practices. Wellcome Open Res, 6:214 https://doi.org/10.12688/wellcomeopenres.17000.1
- 20. Involve (2018) Data for Public Benefit Balancing the risks and benefits of data sharing. Retrieved 5 July, 2022, from https://understandingpatientdata.org.uk/sites/default/files/2018-04/Data%20for%20public%20good_0.pdf
- 21. Jenkner, C. S., Ravi, N., Gabel, M., & Vogt, J.-C. (2022) Trust in data-requesting organizations—A quantitative analysis on cultural antecedents and individual-

- level perceptions. *The Electronic Journal of Information Systems in Developing Countries*, 88(4), e12208. https://doi.org/10.1002/isd2.12208
- 22. Fischli, R. (2022) Data-owning democracy: Citizen empowerment through data ownership. European Journal of Political Theory, doi:10.1177/14748851221110316
- 23. Laurie, G., & Stevens, L. (2016) Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom. *Journal of Law and Society*, *43*(3), 360–392. http://www.jstor.org/stable/45179947
- 24. Liddell, K., Simon, D., Lucassen, S. (2021) Patient data ownership: who owns your health?, *Journal of Law and the Biosciences*, 8 (2) https://doi.org/10.1093/jlb/lsab023
- 25. Lucas, J.C. et al. (2013) Donating Human Samples: Who Benefits? Cases from Iceland, Kenya and Indonesia. In D. Schroeder & J.C. Lucas (Eds.) *Benefit-sharing*. Springer. https://doi.org/10.1007/978-94-007-6205-3_5
- 26. Lusch, R. F., & Nambisan, S. (2015) Service innovation. *MIS quarterly* 39(1), 155-176. https://doi.org/10.25300/MISQ/2015/39.1.07
- 27. Martens, B. & Duch Brown, N. (2020) The economics of business to-government data sharing, JRC Digital Economy Working Paper, No. 2020-04, European Commission, Joint Research Centre. Retrieved 10 July, 2022, from https://www.econstor.eu/bitstream/10419/227699/1/1691869767.pdf
- 28. Mészáros, J., & Ho, C. -. (2018) Big data and scientific research: The secondary use of personal data under the research exemption in the GDPR. *Hungarian Journal of Legal Studies*, *59*(4), 403-419. doi:10.1556/2052.2018.59.4.5
- 29. Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020) Emerging models of data governance in the age of datafication. *Big Data & Society* 7(2). https://doi.org/10.1177/2053951720948087
- 30. Mockus, M. (2017) Open Government Data Licensing Framework: An Informal Ontology for Supporting Mashup. PhD Thesis defended at the University of Bologna. Retrieved 1 August, 2022, from http://amsdottorato.unibo.it/8215/1/Mockus Martynas tesi.pdf
- 31. Munné, R. (2016) Big Data in the Public Sector. In J. Cavanillas, E. Curry & W. Wahlster (Eds.) *New Horizons for a Data-Driven Economy*. Springer. https://doi.org/10.1007/978-3-319-21569-3_11
- 32.OECD (2015) Assessing government initiatives on public sector information: A review of the OECD Council Recommendation. OECD Digital Economy Papers No. 248. Retrieved 13 August, 2022, from https://doi.org/10.1787/5js04dr9l47j-en
- 33. OECD (2019) Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. Retrieved 22 July, 2022, from

- https://www.oecd-ilibrary.org/sites/b4d546a9-en/index.html?itemId=/content/component/b4d546a9-en
- 34.OECD (2020) Statement by the Committee on Digital Economy Policy. Retrieved 22 July, 2022, from https://www.oecd.org/digital/trusted-government-access-personal-data-private-sector.htm
- 36.OECD (2022) Recommendation on Enhancing Access to and Sharing of Data. OECD instruments. Retrieved 22 July, 2022, from https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463
- 37. Ormondroyd, E., Border, P., Hayward, J. *et al.* (2022) Genomic health data generation in the UK: a 360 view. *Eur J Hum Genet* 30, 782–789. https://doi.org/10.1038/s41431-021-00976-w
- 38. Research Data Alliance (2016) RDA Principles on the Legal Interoperability of Research Data. Retrieved 2 August, 2022, from https://www.rd-alliance.org/rda-codata-legal-interoperability-research-data-principles-and-implementation-guidelines-now
- 39. Christine Rinik (2020) Data trusts: more data than trust? The perspective of the data subject in the face of a growing problem, International Review of Law, Computers & Technology, 34:3, 342-363, DOI: 10.1080/13600869.2019.1594621
- 40. Schaefer, G.O., Laurie, G., Menon, S. *et al.* (2020) Clarifying how to deploy the public interest criterion in consent waivers for health data and tissue research. *BMC Med Ethics* 21, 23, https://doi.org/10.1186/s12910-020-00467-5
- 41. Scharfbillig, M. et al. (2021) Values and Identities a policymaker's guide. Publications Office of the European Union. Retrieved 18 September, 2022, from https://publications.irc.ec.europa.eu/repository/handle/JRC126150
- 42. Schroeder, D. (2007) Benefit-sharing: It's Time for a Definition. *Journal of Medical Ethics*, 33(4), 205–209. http://www.istor.org/stable/27719838
- 43. Science Europe (2021) Science Europe Response to the European Commission's Consultation on a Future Data Act. Retrieved 2 August, 2022, from https://www.scienceeurope.org/media/nd1jsyl3/20210827_dataact_seresponse_final.pdf
- 44. Shaw, J.A., Sethi, N. & Cassel, C.K. (2020) Social license for the use of big data in the COVID-19 era. *npj Digit. Med.* 3, 128. https://doi.org/10.1038/s41746-020-00342-y

- 45. Stephenson, N., Smith, C. & Vajdic, C.M. (2022) Health and public sector data sharing requires social licence negotiations. *Australian and New Zealand Journal of Public Health* 46: 426-428. https://doi.org/10.1111/1753-6405.13261
- 46. Tzanou, M. (2022) Modern Servitude and Vulnerable Social Groups: The Problem of the Al Datafication of Poor People and Women. *IEEE Technology and Society Magazine*, 41 (2): 105-108,
- 47. Van Eechoud, M. & van der Wal, B. (2008) Creative commons licensing for public sector information Opportunities and pitfalls. IVIR. Retrieved 1 June, 2022, from https://www.ivir.nl/publicaties/download/CC_PublicSectorInformation_report_v3.p df
- 48. Vargo, S. L., & Lusch, R. F. (2008) Service-dominant logic: continuing the evolution. *Journal of the Academy of marketing Science* 36(1), 1-10. https://link.springer.com/article/10.1007/s11747-007-0069-6
- 49. Vezyridis, P., & Timmons, S. (2021) E-infrastructures and the divergent assetization of public health data: Expectations, uncertainties, and asymmetries. *Social Studies of Science*, *51*(4), 606-627. doi:10.1177/0306312721989818
- 50. Vezyridis, P., & Timmons, S. (2017) Understanding the care.data conundrum: New information flows for economic growth. Big Data and Society, 4(1) doi:10.1177/2053951716688490
- 51. Waind, E. (2020) Trust, security and public interest: Striking the balance a narrative review of previous literature on public attitudes towards the sharing, linking and use of administrative data for research. *International Journal of Population Data Science*, *5*(3) doi:10.23889/IJPDS.V5I3.1368

Governmental/legal documents

Regulation EU 2016/679 (General Data Protection Regulation of 2016) - https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

The Australian Government Data Availability and Transparency Act (DAT Act, 2022) - https://www.legislation.gov.au/Details/C2022A00011

The EU Data Act (2022) - https://digital-strategy.ec.europa.eu/en/policies/data-act

The European Strategy for Data (2022) - https://digital-strategy.ec.europa.eu/en/policies/strategy-data

The UK Government Licensing Framework (2016) - https://cdn.nationalarchives.gov.uk/documents/information-management/uk-government-licensing-framework.pdf

The UK Information Communication Office's "Data sharing across the public sector: the Digital Economy Act codes" – (accessed September 1, 2022) https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/data-sharing-across-the-public-sector-the-digital-economy-act-codes/

BIOS OF THE AUTHORS

Anna Berti Suman is Marie Skłodowska-Curie researcher at the European Commission Joint Research Centre. Through the project "SensJus", she is performing research on the potential of civic monitoring as a source of evidence for environmental litigation and as a tool to foster environmental mediation. She is a Qualified Barrister in environmental and climate law under the Bar Association of Rome, following cases at Systasis - Study Centre for the Management of Environmental Conflicts, Milan. Anna obtained her PhD from the Tilburg Institute for Law, Technology and Society (TILT) with a project aimed at investigating how grassroots-driven monitoring initiatives based on sensor technology influence the governance of environmental health risk. She has a background in Law from the University of Bologna and Transnational Law from the University of Geneva. Her specialisations are International and European Environmental Law, Health Law and Technology, Data Governance and Participatory Data Spaces, Social Innovation. She has work and research experience in the health sector, environmental sector and water law. Her professional page can be found at:

https://joint-research-centre.ec.europa.eu/anna-berti-suman en

Stephanie Switzer is a senior lecturer in the University of Strathclyde's Law School and has extensive consultancy and research experience. She holds an LLB (first class Honours) degree in Law and a PhD in international law. She is a founding member of the Strathclyde Centre for Environmental Law and Governance and has expertise in undertaking qualitative research. She has delivered on several knowledge exchange projects and was also involved in a consultancy for the UN Convention on Biological Diversity which looked at a particular aspect of the benefit-sharing regime internationally, a subject matter that aligns closely with the thematic content of this project on data value and public benefit. She therefore has considerable experience and expertise in looking at issues relating to how data is valued and how the benefits from such data can be shared. Her professional profile can be found here: https://www.strath.ac.uk/staff/switzerstephaniedr/

Views and opinions expressed in this study are those of the authors only and do not by any means reflect those of the authors' institutions.



© Crown copyright 2023



This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit **nationalarchives.gov.uk/doc/open-government-licence/version/3** or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: **psi@nationalarchives.gsi.gov.uk**.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at www.gov.scot

Any enquiries regarding this publication should be sent to us at

The Scottish Government St Andrew's House Edinburgh EH1 3DG

ISBN: 978-1-80525-777-6 (web only)

Published by The Scottish Government, August 2023

Produced for The Scottish Government by APS Group Scotland, 21 Tennant Street, Edinburgh EH6 5NA PPDAS1238742 (08/23)

www.gov.scot