GOVERNMENT-TO-BUSINESS (G2B) RESEARCH DATA SHARING AND THE GDPR: RECONCILING THE ‘PUBLIC’ WITH THE ‘PRIVATE’?

Dr Stergios Aidinlis

This chapter will appear in Eleni Kosta and Ronald Leenes (eds), *Research Handbook of EU Data Protection Law* (Edward Elgar, 2022 forthcoming)

Abstract

Governments across the EU are increasingly turning their attention to advanced big data analytics, aiming to use their data to inform the design and implementation of public policies. Due to limitations in expertise and resources, this is often impossible without the formation of data sharing partnerships with private actors. Yet, the prevailing view in EU data protection regulatory guidance is that the ‘public interest’ and private interests as lawful grounds for data processing under article 6 GDPR find themselves in a zero-sum relationship. The ‘public interest’ under article 6(1)(e) GDPR is construed as the exclusive realm of public authorities, which are often advised against relying on other grounds for processing, associated with private interests, such as ‘legitimate interests’ under article 6(1)(f) GDPR. This chapter argues against the presently dominant divide between public and private interests under lawful grounds for processing, sketching the emergence of Government-to-Business (G2B) research data sharing in the EU. A conceptualisation of the ‘public’ interest as not incompatible with private interests, as long as a contribution to societal well-being is made through data processing, is offered in that regard. The chapter elaborates on this conceptualisation and the requirements for ensuring protection of the fundamental rights of data subjects, while reflecting on the research questions that should concern future EU data protection law researchers with regard to its adoption.

Key words: research data sharing; government data; G2B data sharing; public interest; legitimate interests; GDPR

I. Introduction: COVID-19 and public-private research data sharing

In more ways than one could possibly enumerate, the COVID-19 pandemic took the world by surprise. Within an unprecedentedly short period of time, i.e. a couple of months, well-established for decades social and economic patterns changed and gave rise to novel approaches to doing business and engaging in social interaction. At its clearest, the impact of the pandemic can be seen in the rapid digitalisation of both goods and services, with more and more types of labour being moved online in an effort to curb the aggressive spread of the disease.¹ Taking into account such major implications, and other ones e.g. the amplification of pre-existing tensions between global power blocs, business leaders, politicians and

¹ Deloitte, ‘How Covid-19 contributes to a long-term boost in remote working’ <<https://www2.deloitte.com/ch/en/pages/human-capital/articles/how-covid-19-contributes-to-a-long-term-boost-in-remote-working.html>> accessed 29 November 2020 (applies to all links in this chapter).

international organisations have gone as far as characterising COVID-19 as the ‘great accelerator’.²

When it comes to data sharing, the pandemic, arguably, provides one of the clearest illustrations of not only the benefits, but also the necessity, in some cases, of sustained collaboration between the private and the public sector. Scholars working on the datafication of modern societies have been making compelling arguments for such partnerships from multiple disciplinary perspectives. Economists have argued that governments should treat their data as a strategic asset,³ pointing out the substantial increase of the economic value of such data in recent years.⁴ Sociological work has brought to the fore the experiences of public-sector administrators and the operational challenges they face in accessing privately-held data, highlighting the potential of public-private data partnerships to serve a ‘public interest mission’.⁵ More and more scholars and policy-makers have been converging on the view that the appropriate use of data is, in OECD’s words, a ‘fundamental means of creating public value’.⁶

How does COVID-19 affect the production of such value? The pandemic’s game-changing impact in this context consists in shifting the discussion from a strategic level,⁷ to the concrete, and unfortunately dramatic, need for more effective health data sharing as a life-and-death matter. The outbreak of the virus does not discriminate between public and private organisational apparatuses. Governments often find themselves relying on e.g. telecommunications or transport data, largely held by private operators, to understand population distribution and mobility patterns and, thus, inform their national containment strategies.⁸ The recent statement by the World Health Organisation that ‘enabling the greatest global good will require solidarity and collaboration, establishing (...) public-private partnerships, and facilitating (...) information sharing’,⁹ reflects the rapidly increased demand for wider data sharing that will guide global and national decision-making.¹⁰

In spite of its accelerating capacity, even the C-19 pandemic was not sufficient to move some regulatory mountains in the context of EU data protection law and, more specifically, the EU

² Morgan Stanley, ‘The COVID-19 Accelerator’ <<https://www.morganstanley.com/im/en-us/individual-investor/insights/articles/the-covid-19-accelerator.html>> ; P Wolf and I Bicu, ‘COVID-19 as an accelerator for information operations in elections’ <<https://www.idea.int/news-media/news/covid-19-accelerator-information-operations-elections>>.

³ B Engels, ‘Data Governance as the Enabler of the Data Economy’ (Intereconomics) <<https://www.intereconomics.eu/contents/year/2019/number/4/article/data-governance-as-the-enabler-of-the-data-economy.html>>.

⁴ M Savona, ‘The Value of Data: Towards a Framework to Redistribute It’ (SPRU Working Paper Series) <<https://zenodo.org/record/3967203#.X7asUS0Rry9>>.

⁵ M Micheli, ‘Accessing privately held data: Public/private sector relations in twelve European cities’ (Data for Policy 2020) <<https://zenodo.org/record/3967044#.X7as6i0Rry9>> 1.

⁶ OECD, ‘The Path to Becoming a Data-Driven Public Sector’ <https://www.oecd-ilibrary.org/governance/the-path-to-becoming-a-data-driven-public-sector_059814a7-en> 2.

⁷ B-C Ubaldi, F Curina and B Welby, ‘Towards a new data-driven service design and delivery model?’ (Data for Policy 2020) <<https://zenodo.org/record/3978179#.X7auPC0Rry9>>.

⁸ A Aral et al, ‘Building a data ecosystem for using telecom data to inform the COVID-19 response efforts’ (Data for Policy 2020) <<https://zenodo.org/record/3992420#.X7auxi0Rry9>>.

⁹ World Health Organisation, ‘COVID-19 Strategy Update, World Health Organisation’ <<https://www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020.pdf>>.

¹⁰ C Emilsson, ‘The interdependency of data governance and open government data: lessons from COVID-19’ (Data for Policy 2020) <<https://zenodo.org/record/3978270#.X7avHS0Rry9>>.

General Data Protection Regulation (hereafter the GDPR).¹¹ As it has been documented in many authoritative studies,¹² one of the main barriers in realising wider use of public-sector data – often including personal data¹³ – within the EU relates to the lack of appropriate governance frameworks for facilitating and monitoring public-private collaborations in data sharing. Of course, ‘appropriate’ shall not be conflated with ‘accommodating at all costs’: despite its significant potential value, the provision of government data access to private-sector entities raises heightened risks of abuse or misuse of personal data, creating challenges for the protection of the fundamental rights of privacy and data protection of individual citizens or groups.¹⁴

With that in mind, one of the most salient challenges in that regard is the reconciliation of the ‘public interest’ and non-public/private interests under lawful grounds for data processing. The ‘public interest’ under article 6(1)(e) GDPR is construed in regulatory guidance as the realm of public authorities, which are often advised against relying on other grounds for processing, associated with private interests, such as ‘legitimate interests’ under article 6(1)(f) GDPR.¹⁵ Relying on ‘legitimate interests’, however, may not be a sufficient basis for private parties seeking to access government data with a view to producing insights that can feed back into government work and serve the public good. In addition, this distinction implies a clear-cut divide that is far from obvious, as C-19 has also made apparent. Yet, such EU data protection authorities as the CNIL in France and the Garante in Italy have confined the legitimate undertaking of data processing activities in the name of ‘public health’ during the pandemic to ‘public’, as opposed to ‘private’, actors. For example, both authorities have advised private sector employers against actively collecting data about their employees’ health status, at the same time that a similar measure would be legitimately employed by border control officers in an airport.¹⁶

The present chapter seeks to challenge this competitive, zero-sum understanding of the relationship between the ‘public interest’ and non-public/private interests, arguing that the EU GDPR accommodates ‘hybrid’ understandings of the public good and private interests under its lawful grounds for processing. A data processing activity, it will be argued, serves the ‘public interest’ under article 6(1)(e) GDPR as long as it contributes to a substantial improvement of social welfare, even if ‘private interests’ are also furthered through its

¹¹ The EU GDPR is the main focus of the present chapter because it is the applicable legislation to the conduct of public-private research partnerships, as opposed e.g. to the Law Enforcement Directive which applies for such purposes as ‘prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties’ by competent law enforcement authorities.

¹² OECD (n 6); High-Level Expert Group, ‘Towards a European strategy on business-to- government data sharing for the public interest. Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing’ (European Union)

<<https://www.euractiv.com/wp-content/uploads/sites/2/2020/02/B2GDataSharingExpertGroupReport-1.pdf>> 25.

¹³ While the interest in big data analytics is in aggregates and patterns, not individuals, personal data will often need to be processed to produce meaningful insights, even if data are anonymised after some point within a particular analytical process.

¹⁴ These risks are elaborated upon *infra* in sections IV and V.

¹⁵ See e.g. the guidance by the UK Information Commissioner’s Office, ‘When can we rely on legitimate interests?’ <<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/when-can-we-rely-on-legitimate-interests/>>. The same view has been adopted by the French and Italian Data Protection authorities with regard to data processing for public health purposes, see A Mole et al, ‘COVID-19 in the workplace: differing guidance from data protection authorities’ <<https://www.twobirds.com/en/news/articles/2020/global/covid19-in-the-workplace-guidance-from-data-protection-authorities>>.

¹⁶ Mole et al *ibid*.

execution in a proportionate manner. 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. To elaborate on this argument, this chapter will be referring to a particular type of public-private data sharing partnership: government-to-business (G2B) data sharing for research and evaluation purposes (hereinafter referred to as ‘G2B research data sharing’). This type of partnership envisages continuous data sharing flows between a public and a private body, in the interest of sharing public data and analysing it to improve collective knowledge about economic or societal challenges and, thus, to inform public policy-making and delivery.¹⁷ G2B data sharing is an, underexplored compared to business-to-government (B2G) data sharing, 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.¹⁸

This chapter proceeds from a background discussion of the discourse about making better use of government data to serve the public good in the EU, before fleshing out the economic incentives that gave rise to the practice of G2B data sharing. Then, it analyses the legal concerns surrounding G2B data sharing and stemming from the, so-called, ‘public’ – ‘private’ divide under the GDPR lawful grounds for processing.¹⁹ The chapter argues against this divide, offering instead its proposed conceptualisation of the ‘public interest’. Reflections about the limitations of the proposed conceptualisation and the requirements for applying it in practice to ensure the protection of the fundamental rights of data subjects, as well as the relevant research questions that should concern future EU data protection law scholars, are offered in conclusion.

II. Better use of EU government data in the public interest?

While the pandemic offered us some of the most dramatic illustrations of the significance of big data analytics for public policy-making, governments in the EU and beyond have been investing into maximising the benefits from analysing their data for quite some time. Such challenges as natural disasters, migration flows, traffic congestion and waste management have been increasing in both scale and complexity, rendering traditional policy tools inadequate to address them.²⁰ Timely and reliable population data, of the kind that government departments

¹⁷ J Manzonì, ‘Big data in government: the challenges and opportunities’ (21 February 2017) <<https://www.gov.uk/government/speeches/big-data-in-government-the-challenges-and-opportunities>>; L Einav and J Levin, ‘The Data Revolution and Economic Analysis’ (2014) 14(1) IPE 1, 7- 8.

¹⁸ S Aidinlis, H Smith, A Adams-Prassl and J Adams-Prassl, ‘Building a Justice Data Infrastructure’ <<https://www.law.ox.ac.uk/news/2020-10-08-building-justice-data-infrastructure>>; Welsh Government, ‘The Fuel Poverty Data Linking Project’ <<https://gov.wales/fuel-poverty-data-linking-project-impact-health-welsh-government-warm-homes>>; D Mueller and J Moeller, ‘Giving the International Scientific Community Access to German Labor Market Data: a Success Story’ in N Crato and P Paruolo (eds), *Data-Driven Policy Impact Evaluation* (Springer 2018) 101.

¹⁹ An important caveat to mention is that this chapter will not consider public-private research data sharing in the context of law enforcement, where the existence of the EU Law Enforcement Directive (LED) creates some particular legal complications, see N Purtova, ‘Between the GDPR and the Police Directive: navigating through the maze of information sharing in public-private partnerships’ (2018) 8(1) International Data Privacy Law 52-68.

²⁰ M Barbero et al, ‘Big data analytics for policy making’ (EU Commission DG DIGIT) <https://joinup.ec.europa.eu/sites/default/files/document/2016-07/dg_digit_study_big_data_analytics_for_policy_making.pdf> 2.

necessarily collect while delivering public services, have been seen as a key ‘enabler for a better society and a more efficient public sector’.²¹

According to the OECD, there are three main types of activity within which scientific research and analysis of public-sector data bears great potential to generate public value.²² First, through data analysis, governments can better anticipate and plan the design of their policies and interventions, remaining responsive to citizen needs and adapting their strategies in case of changes in circumstances. Second, with analytical insights from public-sector data, civil service bodies can achieve better implementation and delivery of policies, prioritising the engagement with the citizens most in need of government support. Third, the use and analysis of public-sector data is integral to gauging the actual, on-the-ground impact of government policies, allowing for a systematic, rigorous and empirically-informed evaluation of how high-level decisions affect real lives.

The untapped potential of government data to improve public-policy making and implementation is reflected in the fact that these three categories are neither exhaustive, nor strictly demarcated from one another. Within the three broad categories, there are many sub-categories. For example, in the context of law enforcement, many law enforcement authorities have been investing into cutting-edge data-driven technologies that will enhance situational awareness of their officers, i.e. the latter’s understanding about the environment within which they are operating.²³ Another popular type of activity concerns the use of data in AI systems to achieve automated decisions that will augment or substitute human decision-making, particularly within the increasingly digitalised justice system.²⁴ The boundaries between the three broad categories are also quite porous: many data analysis initiatives may both e.g. monitor performance of a government policy and inform strategic policy-making to pay heed to particular citizen needs. A prime example of this is the Welsh-government funded ‘Fuel Poverty’ study which seeks to produce an evidence-based evaluation of the government energy efficiency schemes for low-income households by linking relevant data with hospital admissions data.²⁵

Various governments around the world have observed this ground-breaking potential in the last two decades progress of big data analytics and have been striving to remove barriers to data-driven, digital transformation of their strategies and services. South Korea has created an open competition for ‘public sector big data analysis projects’, encouraging interested researchers to help the government apply the ‘insights provided by the country’s data to its policy challenges’.²⁶ Singapore’s government data strategy defines government data-driven decision-making as ‘entrepreneurship at the national level’ and a safeguard for the nation’s long-term ‘competitiveness and liveability’.²⁷ In the context of the C19 pandemic, the Qatari government has been using dashboards within which publicly available data, such as the ‘popular locations’ data on Google Maps, can be ‘collected, visualized and analyzed to support monitoring of stay-at-home recommendations, as well as to support fine-grained estimates of

²¹ High-Level Expert Group (n 12) 7.

²² OECD (n 6).

²³ E.g. see the EU H2020-funded project DARLENE at www.darleneproject.eu on the development of relevant data-driven technologies.

²⁴ Aidinlis et al (n 18) 1.

²⁵ Welsh Government (n 18).

²⁶ OECD (n 6).

²⁷ D Lim Yew Mao, ‘Bringing Data into the Heart of Digital Government’ (Singapore Civil Service College) <<https://www.csc.gov.sg/articles/bring-data-in-the-heart-of-digital-government>>.

economic impact’.²⁸ In the US, the Federal Data Strategy reaffirms the intention of the federal government to adopt a national strategy towards improving use of its data for ‘evidence-based policies, data-backed services, organisational management and innovation’.²⁹

EU Member States could hardly stay idle and merely observe this trend, particularly when considering the geopolitical importance of securing a competitive advantage in developing data-driven insights and technologies.³⁰ Long before the very recent proposal of the European Commission for a ‘European Data Governance’ regulation that will facilitate ‘making public sector data available for re-use’,³¹ several governments within the EU had heeded the call to explore the potential of big data analytics for policy-making. They had done so by investing resources into the creation of public-sector research data sharing infrastructures. Government departments in Germany have been linking data from administrative processes and the job market and encouraging external researchers to evaluate the efficiency of federal unemployment policy interventions.³² In the Netherlands, the Bureau for Economic Policy Analysis encourages academic researchers to use microdata it has compiled or collected from other government data providers to maximise the public good benefits from using government data.³³ The Greek government recently launched its own open data platform, seeking to provide interested parties with access to anonymised administrative datasets for the purposes of research and evaluation.³⁴

Relevant initiatives often form part of broader, national strategies such as the ‘Data Agenda’ of the Dutch Government,³⁵ or the ‘Public Service Data Strategy 2019-2023’ of the Irish government.³⁶ In some cases, data analysis to address challenges in the public interest can also be driven by a more bottom-up initiative, such as the so-called ‘self-data territorial’ in the French cities of La Rochelle, Nantes and Lyon.³⁷ Local authorities in these cities have developed data sharing initiatives to enable citizens to share their personal data with the government for public interest purposes, including research, while at the same time retaining control over it. Through these initiatives, local authorities in the three cities have improved their energy transition schemes and public transport services, as well as made steps towards combatting the marginalisation of non-French speakers and enhancing social cohesion.

Regardless of the particular type of the initiative, a common difficulty arises sooner or later during the lifecycle of an ambitious research data sharing endeavour. Data sharing infrastructures require long-term and sustained investment of time and resources, if they are to

²⁸ I Weber et al, ‘Real-Time Mobility Analysis Through Google Maps’ (Data for Policy 2020) <<https://zenodo.org/record/3965386#.X7fmai0RrRZ>>.

²⁹ US Government, ‘Federal data strategy’ <<https://strategy.data.gov>>.

³⁰ Naomi Klein, ‘How big tech plans to profit from the pandemic’, The Guardian (13 May 2020) <<https://www.theguardian.com/news/2020/may/13/naomi-klein-how-big-tech-plans-to-profit-from-coronavirus-pandemic>>.

³¹ European Commission, ‘Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)’ (25 November 2020) <<https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-data-governance-data-governance-act>>.

³² Mueller and Moeller (n 18).

³³ R Douven et al, ‘Microdata and Policy Evaluation at CPB’ in N Crato and P Paruolo (eds), *Data-Driven Policy Impact Evaluation* (Springer 2018).

³⁴ Greek Government, ‘data.gov.gr’ <<http://www.data.gov.gr>>.

³⁵ Ministry of the Interior and Kingdom Relations of the Netherlands, ‘NL DIGITAAL Data Agenda Government’ <<https://www.nldigitalgovernment.nl/wp-content/uploads/sites/11/2019/04/data-agenda-government.pdf>>.

³⁶ Department of Public Expenditure and Reform of the Republic of Ireland, ‘Public Service Data Strategy 2019-2023’ <<https://www.gov.ie/en/publication/1d6bc7-public-service-data-strategy-2019-2023/>>.

³⁷ OECD (n 6).

pay off and facilitate the production of a high-quality evidence-basis for public-policy making. Governance structures, as well as sufficient legal and technical expertise, need to be committed to the work of the infrastructure. Various stakeholders, including civil society organisations and other, public or private, end-users of the produced body of knowledge need to be engaged with by government actors to maximise the utility of such knowledge.³⁸

However, public-sector bodies generally ‘lag behind’ in developing and implementing data analytics as compared to the private sector.³⁹ This has traditionally been the reason behind the underuse of the ‘vast amount of valuable data’ that public sector bodies in Europe have been producing ‘for centuries’.⁴⁰ When also tasked with the full responsibility of responding to pressing social challenges and safeguarding the well-being of the general public, the public sector often buckles under other pressures that make research and evaluation data sharing an unneeded luxury. In response to the conundrum of aborting much-needed for the public good data sharing endeavours, the question arose: should governments allow private-sector access to public datasets, with a view to sharing the economic and organisational burdens? If they do, will this type of data sharing partnership serve the ‘public interest’ or ‘private’ interests? Government-to-business (G2B) research data sharing slowly started to emerge as a practice that would have to navigate towards a response to these crucial questions.

III. The slow emergence of G2B research data sharing: organisational challenges and economic incentives

The progress of governmental research data sharing is unequal across the EU, with some countries making significant investments in their ‘capacity to use data strategically to improve policy making’,⁴¹ while other countries are making less progress. Why is that? There is a number of well-known factors in the literature which operate as ‘barriers’ to the wider creation of research data sharing infrastructures in public sectors across the EU (and elsewhere). One factor relates to so-called ‘cultural’ barriers, referring to the lack of a ‘data sharing culture’, i.e. the absence of sufficient appreciation of the value that can be created by public sector organisations through data analysis.⁴² Another similar barrier stems from the lack of relevant appreciation on behalf of the broader public, with citizens often demonstrating low levels of trust as to digital experimentation in the public sector when their personal data is involved.⁴³ Additionally, the lack of specific governance and legal frameworks when it comes to using public-sector data for research and evaluation purposes has often been pointed out.⁴⁴ The lack of such frameworks does not only create uncertainty that may stifle data-driven innovation, but also raises questions as to the conformity of any data sharing that may be happening with the fundamental rights of data subjects.

³⁸ Aidinlis et al (n 18) 31.

³⁹ High-Level Expert Group (n 12) 14.

⁴⁰ M Barbero et al, ‘Study to support the review of Directive 2003/98/EC on the re-use of public sector information’ (EU Commission) <<https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en>> 16.

⁴¹ OECD (n 6).

⁴² Micheli (n 5) 2.

⁴³ High-Level Expert Group (n 12) 27; as it will be discussed *infra* under sections V and VI, such digital experimentation in the public sector has often raised concerns from a privacy, data protection and non-discrimination perspective, with powerful private actors often exercising significant influence over the design of data collection and analysis, as well as governments seeking to develop invasive technologies and predictive analytics.

⁴⁴ *Ibid* 7.

Nonetheless, the barriers that seem to be the most common among EU countries, especially in times when public authorities are resource-constrained like the current ones, boil down to essentially economic factors: the lack of adequate resources and expertise. As the OECD has noted, encouraging better uses of public-sector data often stumbles upon the ‘deficiency of skills among public servants’,⁴⁵ as well as the lack of hardware infrastructure and funds to support the overall implementation of a research data sharing strategy.⁴⁶ Efforts to use government data as part of the COVID-19 response have been hampered by the lack of ‘data infrastructures, standards and interoperability (including scalable tools), and the identification of what data to prioritise for release’,⁴⁷ as well as the requisite skills among government operational and technical staff.

The public-sector drive to share the logistical burdens of research data sharing with the private sector can be very well-understood when considering the particularities of this activity. By contrast to one-off types of information disclosure that public bodies have been used to in the past, e.g. by virtue of freedom of information legislation,⁴⁸ research data sharing is a long-term activity. It involves the creation and monitoring of data sharing agreements, the sustained technical collaboration of data providers and researchers, as well as the on-going involvement of many individuals and groups within the data provider group. While the sophistication and the technical complexity of the activity is reflected in the robustness and the accuracy of the produced evidence-basis, the demands in terms of resources and time are not to be underestimated. Simply put, holding a very valuable resource is not really meaningful, if you do not have the tools or the personnel that will be able to make the best use of it and ‘unlock’ it. Both educated personnel and up-to-date hardware infrastructure are necessary to ‘ensure the capabilities to use and analyze the data’.⁴⁹

Considering these logistical burdens more holistically, it is not sufficient to even secure a leadership commitment to invest some money into technical infrastructure or educate civil servants on how to make better use of government data. Like businesses, public sector organisations need a broader strategic vision of understanding the contributions of data analysis to their mission. This will inform the creation of concrete ‘business cases’ for data-related infrastructure investments that will make the best use of potentially available funds, either these funds come directly from within the government or from external sources.⁵⁰ This is not to say that such a ‘business case’ should be construed in the same way as a market-driven case. Beyond improving government efficiency, there are many intangible benefits from better uses of public-sector data, e.g. increasing accountability, transparency and trust in the democratic institutions by facilitating independent, evidence-based scrutiny of their work.⁵¹ Nevertheless, without a ‘compelling value proposition’ and, thus, financial sustainability,⁵² such an ambitious and long-term investment as a public-sector research data infrastructure is very likely to face serious economic hurdles as it is being implemented.

⁴⁵ Micheli (n 5) 2.

⁴⁶ OECD (n 6).

⁴⁷ Emilsson (n 10) 3.

⁴⁸ A Savage and R Hyde, ‘Using freedom of information requests to facilitate research’ (2014) 17(3) *International Journal of Social Research Methodology* 303.

⁴⁹ Aral et al (n 8) 6.

⁵⁰ Aidinlis et al (n 18) 21.

⁵¹ E.g. HM Government, ‘Open Data White Paper: Unleashing the Potential’ (June 2012) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf> 5: ‘Data is the 21st century’s new raw material. Its value is in holding governments to account’.

⁵² Ubaldi et al (n 7) 3.

Against the backdrop of this reality, public-sector bodies have gradually started to turn to the private sector with a view to participating in data sharing partnerships that will enable them to improve the design and delivery of their policies.⁵³ By contrast to public bodies, private companies have been investing more consistently into their capability to analyse data with a view to improving business decisions and maximising profit.⁵⁴ IT infrastructure providers and data analytics companies have been at the forefront of data-driven public-private partnerships. For example, a multinational telecommunications company, Telefónica, agreed with the United Nations (UN) Food and Agriculture Organisation (FAO) to assist Colombian authorities with estimating the rate of forced migration in the country as a result of extreme drought.⁵⁵ Long before COVID-19, Vodafone, had created a data-driven solution, using only aggregated anonymised mobile data, to help governments ‘track trends in population movements and control epidemics so as to prevent widespread outbreaks’.⁵⁶

These initiatives are not necessarily one-off or short-term; private companies often commit to engaging with the public sector so as to both educate public sector personnel in terms of using data and modernising the public sector data infrastructure. In some cases, public and private bodies form so-called ‘data research partnerships’ within which they seek to identify projects of mutual interest. What does data sharing within such partnerships entail in practice? As previously mentioned, G2B research data sharing differs from other types of information disclosure due to its sustained and complex character. Essentially, a public body interested in engaging in such data sharing would have to identify the appropriate datasets, prepare them for analysis by ensuring data quality (e.g. not many missing variables) and ‘cleaning’ them so as to remove unnecessary information and noise, as well as perform statistical disclosure controls to minimise the amount of identifying information that would be passed on to researchers. The public body would also need to create and oversee the enforcement of a data sharing agreement, including provisions on the retention period of the data by the private entity. Such data research partnerships may not always require personal data. For example, a private entity may require access to anonymised court hearings audio data in the course of developing a transcription algorithm that will provide litigants with a written record of trials.⁵⁷ In most cases, however, as also noted above,⁵⁸ technologies like machine learning algorithms might need personal data of actual individuals, compared e.g. to synthetic data, to yield analytical insights of relevance to the real world.⁵⁹

Although public-private research data partnership have mostly involved analysing privately held data so far,⁶⁰ there have also been such initiatives as ‘data sharing pools’, where various entities, i.e. government, commercial and independent research institutions, exchange data and identify knowledge gaps while minimising duplicative efforts.⁶¹ While not providing names of

⁵³ B Ubaldi et al, ‘A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance’ (2019) <<https://www.sipotra.it/wp-content/uploads/2019/05/A-data-driven-public-sector-Enabling-the-strategic-use-of-data-for-productive-inclusive-and-trustworthy-governance.pdf>> 34.

⁵⁴ High-Level Expert Group (n 12) 5.

⁵⁵ OECD (n 6).

⁵⁶ *Ibid.*

⁵⁷ A similar service is performed by JUST: Access, ‘JUST: Transcription’ <<https://www.just-access.org/just-transcription>>.

⁵⁸ *Supra* (n 13).

⁵⁹ D Rankin et al, ‘Reliability of Supervised Machine Learning Using Synthetic Data in Health Care: Model to Preserve Privacy for Data Sharing’ *JMIR Med Inform* 2020;8(7):e18910: ‘A total of 92% of models trained on synthetic data have lower accuracy than those trained on real data’.

⁶⁰ High-Level Expert Group (n 12) 60.

⁶¹ J Shkabarur, ‘The Global Commons of Data’ (2019) 22 *Stanford Technology Law Review* 1.

organisations due to confidentiality reasons, Micheli elaborates on similar partnerships between private entities and twelve (12) local administrations across the EU.⁶² She elaborates on how local administrations, after identifying ‘shared interests’ with the private companies, provide access to administrative data for such companies by ‘creating a data sharing pool’ in exchange for the expertise and resources contributed by the private sector.⁶³ This approach is hailed as resulting in a ‘co-creation’ relationship instead of creating relationships between ‘buyers’ and ‘sellers’ of data.⁶⁴

In recognition of this potential of private-public partnerships to realise innovative data analytics in the public interest, the European Council has been including such ‘multi-stakeholder models’ as core elements of a ‘new European approach’ in such fields as network and information security since the late 2000s.⁶⁵ This is also reflected in the EU directive 2016/1148 on the security of Network and Information Systems (NIS)⁶⁶ which makes the consideration of public-private collaborations a necessary element of Member States cybersecurity strategies. More specifically, article 7(1)(c) of the directive requires that ‘the national strategy on the security of network and information systems’ of Member States addresses public-private cooperation in identifying measures related to ‘preparedness, response and recovery’.⁶⁷ When it comes to other public sector areas, however, the practice of data sharing between EU governments and businesses is ‘currently sporadic and lacks sustainability’.⁶⁸ How should we explain the difficulties of scaling up G2B as a practice across the EU? While the potential barriers for G2B research data sharing should not be confined to legal considerations, this paper suggests that regulatory guidance about the conceptualisation of ‘public interest’ in EU data protection law creates complications for this type of data sharing. The following section elaborates on these complications.

IV. Legal concerns about G2B research data sharing: the ‘public interest’ vs ‘private interests’ in the GDPR

The GDPR seems to be of two minds when it comes to the status and governance of G2B research data sharing. On the one hand, the regulation adopts an arguably broad definition of ‘research’ that does very well accommodate research endeavours undertaken by private enterprises.⁶⁹ In their ‘Preliminary Opinion on data protection and scientific research’,⁷⁰ the

⁶² Micheli (n 5).

⁶³ *Ibid* 2.

⁶⁴ Although Micheli, *ibid* points out that ‘The results highlight new divides caused by the unequal possibilities to access private data, as cities with more experience and better reputation are in a favourable position both to establish partnerships and receive data at no cost’.

⁶⁵ Purtova (n 19).

⁶⁶ EU Commission, ‘The Directive on security of network and information systems (NIS Directive)’ <<https://ec.europa.eu/digital-single-market/en/network-and-information-security-nis-directive>>.

⁶⁷ Article 7 of the Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union, (2016) OJ L 194/1.

⁶⁸ Micheli (n 5) 1. Operation Tovar (Wikipedia) is a particularly interesting example here of a partnership between law enforcement agencies and companies like Microsoft: ‘Operation Tovar is an international collaborative operation carried out by law enforcement agencies from multiple countries against the Gameover Zeus botnet, which is believed by the investigators to have been used in bank fraud and the distribution of the CryptoLocker ransomware’.

⁶⁹ See specifically GDPR Recital 159: ‘(...) For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research (...)’.

⁷⁰ European Data Protection Supervisor, ‘A preliminary opinion on data protection and scientific research’ <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf>.

European Data Protection Supervisor (EDPS) confirmed that the GDPR definition of research encompasses different research organisations, including academic, not-for-profit, government and for-profit commercial entities.⁷¹ On the other hand, as the following analysis will indicate, it seems that government data may be off-limits for any research endeavour involving commercial entities, even if it seeks to serve the public interest. In other words, there is not enough clarity as to whether and under which principles private actors could be able to process public-sector data with a view to producing knowledge about the efficiency and impacts of government policies.

To elaborate on this ambivalence, the first and the second data protection principles are central, i.e. lawful basis for data processing under article 6(1) GDPR and purpose limitation under article 6(3) GDPR. On the question of lawful basis, considering the difficulties with obtaining consent from data subjects when using such large datasets as the ones held by the government,⁷² the following two bases are relevant:

6(1)(e) Public interest - processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller

6(1)(f) Legitimate interests - processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child

The prevailing view in regulatory guidance is that the ‘public interest’ should be invoked by public authorities, whereas ‘legitimate interests’ should be invoked by private parties including commercial enterprises. In its opinion 06/2014, the Article 29 Working Party had essentially confined reliance on the ‘public interest’ ground to public-sector data sharing uses, either for public bodies with official authority (e.g. ‘tax authority’ or ‘local government body’) or for private bodies which are required to disclose data to public bodies with official authority.⁷³ This interpretation, which considered the draft EU GDPR,⁷⁴ is echoed in post-GDPR national regulatory guidance by the UK, Italian and French Data Protection Authorities (DPAs).⁷⁵ In that sense, the common good and private interests are seen as being in a *zero-sum* relationship; when the former is served, the latter cannot be served at the same time, and vice versa. The law does not seem to envisage a ‘win-win’ relationship within which the public good and private interests are overlapping and being furthered at the same time by research data processing. As a result, a divide is created between government bodies who hold the data and disclose them for research purposes on ‘public interest’ grounds and private entities who may not, according to this view, rely on such grounds.

⁷¹ *Ibid*

⁷² Relying on consent in G2B research data sharing would be highly impractical considering the need to contact a massive number of individuals and also quite difficult to justify legally due to the imbalance of power between public authorities and individual citizens, see J Bell et al, ‘Balancing Data Subjects’ Rights and Public Interest Research’ (2019) 5 European Data Protection Law Review 43.

⁷³ Article 29 Working Party, ‘Opinion 06/2014 on the “Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC” 21-23.

⁷⁴ *Ibid* 5.

⁷⁵ Information Commissioner’s Office (n 15); Garante, ‘Coronavirus: No do-it-yourself (DIY) data collection, says the Italian DPA’ <<https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9282117#1>> ; CNIL, ‘Coronavirus (COVID-19): reminders from the CNIL on the collection of personal data by employers’ <[https://www.cnil.fr/fr/coronavirus-covid-19-les-rappels-de-la-cnil-sur-la-collecte-de-donnees-personnelles-par-les](https://www.cnil.fr/fr/coronavirus-covid-19-les-rappels-de-la-cnil-sur-la-collecte-de-donnees-personnelles-par-les-)>.

Significant legal implications, however, stem from relying on one of these two legal bases rather than the other. First, the ‘public interest’ is a legal basis that offers more powers to data controllers compared to ‘legitimate interests’. For example, under article 21(6) GDPR, research data processing that is ‘necessary for the performance of a task carried out for reasons of public interest’ benefits from an exemption to the data subject right to object to ‘processing of personal data concerning him or her’.⁷⁶ Another example relates to the processing of sensitive personal data by researchers under article 9 GDPR; looking at the clauses of this provision, there is much more leeway for domestic laws to enable such processing if there is a ‘public interest’ reason compared to the ‘legitimate interests’ of a commercial enterprise.⁷⁷ Second, there are distinct applicable legal considerations related to each of the two legal bases. This may create fragmentation and confusion in terms of protecting data subjects rights if different partners in the same data sharing partnership are subject to different standards of protection. The clearest difference relates to the application of the so-called ‘balancing test’ within article 6(1)(f),⁷⁸ i.e. the determination of whether the legitimate interests ‘override’ the data subjects ‘rights and freedoms’. This is not a requirement for the application of the ‘public interest’ basis under article 6(1)(e), where the ‘necessity’ of the data processing for the performance of a task in the public interest suffices.⁷⁹ This might be a crucial difference in practice, particularly when considering the symbolic weight of ‘rights’ *vis a vis* ‘interests’ in legal reasoning.⁸⁰ Hence, invoking ‘legitimate interests’ for research data processing may be both insufficient for data controllers and processors, as well as not transparent enough for data subjects.

Could this risk of divergence be resolved by reference to the *controllership* of the data? Under article 4(7) and (8) GDPR, the data controller is the entity which ‘determines the purposes and means of the processing of personal data’, whereas the data processor ‘processes personal data on behalf of the controller’.⁸¹ Controllership bears important implications both for the first data protection principle, i.e. legal basis, and for the second data protection principle, i.e. purpose limitation. With regard to legal basis, one could argue that public authorities can invoke ‘public interest’ grounds as the data controllers and private entities could ‘borrow’ this legal basis operating as data processors, provided that there is a comprehensive data sharing agreement between the parties.⁸² This is, actually, an often invoked solution in the context of public-private data sharing for law enforcement (e.g. police sharing datasets to external forensics experts), where data processing takes place to ‘protect the public’ and the vital interests of data subjects.⁸³ Extending this solution to generic G2B research data sharing seems to be facing certain difficulties though. A main difficulty relates to the exploratory nature of data-driven research; it is easier to pre-determine the scope of data processing when it comes to a business delivering a specific technical task (e.g. a DNA analysis) compared to when data

⁷⁶ Article 21(6) GDPR.

⁷⁷ E.g. articles 9(2)(g) and (j): ‘processing is necessary for reasons of substantial public interest (...)’ and ‘processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued’.

⁷⁸ Information Commissioner’s Office (n 15).

⁷⁹ Article 6(1)(e) GDPR: ‘processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller’.

⁸⁰ G Webber, ‘On the loss of rights’ in G Huscroft et al (eds), *Proportionality and the Rule of Law Rights, Justification, Reasoning* (CUP 2014) 123; S Tsakyrakis, ‘Proportionality: An assault on human rights?’ (2009) 7(3) *International Journal of Constitutional Law* 468.

⁸¹ Article 4(7)GDPR.

⁸² Purtova (n 19) 45.

⁸³ *Ibid.*

processing aspires to unravel hitherto unknown insights about the design and implementation of government policies.⁸⁴ In seeking to produce such insights, private entities may need to reconsider the originally agreed ‘scope’ and ‘means’ of processing. In practice, this means that the boundaries between ‘controllers’ and ‘processors’ are quite porous in G2B research data sharing partnerships, especially when considering their long-term and sustained collaborative nature.

The second data protection principle, i.e. purpose limitation, also raises certain difficulties in this context. In other examples of public-private research data sharing, private actors may easily demonstrate that they process the data for a compatible purpose with the original purpose of collection. For example, public hospitals without adequate technical facilities to refrigerate the coronavirus vaccine might share the personal information of individual patients to a private entity, agreeing to undertake the freezing, that will process this data to organise the patients’ vaccination appointments and dispatch the relevant vaccine doses as needed. In this case, there is a clear connection between the purpose behind the task carried out in the public interest, and the purpose behind the data processing conducted by the private entity. Both are types of data processing serving the purpose of the protection of public health.⁸⁵ Many cases of generic G2B research data sharing, however, may not fit this picture very well. If, for example, a judicial administration authority discloses judgment data to a commercial enterprise for the purposes of researching the main technological barriers in developing natural language processing algorithms that could be used in predictive analytics,⁸⁶ there might be some distance between research and praxis. In principle, the processing by the private party does contribute to a better understanding of the justice system which serves the public interest. It does so in a number of ways: it models the knowledge of a very experienced (and thus expensive) attorney onto a technological solution that will be available to many more clients, it stimulates open competition between lawyers, including younger attorneys, and promises to reduce the backlog of courts which will positively impact on overstretched legal aid systems.⁸⁷ Yet, it is less clear whether the outcomes of the research will directly impact the actual delivery of justice, at least in the short term.

One could potentially point out that G2B research data sharing might be exempt from the purpose limitation principle, by reference to Article 5(1)(b) and Recital 50 GDPR. Article 5(1)(b), establishing the purpose limitation principle, stipulates that ‘further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes’.⁸⁸ Recital 50 GDPR similarly mentions that ‘further processing for (...) scientific (...) research purposes or statistical purposes should be considered to be compatible

⁸⁴ D Erdos, ‘Systematically Handicapped? Social Research in the Data Protection Framework’ (2011) 20 ICTL 83: pre-determining the scope of processing would be hard to reconcile with the ‘fluid and norm- challenging nature of a social science research endeavour’ supported by big data analytics.

⁸⁵ Note that the private entity is only a data processor in this case, thus this example does not raise difficulties under the first data protection principle as G2B research data sharing does.

⁸⁶ N Aletras et al, ‘Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective’ (2016) PeerJ Computer Science.

⁸⁷ Although note that the legitimate scope of litigation analytics is a controversial issue, with France recently banning the use of such technologies when they use ‘identity data’ of judges or jurors, see R Tromans, ‘France Bans Judge Analytics, 5 Years In Prison For Rule Breakers’ (2019) <<https://www.artificiallawyer.com/2019/06/04/france-bans-judge-analytics-5-years-in-prison-for-rule-breakers/>>.

⁸⁸ Article 5(1)(b) GDPR.

lawful processing operations’.⁸⁹ If we strictly followed the wording of these provisions, and provided that article 89 GDPR safeguards like pseudonymisation are applied, the purpose limitation principle would override the need for a lawful basis invoked by the private actors in G2B research data sharing. Provided that public bodies have collected data lawfully, as will often be the case, they may further disclose it for research purposes.⁹⁰ This solution, though, is fraught with its own challenges. These provisions are confined to the second data protection principle (purpose limitation) rather than all data protection principles, including legal basis.⁹¹ Considering the sensitive character of the question of allowing business access to government data, it would be preferable to adopt a more secure legal position and require a separate lawful basis for private entities processing data within a G2B partnership. This would also be consistent with persuasive pre-GDPR legal authority, such as the Article 29 Working Party ‘Opinion 03/2012 on purpose limitation,’⁹² as well as CJEU case-law treating the two as distinct legal tests.⁹³

What are the practical implications of this legal analysis for G2B research data sharing? Essentially, the conclusion is that the ‘public interest’ is the preferable and most secure legal basis: unless all parties engaged in G2B research data sharing can invoke it, it is very likely that legal complications may block smooth public-private collaborations. Practically, this is quite likely to result in either resource-constrained public bodies being prevented from accessing state of the art data-driven insights, or, equally alarmingly, non-regulated and dodgy public-private research data sharing partnerships. The following section proposes a way to show the fly out of the fly-bottle: detaching the interpretation of ‘public interest’ as a lawful basis from organisational identity and defining it by reference to the contribution of research data sharing to collective societal or economic welfare.

V. Reconciling the ‘Public’ with the ‘Private’: Defining the ‘Public Interest’ by reference to Societal Welfare

As the analysis has suggested thus far, G2B research data sharing might face significant hurdles unless private entities are allowed to rely on the ‘public interest’ as a lawful basis for processing. The currently existing divide in regulatory guidance about article 6 GDPR seemingly prevents this, requiring a strict connection between an organisation’s identity as a public body and their legitimate capacity to contribute to the common good.⁹⁴ Scholars often invoke a number of reasons to argue for the legitimacy of treating public and private interests

⁸⁹ Recital 50 GDPR.

⁹⁰ The UK DPA, the Information Commissioners Office (‘ICO’) seemingly endorse such an interpretation: On ‘further processing’ (but not in the context of consent) the ICO state: ‘If your purposes change over time or you have a new purpose which you did not originally anticipate, *you may not need a new lawful basis as long as your new purpose is compatible with the original purpose.*’ [Emphasis added] See ICO, ‘Lawful basis for processing’ <<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/>>.

⁹¹ Bell et al (n 72) 50.

⁹² In this opinion, ‘Opinion 03/2013 on purpose limitation’ <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> 33 the Article 29 Working Party rejected the idea of conflating the notion of ‘compatibility’ under the purpose limitation principle with the notion of ‘lawful grounds’ for processing under Article 6: ‘Further, these provisions also confuse two different concepts: the notion of ‘compatibility’ under Article 5(b) of the proposed Data Protection Regulation and the notion of ‘lawful ground’ under Article 6. As explained earlier, these two requirements are cumulative’.

⁹³ Joined Cases C-465/00, C-138/01 and C-139/01 *Rechnungshof v Österreichischer Rundfunk* [2003] 3 C.M.L.R 10, paras 65-67 clearly distinguishing between the need for all processing to ‘first’ satisfy purpose limitation ‘and’ ‘second’ one of the lawful grounds for processing.

⁹⁴ Article 29 Working Party (n 73); Information Commissioner’s Office, Garante and CNIL (n 75).

differently. First, it is a general principle, common in European data protection and human rights law, that, when balanced against qualified (i.e. non-absolute) individual rights like privacy, corporate interests are to receive less normative weight compared to public goods.⁹⁵ Second, in the advent of so-called ‘surveillance capitalism’,⁹⁶ there is increasing worry about the collection and aggregation of massive datasets in private hands. Third, regardless of the volume of data, proponents of confining government data access to public-sector and academic researchers point out to empirical findings which reflect citizen uneasiness with private companies processing their data for research purposes.⁹⁷ A prime example in this context is the Facebook emotional contagion study, which was conducted within a public-private partnership between Cornell University and Facebook, involving manipulating the news feed of Facebook users to assess, among other things, whether showing negative posts to users would lead to greater user expressions of sadness.⁹⁸ This study triggered public backlash as users were not informed about the manipulation of their news feed by the company.⁹⁹

Caution is certainly justified and there are salient reasons for ensuring that G2B research data sharing and further processing are accompanied by robust safeguards and ongoing monitoring by public bodies. Nonetheless, if one insists that the organisational identity of the research data sharing participants should be a conclusive consideration when it comes to serving the ‘public interest’, one might be throwing the baby out with the bathwater. Recent high-profile research data processing debacles, like Cambridge Analytica, illustrate that mishandling and manipulation of data are not the exclusive province of commercial entities.¹⁰⁰ Furthermore, if we look at other empirical studies, citizen trust does not necessarily hinge on organisational identity. Other characteristics, such as transparency and clarity about the benefits from research data processing, as well as the existence of sufficient safeguards for data subjects are arguably more central.¹⁰¹ Data subjects are particularly interested in seeing a ‘tangible benefit’ of public nature when data are used by commercial organisations,¹⁰² often demanding a specific articulation of how using data for research will impact the ‘everyday operation or activities of a public body or service’.¹⁰³

For these reasons, the currently prevailing approaches to creating and implementing a satisfactory definition of the ‘public interest’ as a criterion for classifying the users of a public-

⁹⁵ See e.g. Joined Cases C-293/12 and C-594/12, *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others* (8 April 2014) paras 44 and 47 about the importance of the ‘object pursued by the interference’ being ‘an objective of general interest’; in the European human rights law area, see *Eweida v United Kingdom* (2013) 57 EHRR 8 para 94: ‘too much weight was given by the domestic court to BA’s wish to project a certain corporate image and too little to the applicant’s desire to manifest her religious belief’.

⁹⁶ S Zuboff, *The Age of Surveillance Capitalism* (Public Affairs 2019).

⁹⁷ IPSOS Mori, ‘Commercial Access to Health Data’ (2016) <<https://perma.cc/9TK2-JWEB>>.

⁹⁸ A Kramer et al, ‘Experimental evidence of massive-scale emotional contagion through social networks’ (2014) 111 PNAS 8788.

⁹⁹ M Murray, ‘Users angered at Facebook emotion-manipulation study’ <<https://www.today.com/tech/users-angered-facebook-emotion-manipulation-study-1D79863049>>.

¹⁰⁰ P Lewis et al, ‘Cambridge Analytica academic's work upset university colleagues’ (2018) <theguardian.com/education/2018/mar/24/cambridge-analytica-academics-work-upset-university-colleagues>.

¹⁰¹ E Waind, ‘Trust, Security & Public Interest: Striking the Balance. A review of previous literature on public attitudes towards the sharing, linking and use of administrative data for research’ (Data for Policy 2020) <<https://zenodo.org/record/3971006#.X76I2C0RrRZ>> 1 on the importance of the public perceiving ‘benefits of some sort’.

¹⁰² M Oswald, ‘Share and share alike? An examination of trust, anonymisation and data sharing with particular reference to an exploratory research project investigating attitudes to sharing personal data with the public sector’ (2014) 11(3) *scripted*.

¹⁰³ Waind (n 101) 6.

sector research data infrastructure are unduly narrow. One prevalent approach is to use the independent character of research, often associated with academic or civil society/NGO researchers,¹⁰⁴ as a proxy for research in the public interest. Many existing research data sharing initiatives distinguish between academic and commercial researchers,¹⁰⁵ either precluding access to the latter¹⁰⁶ or imposing much stricter requirements on them e.g. requiring that their research project is commissioned by a government department¹⁰⁷ or that they demonstrate ‘autonomy’ and ‘independence’ in ‘formulating scientific conclusions’.¹⁰⁸ A similar approach has been to classify parties requesting access based on their intended use of the data, i.e. their purpose behind accessing the infrastructure’s resources and performing data analysis on them. As with organisational identity, ‘commercial’ and ‘non-commercial’ uses are used to delineate which data sharing requests aim to ‘generate revenue’ as opposed to ‘supporting a public good’.¹⁰⁹ In both cases, the realities of multi-stakeholder research data sharing may obfuscate a clear delineation of the public good from private interests. Consider the following scenario:

A start-up, currently operating as a not-for-profit social enterprise, approach a public body and propose a research data sharing partnership. They are interested in developing data-driven insights to improve accessibility of the services offered by the public body to socially-excluded citizens. Their mission is to alleviate power imbalances in society and strengthen social cohesion. In the long run, the start-up would consider the marketisation of their data-driven solutions, with governments, international organisations and non-Government organisations amongst their potential clientele.¹¹⁰

If the ‘public’ – ‘private’ divide were to persist, public bodies would be faced with a harsh dilemma in similar cases: turn down requests that can produce high-quality research in the public interest without overburdening the public purse or interpret the law more creatively, at the risk of severe reputational harm and legal liability in case of data misuse? The very fact of arriving at this dilemma is a problem in itself: with data providers being expected to exercise a certain amount of discretion in deciding access to public data, their decisions are quite likely to suffer from lack of consistency and transparency. To overcome this unsatisfactory state of affairs and dissociate the ‘public’ interest from the organisational identity of data providers, the following classification is proposed to classify the research users of government data:

- A. *Public interest research*** – when the research project exclusively serves political aims that pursue the collective or societal welfare
- B. *Proprietary research*** – when the research project exclusively serves private interests of a proprietary nature

¹⁰⁴ E.g. researchers affiliated with academic organisations or civil society organisations like Privacy International or Amnesty International.

¹⁰⁵ S Price, ‘Academic and commercial research: Bridging the gap’ (2015) 12(2) Participations.

¹⁰⁶ Mueller and Moeller (n 18) refer to the example of the Federal Institute for Employment Research in Germany which relies on sections 75 of the German Social Code Book X and 282 of the German Social Code Book III as the legal basis for sharing data only for the purposes of ‘independent scientific research’. Also see Research Data Centre of the German Federal Employment Agency, ‘Data Access: Background Information’ <https://fdz.iab.de/en/FDZ_Data_Access/background.aspx>.

¹⁰⁷ HMRC, ‘Research at HMRC’

<<https://gov.uk/government/organisations/hm-revenue-customs/about/research>>.

¹⁰⁸ This is required by Eurostat for access to their ‘safe centre’ in Luxembourg, see Eurostat, ‘How to apply for microdata?’

<https://ec.europa.eu/eurostat/documents/203647/771732/How_to_apply_for_microdata_access.pdf>.

¹⁰⁹ E.g. UK Data Service, ‘Registration’

<<https://ukdataservice.ac.uk/get-data/how-to-access/registration/commercialusers.aspx>>.

¹¹⁰ This scenario is inspired by reference to a similar scenario discussed by Aidinlis et al (n 18) 13.

C. *Hybrid research* – when the research project serves *both* the collective welfare and private interests of a proprietary nature

This conceptualisation is structured around the notion of collective societal or economic welfare, as reflected e.g. in such broader political goals as the standard of living, productivity or other metrics of collective well-being.¹¹¹ Crucially, the conceptualisation detaches contribution to the public good from the identity of an organisation as a public-sector one and enables better use of EU government data by drawing on private-sector resources. This is not to say that the conceptualisation is to be uniformly adopted by all public bodies or to apply to all types of data, nor that all types of research should be granted access to government data. As a matter of fact, purely proprietary research (of type B) using government data would be difficult to justify under the current EU data protection legal framework.¹¹² Furthermore, it might very well be the case that specific bodies, e.g. bodies with competence in national security, or bodies holding specific very sensitive datasets, need to tailor the conceptualisation according to their needs.

Nevertheless, adopting the conceptualisation would enable more ‘hybrid’ types of research data sharing (of type C), like G2B research data sharing, to proceed. This would require both public and private data sharing partners to clearly demonstrate that a substantial contribution is being made to the common good through their partnership. It would essentially mean that the private parties accessing government data need to show that they are contributing to the performance of ‘partially or traditionally public activities’ through producing analytical insights.¹¹³ Furthermore, it should be stressed that the private entities would not be empowered under this framework to repurpose the public-sector data ad infinitum for commercial use, since this would no longer serve the public good. They would be allowed, though, to incorporate the non-identifying insights gained through data analysis into their business models and use them to promote their commercial needs.

Requiring this type of demonstrating contribution to the common good would be a step forward considering that many private-public data sharing partnerships might be currently undergoing under opaque circumstances in the EU and beyond. Government bodies in such countries as Greece and the United Kingdom have recently been at the receiving end of NGO or media criticism for withholding their data sharing agreements with private companies from the public eye.¹¹⁴ Moving forward, best practice, principles and guidance on how to establish contribution to the common good would have to be developed by each government body to stimulate interest in using public datasets for analytical purposes.¹¹⁵ This will certainly not be the most straightforward task. As the Business-to-Government (B2G) High-Level Expert Group of the European Commission has noted, the boundaries of public interest as a notion

¹¹¹ C Cooper, ‘Quantitative Models of Well-Being to Inform Policy: Problems and Opportunities’ (2020) 12 Sustainability 3180.

¹¹² See *supra* IV on the difficulties based on the lack of contribution of such research to the ‘public interest’.

¹¹³ E S Savas, *Privatization and Public-Private Partnerships* (Chatham House, New York 2000) 4.

¹¹⁴ N Giannopoulos, ‘The Greek Government-CISCO data sharing agreement: the lost Atlantis’ (03 October 2020) <<https://www.news247.gr/koinonia/symvasi-cisco-dimosioy-apdpch-tin-zitise-kerameos-den-tin-edose-pote.9005665.html>>; Privacy International, ‘Alexa, what is hidden behind your contract with the NHS?’ <<https://privacyinternational.org/node/3298>>; Naomi Klein, ‘How big tech plans to profit from the pandemic’, The Guardian (13 May 2020) <<https://www.theguardian.com/news/2020/may/13/naomi-klein-how-big-tech-plans-to-profit-from-coronavirus-pandemic>>

¹¹⁵ E.g. the UK Office for National Statistics, ‘ONS research and data access policy’ <<https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy>>.

remain ‘largely undefined’ and whether it includes or not the purpose of ‘improved public services’ is not yet established.¹¹⁶ Collective welfare as a proxy for the public interest does not conclusively solve this problem of definition. It does, however, detach the public interest from the individual organisational missions and interests of public bodies, shifting the discussion towards broader, societal political aims along the lines of such concepts as the one of ‘services of general economic interest’ in EU law.¹¹⁷

How difficult would it be for government bodies interested in enabling G2B research data sharing across the EU to adopt the proposed classification? While there is certainly a lot to be said about the particularities of adopting the classification by specific public bodies, I will confine myself here to three common, overarching challenges. First, public bodies will need to respond to a key question: where to draw the lines in G2B research data sharing? As it has been argued, the proposed classification does allow private entities to access government data for analytical purposes, provided that they contribute to the collective or societal welfare in some way. What kind of contribution, however, is enough and in which cases should this contribution be seen as a marginal and tokenistic excuse to make the wealth of government data accessible to data-savvy private companies? Will resource-constrained public bodies be in the position to assess whether a particular data sharing partnership contributes much more to proprietary private interests compared to the common good? Furthermore, shall public bodies conduct this assessment, what can be the safeguards that they should require from private entities to adhere to for such research data sharing to go ahead?

Even if this is a context-specific enquiry, government bodies could use the principle of ‘public integrity’,¹¹⁸ developed by the OECD with regard to public-private partnerships. Under this principle, public bodies should consistently adhere to ‘ethical values, principles and norms for upholding and *prioritising* the public interest over private interests in the public sector’ (emphasis added).¹¹⁹ Guided by this principle, public bodies shall seek to produce and require clear evidence of public benefits as the major and direct output of a relevant research data sharing partnership, with private benefits only coming as a by-product in the future and under the above-mentioned condition that such benefits rely on non-identifying insights produced via G2B research data sharing. To use the previously mentioned example, if a private entity could produce analytical insights that help the public administration of justice reduce accessibility problems to the justice system in a significant and measurable manner, their potential right to marketise their services in the future using non-identifying parts of those data-driven insights would not be barred by the GDPR.

Secondly, public bodies will need to take data subjects rights seriously to avoid legal liability arising from G2B research data sharing. By data subjects rights, I refer both to the rights in GDPR, e.g. the right to object to data processing,¹²⁰ and the fundamental rights to privacy and data protection as enshrined in the EU Charter of Fundamental Rights and the European Convention on Human Rights. Such key legal principles as proportionality and

¹¹⁶ High-Level Expert Group (n 12) 16.

¹¹⁷ *Ibid* citing Cases C-179/90 *Merci convenzionali porto di Genova* [1991] ECR I-5889, paragraph 27; Case C-242/95 *GT-Link A/S* [1997] ECR I-4449, paragraph 53; and Case C-266/96 *Corsica Ferries France SA* [1998] ECR I-3949, paragraph 45: ‘services that public authorities of the Member States at national, regional or local level classify as being of general interest and, therefore, subject to specific public-service obligations’.

¹¹⁸ OECD (n 6).

¹¹⁹ *Ibid*.

¹²⁰ Article 21(1) GDPR: ‘The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on point (e) or (f) of Article 6(1), including profiling based on those provisions’.

necessity should permeate all the lifecycles of G2B research data sharing, from linkage, cleaning and access, to potential data retention and re-use by both parties.¹²¹ Public bodies must be able to transparently justify why particular datasets need to be processed within a G2B data partnership, as well as how much of the data needs to be personal instead of anonymous.

Thirdly, these safeguards shall not only be considered in terms of general principles, but their application should be translated into the introduction of particular data security safeguards on an operational level to ensure that G2B research data sharing is lawful and proportionate by-design. While G2B research data sharing cannot be considered an exercise of ‘academic’ expression,¹²² it will, in most cases, be considered as falling under the ambit of ‘scientific research’ under article 89 GDPR. Article 89 GDPR confines ‘archiving’ purposes to ‘public interest’ applications, but leaves space for ‘hybrid’ scientific research to benefit from certain research exemptions (realised via domestic law provisions) that would render its realisation ‘impossible or seriously impair’ its purposes under article 89(2) GDPR e.g. limiting the data subjects’ rights to have their personal data rectified (article 16 GDPR) or restrict its processing (article 18 GDPR).¹²³ The ‘appropriate safeguards’ indicated by article 89(1) GDPR include ‘technical and organisational measures’ to ensure respect of the data minimisation principle,¹²⁴ as well as pseudonymisation to mitigate re-identification risks and ensure robust protection of data subjects’ rights. For transparency and accountability reasons, public bodies should develop and publicise a principled position on such data protection by-design (e.g. pseudonymisation) and by-default (e.g. strictly processing only the necessary datasets from the first instance) measures, elaborating on their preferred methods e.g. homomorphic encryption or hashing, rather than confining themselves to abstract commitments that article 89 GDPR safeguards will be applied in the course of G2B research data sharing.

While there is not enough publicly available information on how existing G2B research data sharing partnerships may have already addressed similar data protection aspects, important lessons can be learned by looking at cases of public-private research partnerships utilising privately held data. Some well-known examples come from the United States and are relevant here in terms of best practices. Telenor, a large telecommunications private entity, provided its datasets for research use by academics affiliated with the Public Health department of Harvard. The company’s researchers collaborate with their academic counterparts to ensure respect for data minimisation through hashing and encryption de-identification techniques, as well as data aggregation. More specifically, they sought to identify the ‘maximum spatial and temporal aggregation level’ for the data that would make re-identification risks very remote while at the same time allowing the research questions of the academic study to be answered.¹²⁵ This is a telling example of how important it is for the legal departments of public bodies to communicate effectively with data scientists both within their own organisation, and within the private entity, to ensure the introduction of robust data protection safeguards. MasterCard has also been providing data to research institutions and universities, allowing them to conduct research on their own premises via remote access interfaces.¹²⁶ In this case, the privacy and

¹²¹ Aidinlis et al (n 18) 3.

¹²² So as to benefit from exemptions that would rely on the ‘academic’ derogation under article 85 GDPR: ‘(...) including processing for journalistic purposes and the purposes of academic, artistic or literary expression’.

¹²³ M Mourby et al, ‘Governance of academic research data under the GDPR—lessons from the UK’ (2019) 9(3) International Data Privacy Law 192, 200.

¹²⁴ Article 5(1)(c) GDPR: ‘Personal data should be (...) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed’.

¹²⁵ C Buckee and K Engø-Monsen, ‘Mobile phone data for public health: towards data-sharing solutions that protect individual privacy and national security’ (2016) arXiv:1606.00864.

¹²⁶ J Shkabatur (n 61) 406.

data protection counsel of the company is responsible for assessing the re-identification risks before approving of the data sharing, as well as requiring that researchers will be developing a process that guarantees data destruction after the end of their study. Furthermore, a secure file transfer mechanism had to be developed for the purpose of ensuring data security.

Fourthly, and very importantly, research data sharing practice has shown that legal compliance alone is often not sufficient. To avoid jeopardising the operation of a G2B research data sharing partnership, public bodies should prioritise public engagement strategies to build societal acceptability for the ethical use of government data within public-private partnerships.¹²⁷ The recent case of *care.data* in the UK is telling: the initiative purported to use primary health care medical records in England, in part for research purposes, unless patients explicitly opt out.¹²⁸ Despite being perfectly lawful, the initiative triggered a strong public backlash for not having consulted health practitioners and patients, resulted in resistance and massive opt-outs. It was subsequently indefinitely suspended. Regardless of whether there is a legal requirement for data subjects' consent, the sustainability of G2B research data sharing requires an understanding of what it is that social communities affected by the research 'perceive as the benefits of the work'.¹²⁹ Although there are some great examples of public engagement in the context of data-driven medical research,¹³⁰ there is still a lot of progress to be made when it comes to engaging the public while improving broader public-sector policy-making through G2B data sharing.

The common denominator of these three challenges is the need for what has been defined by the OECD as a 'comprehensive model for data governance'.¹³¹ The Organisation has urged its member countries to develop a 'cross-government, coherent approach' to data governance that 'reflects the critical elements for achieving system-wide benefits in government'.¹³² EU governments will need to equip their public-sectors with bespoke and devoted governance structures that will be tasked with addressing both strategic and operational challenges of G2B research data sharing. These challenges may range from delineating acceptable uses of data to monitoring whether data sharing agreements are complied with. As with the current status quo in trying to make better use of government data, the distance from principle to practice can be quite long, with organisational culture, leadership decisions and resource constraints creating complications. Nonetheless, overcoming barriers to G2B research data sharing stemming from the GDPR is certainly a necessary, even if not a sufficient, condition of building a more effective, inclusive and accountable 21st century public sector through data analytics.

VI. Conclusion: G2B research data sharing in the 21st century public-sector

While the literature is awash with metaphors of data as the new oil,¹³³ the truth is that data analysis is more than a refinery. Through data analysis, original value is created anew by the unique combination of insights that can emerge through comparative analyses of datasets via

¹²⁷ P Carter et al, 'The Social Licence for Research: Why Care.data Ran into Trouble' (2015) 41 JME <<https://jme.bmj.com/content/41/5/404>>.

¹²⁸ P Vezyridis and S Timmons, 'Understanding the care.data conundrum: New information flows for economic growth' (2017) 4(1) BDS <<https://journals.sagepub.com/doi/pdf/10.1177/2053951716688490>> 7.

¹²⁹ Waind (n 101) 3.

¹³⁰ Understanding Patient Data, 'Putting people at the centre of decisions about patient data' <<https://understandingpatientdata.org.uk/>>.

¹³¹ OECD (n 6).

¹³² *Ibid.*

¹³³ E.g. Purtova (n 19); T Gottschalk, 'The Data-Laundromat? Public-Private-Partnerships and Publicly Available Data in the Area of Law Enforcement' (2020) 6(1) European Data Protection Law Review 21.

the lens of advanced statistical methods.¹³⁴ Both with regard to its promise and its perils, it is not data *per se* that is powerful; analysis stemming from data is the real power. To harness this power and maximise its positive impact on European societies and economies, this chapter has analysed EU data protection law constraints on a key emerging type of analytical activity related to public-sector data, G2B research data sharing. More specifically, the regulatory guidance divide between the common good and private interests under article 6 GDPR was identified as a potential barrier to G2B research data sharing and an alternative, consistent with the practicalities of research data sharing and the GDPR, interpretation was put forward.

The proposed here reconciliation of the ‘public’ and the ‘private’ within lawful grounds for processing seeks to turn current interpretations of article 6 GDPR on their head. 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. Provided that such a partnership contributes to the realisation of major societal aims through the creation and curation of a high-quality evidence-basis that can inform policy design and implementation, organisational identity shall not be seen as a strictly necessary condition to contribute to the ‘public interest’. There is nothing in the GDPR to confine lawful processing in the ‘public interest’ to public bodies,¹³⁵ and it is in the best interest of EU member states to provide clarity and certainty about this matter by legislating on the acceptable analytical uses of government data and the conditions under which G2B research data sharing is allowed to take place.

Although the chapter has aspired to provide a principled basis to address legal barriers to G2B research data sharing, it is by no means suggested that this activity will always be uncontroversial from both a legal and ethical perspective. Regardless of G2B research data sharing, the question of ownership or other rights over public-sector data is becoming more topical than ever.¹³⁶ As bureaucracies across the EU increasingly rely on data scientists, systems architects and machine learning modelers to develop novel insights that can save resources and increase efficiency, much of the necessary expertise may be contracted-out or developed within public-private partnerships. Civil society organisations and international organisations have been calling for attention to the influence of powerful private actors over government data processing, highlighting the significant risks from a privacy, autonomy and non-discrimination perspective.¹³⁷

The promise of G2B research data sharing shall not make us blind to the equally significant risks stemming from its abuse. Business access to government data is, certainly, not an end in

¹³⁴ V Mayer-Schönberger, K Cukier, *Big Data: A Revolution that Will Transform how We Live, Work, and Think* (Houghton 2013); J Foreman, *Data Smart: Using Data Science to Transform Information into Insight* (Wiley 2013).

¹³⁵ According to the UK Information Commissioner’s Office, ‘Public task’ <<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/>>: ‘it is most relevant to public authorities, but it can apply to any organisation that exercises official authority or carries out tasks in the public interest’.

¹³⁶ The alternative UK, ‘Harari: Owning your data (and your mind) will beat the techno-totalitarians. Us: How local, and how mindful, must this be?’ <<https://www.thealternative.org.uk/dailyalternative/2018/9/1/harari-data-sovereignty-mindful-power>>: ‘So we had better call upon our scientists, our philosophers, our lawyers, and even our poets to turn their attention to this big question: How do you regulate the ownership of data?’.

¹³⁷ Privacy International, ‘All roads lead to Palantir’ <<https://www.privacyinternational.org/report/4271/all-roads-lead-palantir>>; Association for Progressive Communications, ‘Extreme poverty and digital welfare: New report from UN Special Rapporteur on extreme poverty raises alarm about the rise of a digital welfare dystopia’ <<https://www.apc.org/en/news/extreme-poverty-and-digital-welfare-new-report-un-special-rapporteur-extreme-poverty-raises>>.

itself. It might very well be the case that the digitalised public-sector of the 21st century would not need to rely on private assistance to produce much-needed analytical insights. Ideally, the day will come when EU governments possess the data science expertise and are able to invest the required resources to create and curate a data-driven evidence-basis by reference to which policies will be designed and implemented. In the words of a seasoned Welsh government data researcher, on that day, a wealth of government data will be ‘flowing routinely from data management systems into secure settings and out again, as aggregates, into dashboards, desktop data portals, publicly accessible interactive data websites’,¹³⁸ freeing up time for policy-makers and allowing a better targeting of public resources.

Until that day, however, the major challenges facing the EU, such as effectively containing the COVID-19 pandemic and mitigating its socio-economic harms, cannot wait. Embracing the potential of G2B research data sharing to contribute to tackling these common challenges, while theorising on the complex and often sensitive dynamics of G2B research data sharing partnerships from a legal and ethical perspective, is the way forward for the foreseeable future. The present chapter has outlined some of the key research questions that are engendered in these dynamics. Some of these questions are more general and overarching. How can we assess the contribution of a data sharing partnership to collective economic or societal welfare metrics in a realistic and reliable manner? Where should we draw the lines in G2B research data sharing, e.g. with regard to the types of private enterprise that can participate, to conform with the principles of necessity and proportionality in EU data protection law? What levels of transparency and public engagement should we require from G2B research data sharing partnerships? There are also, however, more specific research questions. What are the most scalable and robust data protection safeguards, including pseudonymisation and anonymisation techniques, that government bodies should adopt to safeguard data subjects’ rights within G2B research data sharing? For example, would particular quantitative methods like k-anonymisation offer consistency across the public sector in terms of mitigating identifiability risks? Should public bodies, rather, seek a more context-specific approach to anonymisation and pseudonymisation so as to address particularities of their specific datasets? Future EU data protection law scholars could contribute significantly to the tall task of addressing these questions, as well as expanding them in the light of modern-day developments in bureaucracies across the Union and beyond.

¹³⁸ S Lowe, ‘The value of safe settings in evidence based policy making’ <<https://zenodo.org/record/996037#.X7Zs4S0Rry9>> 1.