# On the need for additional access rights

# Enhancing access to and sharing of data: Striking the balance between openness and control over data

*Christian Reimsbach-Kounatze*[*]

## *A. Introduction*

The effective use of 'big data' and analytics (data and analytics) can help boost productivity and improve or foster new products, processes, organisational methods and markets (data-driven innovation),[1] a phenomenon which is growing in importance with artificial intelligence (AI).[2] As a result, access to data has become a critical factor for the competitiveness and success of businesses.[3] This is reflected in the growing number of mergers and acquisitions (M&As) of data-intensive firms, and most notably the acquisition of smaller, younger data-intensive firms by larger firms. Many of these M&As are motivated by strategic considerations to secure access to data.[4] Between 2013 and 2017, the annual number of acquisitions of data-intensive firms increased by a factor of four, with the average price paid exceeding USD 1 billion in some quarters.[5]

---

* The opinions expressed and arguments employed in this chapter are those of the author and should not be reported as representing the official views of the OECD or of its member countries.

1 See OECD, *Data-Driven Innovation: Big Data for Growth and Well-Being* (OECD 2015).

2 See OECD, *Artificial Intelligence in Society* (OECD 2019) 19–34.

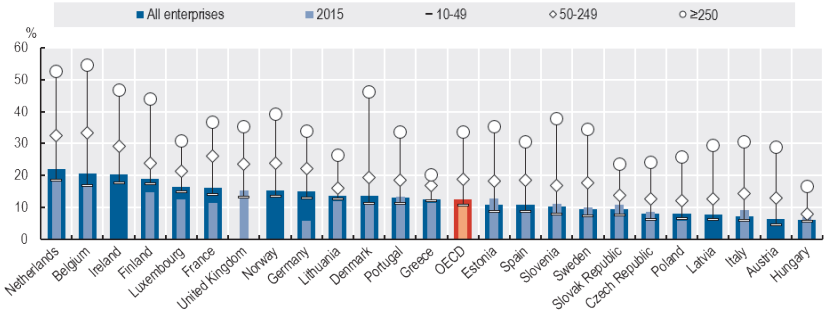
3 Firm-level studies suggest that firms that use data and analytics exhibit faster labour productivity growth than those that do not by approximately 5 % to 10 %. OECD (n. 1) 234. See, for instance, Erik Brynjolfsson and Kristina S. McElheran, 'Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing' (2019) Rotman School of Management Working Paper No. 3422397 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3422397> accessed 31 August 2020.

4 Examples include: Monsanto's acquisition of the Climate Corporation, an agriculture analytic firm, for USD 1.1 billion in 2013; Facebook's acquisition of WhatsApp for USD 14 billion in 2014; and IBM's acquisition of a majority share of the Weather Company, a weather forecasting and analytic company, for over USD 2 billion in 2015.

5 OECD, *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies* (OECD 2019) 16.

This development in M&As, which points to a tendency towards a concentration of data in favour of larger firms, is in line with trends in the use of data and analytics, where firm size is the most significant determining factor.[6] It is estimated that more than 30 % of all large firms (with 250 or more employees) in the OECD area used data and analytics in 2017 compared to only roughly 12 % of all small and medium-sized enterprises (SMEs) (with 10 to 249 employees) (Figure 1). Adoption has increased in particular among large firms in Germany, France, Finland, Korea and Portugal, although with significant variation by sector. That said, information and communication technology (ICT) firms remain the dominant users of data and analytics: more than 25 % of all ICT firms used data and analytics in the EU 28 in 2018 compared to less than 12 % of all firms (Figure 2).

*Figure 1. Business use of data and analytics by country and firm size, 2017*
*As a proportion of enterprises in each group*

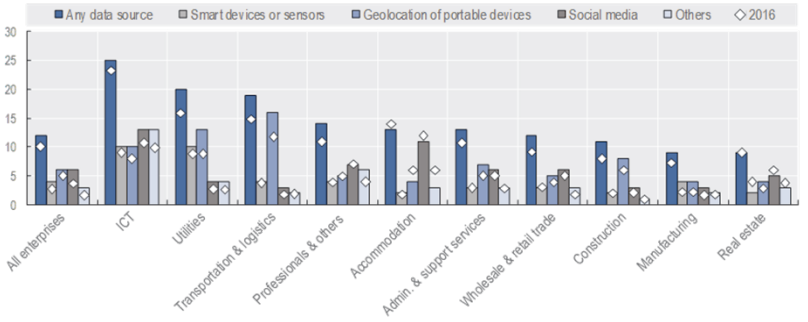



*Note*: The number of full-time employees defines the firm size. For the United Kingdom, data relate to the year 2015. OECD data figures are based on a simple average of the available OECD countries.

*Source*: OECD, *Digital Economy Outlook 2020* (OECD, 2020) 108 <https://doi.org/10.1787/888934191825> accessed 20 January 2021.

6 OECD, *Digital Economy Outlook 2020* (OECD 2020) 107-9.

*Figure 2. Business use of data and analytics by data type and industry in the EU 28, 2018*
*As a proportion of enterprises in each group*




*Source:* Ibid. 133 <https://doi.org/10.1787/888934192129> accessed 20 January 2021.

The importance of data access goes beyond the positive economic effects on productivity growth as data can also contribute directly to the well-being of citizens. Quantification however remains challenging because many if not most of the social benefits related to the use of data are poorly captured by market transactions.[7]For example, data can be shared and re-used to enhance public service delivery and to address societal needs and emergencies. Data were for instance critical for effective emergency responses during the 2011 Fukushima incident and the 2014–16 Ebola outbreak in West Africa,[8] and recently during the COVID-19 crisis.[9]

Despite the economic and social potential of data, data access and data sharing (data access and sharing), including the commercialisation of data, remain below their potential, even among data-intensive firms. In a survey by Forester Research of almost 1,300 data and analytics businesses across the globe, only a third of the respondents reported commercialising their data.[10] High tech, utilities and financial services rank among the top industries commercialising their data, while pharmaceuticals, government and

7 OECD (n. 1) 29.
8 Ibid. 335.
9 See Section D.III.2.b.
10 Jennifer Belissent, Gene Leganza and Jeremy Vale, 'Top Performers Commercialize Data Through Insights Services' (2017) <https://d3w3ioujxcalzn.cloudfront.net/item_files/206c/attachments/779835/original/forrester_infographic_top_performers_appoint_data_insights_leaders_jennifer_belissent.pdf> accessed 31 August 2020.

healthcare are at the bottom of the list. There are still significant barriers to data sharing and re-use. The risks associated with the revelation of confidential information (e.g. personal data and trade secrets) are often indicated as the main rationale for individuals and organisations not to share their data. This remains true even in cases where commercial and other private interests do not oppose data sharing.[11]

This chapter discusses how data access and sharing can be effective means for maximising the social and economic value of data, and possible venues to address related data governance challenges. Section B first introduces the theoretical foundation for understanding the social and economic potential of data, presenting data as an infrastructural resource, i.e. a general-purpose, non-rivalrous, partially excludable capital good. This functional perspective on data, which is inspired by Frischmann's work on infrastructures,[12] may seem counter-intuitive for some readers at first. However, it is helpful to better understand and explain: (i) how data can support downstream social and economic activities, (ii) why access is a key lever through which data use and value extraction can be controlled (irrespective of the existence of intellectual property rights, IPRs) and (iii) why commons present a promising solution to the collective data governance challenges of data access and sharing. Section C focusses on some of these data governance challenges.[13] Section D then presents a few promising means to address these challenges. It suggests that more differentiated data

11 AIG, 'The Data Sharing Economy: Quantifying Tradeoffs That Power New Business Models' (2016) <www.aig.com/content/dam/aig/america-canada/us/documents/brochure/the-data-sharing-economy-report.pdf> accessed 31 August 2020.

12 See Brett M. Frischmann, *Infrastructure: The Social Value of Shared Resources* (Oxford University Press 2012).

13 Other important data governance challenges had to be omitted due to space constraints, which does not mean that they were considered unimportant. These include: (i) digital security risks, (ii) liability risks in particular in respect to data quality, (iii) the risk of anti-competitive data sharing agreements (collusion), (iv) the role and limitations of data ethical frameworks, (v) the development and adoption of standards for improved interoperability, (vi) the sustainability of open data and last, but certainly not least, (vii) issues related to cross-border data access and sharing, and the interoperability of legal and regulatory frameworks affecting data access and sharing. Furthermore, issues related to IPRs, which are discussed in more detail in other chapters of this publication, are not addressed, although this chapter does discuss issues related to the concept of 'data ownership'. The same applies for issues related to privacy and data protection as well as competition regulation, which are rather superficially addressed to focus on the bigger picture, knowing that these topics are discussed in more detail in other chapters. Readers interested in the work of the Organisation for Economic Co-operation

governance approaches for data access and sharing, as implemented by data commons in combination with technological means for re-establishing control over data and information, are needed to better reflect the various interests of stakeholders and the risks they face. Section E concludes with a few public policy implications.

### *B. Data as infrastructural resource and the spillover benefits of its shared access*

The economic properties of data suggest that data may be considered as an infrastructure or, more correctly, an infrastructural resource. This may sound counter-intuitive, since traditionally infrastructures typically refer to large-scale physical facilities provided for public consumption; the classic examples are transportation systems, communication systems and basic services and facilities such as buildings and sewage and water systems. However, as for example recognised by the US National Research Council, the notion of infrastructure also refers to non-physical facilities, such as education systems and governance systems (including for example the court system).[14] This is in line with Merriam-Webster, which defines infrastructures as 'the resources (such as personnel, buildings, or equipment) required for an activity' and 'the underlying foundation or basic framework (as of a system or organization)'.[15] For Frischmann, infrastructures are 'shared means to many ends'[16] that satisfy the following three criteria:[17]

- the resource may be consumed in a non-rivalrous fashion for some appreciable range of demand (i.e. the non-rivalrous criterion);
- social demand for the resource is driven primarily by downstream productive activities that require the resource as an input (i.e. the capital good criterion); and

---

and Development (OECD) on data governance, and more specifically on data access and sharing, which has informed this work, are advised to consult the website of the OECD Working Party on Data Governance and Privacy in the Digital Economy (http://oe.cd/datagovernance) besides the relevant OECD publications referenced in this chapter.

14 National Research Council, *Infrastructure for the 21st Century: Framework for a Research Agenda* (National Academy Press 1987).

15 Merriam-Webster, 'Infrastructure' (*Merriam-Webster.com Dictionary*) <www.merriam-webster.com/dictionary/infrastructure> accessed 31 August 2020.

16 Frischmann (n. 12) 4.

17 Ibid 62–66.

- the resource may be used as an input into a wide range of goods and services, which may include private goods, public goods and social goods (i.e. the general-purpose criterion).

As discussed in the following three sections, most (though not all) data are indeed 'shared means to many ends' and satisfy these three criteria. Therefore, data can in principle be considered an infrastructural resource.

### *I. Data as a non-rivalrous although partially excludable good*

(Non-)rivalry of consumption describes the degree to which the consumption of a resource affects (or does not affect) the potential of the resource to meet the demands of others. It thus reflects the marginal cost of allowing an additional consumer of the good. A rivalrous good such as oil can only be consumed once. A non-rivalrous good such as knowledge, in contrast, can be consumed in principle an unlimited number of times. This property is the source of significant spillovers and that provides the major theoretical link to total factor productivity growth enabled by data, but it also raises questions about how best to allocate data as a resource.

While it is widely accepted that social welfare is maximised when a rivalrous good is consumed by the person who values it the most, and that the market mechanism is generally the most efficient means for rationing such goods and for allocating resources needed to produce such goods, this is not always true for non-rivalrous goods. The situation is more complex in this case, since non-rivalrous goods come with an additional degree of freedom with respect to resource management: Social welfare is maximised not when the good is consumed solely by the person who values it the most, but when everyone who values it consumes it.[18] Maximising access to the non-rivalrous good will thus in theory maximise social welfare, as every additional private benefit comes at no additional cost.

However, data are not always and in every circumstance non-rivalrous. Some data can lose their value as soon as e.g. illegitimate users have access to them. This would be the case, for instance, when the data contain confidential information, such as trade secrets, and/or could be misused for insider trading. Although the data in these cases are not depleted due to the illegitimate use, the information they contain may lose its economic value, at least for those that wish to protect the information. In other words, in

18 Ibid 28.

these particular cases the consumption of data would affect their potential to meet the demands of (a few) others.[19]

The fact that data may not always be perfectly non-rivalrous does not invalidate the non-rivalrous criterion for data to be considered an infrastructure however. This is because: (i) public goods theory recognises that there are few perfectly non-rivalrous goods[20] and, more specifically, (ii) infrastructures can most of the time be consumed in non-rivalrous fashion only within some appreciable range of demand, for instance, when they are not congested.

It is important to note at this point that non-rivalry of consumption of data does not imply that data are a public good. A public good must satisfy an additional criterion: besides being non-rivalrous, the good must also be non-excludable, i.e. the cost for one person to prevent another from consuming the resource must be significant. While the marginal costs of transmitting, copying and processing data can be close to zero, making it easy for others to reproduce and use them, ICTs including for e.g. user access control and encryption[21] have also dramatically reduced the costs of exclusion. Thus, where data are kept within a controlled environment the cost of exclusion will be typically low enough to prevent others from using them. This is why data can be considered at least partially excludable and not a public good.

### *II. Data as a capital good with increasing returns to scale and scope*

Data are still sometimes described as 'the new oil' of the digital economy. However, besides the non-rivalrous nature of data, data are neither a consumption good such as an apple, nor an intermediate good such as oil. In most cases, data should be classified as a capital good.

---

19 See David W. Opderbeck, 'Socially Rivalrous Information: Of Candles, Code, and Virtue' (2007) Seton Hall Public Law Research Paper No. 1008500, 85 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1008500> accessed 31 August 2020, who refers to this particularity as 'social rivalry' with the following argument: 'Because information helps construct communities, those who possess information possess a form of power. Sharing information diminishes the power held by the person who previously restricted access to the information. Information therefore is socially rivalrous.' This argument applies to certain types of information and thus only to certain types of data and this only under certain conditions.

20 See John G. Head, 'Public Goods and Public Policy' in Charles K. Rowley (ed.), *Readings in Industrial Economics*, Vol. II (MacMillan Publishers 1972) 66.

21 See Section D.III.

While consumption goods are consumed to generate direct benefits to the consumer or firm,[22] intermediate goods and capital goods are used as inputs to produce other goods. Intermediate and capital goods are both means rather than ends, and their demand is driven by the demand for the derived outputs. They are thus factors of production.

The difference between intermediate and capital goods is that, while intermediate goods such as raw materials (e.g. oil) are used up, exhausted, or otherwise transformed when used as input to produce other goods, capital goods are not.[23] Furthermore, capital goods 'must have been produced as outputs from processes of production', which explains why 'natural assets such as land, mineral or other deposits, coal, oil, or natural gas, or contracts, leases and licences' are not considered capital goods.[24]

Data, in most cases, are used as an input for goods or services; this is especially true of large volumes of data (i.e. 'big data'), which are means rather than ends in themselves. They are however not an intermediate good, as they are not exhausted when used, given their non-rivalrous nature. This does not mean that data cannot be discarded after they have been used. In many cases, they may be used just once. However, storage costs today have decreased to the point where data can generally be kept for long periods of time, if not indefinitely. This has increased data's capacity to be used as a capital good and production factor.

Furthermore, being a capital good does not mean that data do not depreciate. The value of most capital goods declines 'as a result of physical deterioration, normal obsolescence or normal accidental damage'.[25] In the case of data, depreciation is more complex, however, because it is context-dependent. That is, the value of data depends on the context of their use.[26] Therefore, the relevance, accuracy and timeliness of data will typically af-

22 EC and others, 'System of National Accounts 2008' (2008) 179 <http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf> accessed 31 August 2020.

23 Ibid. 120. See also Eurostat, 'NACE Rev. 2 – Introductory Guidelines' (2006) 39 <https://ec.europa.eu/eurostat/documents/1965800/1978839/NACEREV.2INTRODUCTORYGUIDELINESEN.pdf/f48c8a50-feb1-4227-8fe0-935b58a0a332> accessed 31 August 2020.

24 EC and others (n. 22) 123.

25 Ibid.

26 See OECD, 'Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value' (OECD 2013) OECD Ditigal Economy Papers No. 220 <www.oecd-ilibrary.org/docserver/5k486qtxldmq-en.pdf?expires=1604224239&id=id&accname=guest&checksum=90171D5AA0F516519D87E5DD30974A07> accessed 31 August 2020, which shows that assessing the value of data *ex ante* (before use) is almost impossible, because the information derived is context-

fect that value. Data can depreciate, for instance, when they begin to lose their relevance for an intended use. This explains, for example, why there is a temporal premium for 'real-time' data in the financial sector.

The capital-good nature of data has major economic implications. As data are a non-rival capital, multiple users can in theory use them (simultaneously) for multiple purposes (see next section) as an input to produce an unlimited number of goods and services. In addition, increasing returns to scale and scope are possible as the value of data increases when the data can be linked with, and integrated into, a (larger) big data set.[27] In practical terms, these properties find their application in data-enabled multi-sided markets, i.e. economic platforms in which distinct user groups generate benefits (externalities or spillovers) to other groups.[28] In other words, the re-use of data enables multi-sided markets in which huge returns to scale and scope can lead to positive feedback loops on one side of the market in favour of the business, which in turn reinforces success in the other side(s) of the multi-sided market, overall leading to a potential '"winner takes all" outcome in which monopoly is the nearly inevitable outcome of market success'.[29]

### *III. Data as general-purpose but context-dependent input*

Infrastructures are not inputs that have been optimised for a special limited purpose, but 'they provide basic, multipurpose functionality'.[30] In par-

---

dependent: data that are of good quality for certain applications can thus be of poor quality for other applications. Furthermore, the information and thus value that can be extracted from data is not only a function of the data, but also a function of the (analytic) capacity to link data and to extract insights. This capacity is determined by available (meta-)data, analytic techniques and technologies; however, it is also a function of pre-existing knowledge and skills. See also OECD (n. 1).

27 See Bertin Martens, 'Data access, consumer interests and social welfare: An economic perspective on data' (2020) in this publication.

28 See Jean-Charles Rochet and Jean Tirole, 'Two-sided Markets: A Progress Report' (2006) 37 RAND Journal of Economics 645, defining two- or multi-sided markets 'roughly ... as markets in which one or several platforms enable interactions between end users and try to get the two or multiple sides "on board" by appropriately charging each side'.

29 OECD, *Supporting Investment in Knowledge Capital, Growth and Innovation* (OECD 2013) 170.

30 Frischmann (n. 12) 65.

ticular, infrastructures make possible a wide range of private, public and social goods, which users are free to produce according to their capabilities.

How data are used will typically depend on the initial purpose for which they have been collected. For example, at the outset agricultural data will primarily be used for agricultural goods and services. However, in theory, there are no limits with regard to the purposes for which data can be re-used, and many of the benefits stemming from their re-use are based on the fact that data created in one domain can provide further insights when applied in another domain. A clear illustration is provided by open public-sector data, where data sets used originally for administrative purposes are re-used by entrepreneurs to create services unforeseen when the data were originally created. Another example is the use of anonymised mobile call data records (CDRs) of telecommunications services providers that have been re-used to monitor and control the spread of pandemics such as COVID-19.[31]

The general-purpose nature of infrastructure comes with a key policy implication. The production of (*ex ante* unforeseeable) public and social goods via the infrastructure could lead to the market failure of insufficient provision of the infrastructure, which would call for government intervention in some cases. This is because 'users' willingness to pay [for the infrastructure] reflects private demand – the value that they expect to realise – and does not take into account [the social] value that others might realise as a result of their use.'[32] Where this social value is difficult to measure, a 'demand-manifestation problem' can occur, which in turn may lead to an undersupply of the infrastructure and a 'prioritisation of access and use of the infrastructure for a narrower range of uses than would be socially optimal'.[33] This is why, as a consequence, there can be significant (social) opportunity costs in limiting access to infrastructures. In other words: open (closed) access enables (restricts) user opportunities and degrees of freedom in the downstream production of private, public and social goods, many of which by their nature have significant spillover effects. Non-discriminatory access can therefore be an optimal (private and public) strategy for maximising the benefits of an infrastructure, in particular in environments characterised by high uncertainty, complexity and dynamic changes.

31 See Section D.III.2.b.
32 Frischmann (n. 12) 66.
33 Ibid.

This means that markets through which data are commercialised may not be able to fully serve social demand for data where such a demand manifestation problem would occur. Although the data demand manifestation problem remains poorly documented in literature, there are plausible reasons to believe that such a problem may occur in praxis, in particular, when data are used for the production of social or public goods such as to increase transparency in government or to combat poverty and pandemics. In addition, the context dependency of data and the highly uncertain, complex and dynamic environment in which some data are used (e.g. research) make it almost impossible to fully evaluate ex ante the potential of data, which further exacerbates the demand manifestation problem.

The next section briefly summarises the finding of available empirical studies that assess to what extent non-discriminatory access, and open access to data (open data) particularly, in the public[34] and private[35] sector can create social and economic spillover benefits.

---

34 See Office of Fair Trading, 'The Commercial Use of Public Information' (2006); ACIL Tasman, 'The Value of Spatial Information: The Impact of Modern Spatial Information Technologies on the Australian Economy' (2008); ACIL Tasman, 'Spatial Information in the New Zealand Economy: Realising Productivity Gains' (2009); Graham Vickery, 'Review of Recent Studies on PSI Reuse and Related Market Developments' (2011); Graham Vickery, 'Review of Recent Studies on PSI Reuse and Related Market Developments' (2012); Deloitte, 'Market Assessment of Public Sector Information: A Report to the Department for Business, Innovation and Skills' (2013); Deloitte, 'Assessing the Value of TfL's Open Data and Digital Partnerships' (2017).

35 See McKinsey Global Institute, 'Open Data: Unlocking Innovation and Performance with Liquid Information' (2013) <www.mckinsey.com/business-functions/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information#> accessed 31 August 2020; Nicholas Gruen, John Houghton and Richart Tooth, 'Open for Business: How Open Data Can Help Achieve the G20 Growth Target' (2014) <www.academia.edu/attachments/55569779/download_file?st=MTYwNDIyNDM1OSwxOTMuMTc0LjEzMi42NA%3D%3D&s=swp-splash-paper-cover> accessed 31 August 2020; Nomura Research Institute, 'Research on Spillover Effects of Evolution in Operation and Services by Data Use on the Economy and Society' (2014); Mitsubishi Research Institute, 'De-Ta Ryutuu Purattofo-Mu Ni Kannsuru Tyo-Sajigyo [Study on Platforms for Data Sharing]' (2017); IDC and Lisbon Council, 'The European Data Market Monitoring Tool' (2019).

## IV. Empirical evidence of the spillover social and economic benefits of data access and sharing

Although the quantification of the overall benefits remains challenging, available evidence strongly suggests that non-discriminatory access, including in particular open data, generates positive social and economic benefits for data providers (direct impact), their suppliers and data users (indirect impact), and for the wider economy (induced impact). This is thanks to: (i) greater transparency, accountability and empowerment of users, for instance when open data are used for (cross-subsidising) the production of public and social goods; (ii) new business opportunities, including for the creation of start-ups and in particular for data intermediaries and mobile application (app) developers; (iii) competition and co-operation within and across sectors and nations, and including the integration of value chains, (iv) crowdsourcing and user-driven innovation and (v) increasing efficiency thanks to linkages of data across multiple sources.[36]

The magnitude of the relative effects will vary however depending on the sector (public vs. private sector) and the type of effect. Studies show that, while non-discriminatory access to data can increase the value of data to holders (direct impact), it can help create 10 to 20 times more value to data users (indirect impact), and 20 to 50 times more value for the wider economy (induced impact). In some cases, however, non-discriminatory data access and sharing, and in particular open data, may also reduce the producer surplus of data holders, which is the cause of the incentive problem discussed in Section C.II.[37] Overall, these studies suggest that non-discriminatory data access and sharing can help generate social and economic benefits worth between 0.1 % and 1.5 % of GDP in the case of public sector data, and between 1 % and 2.5 % of GDP when also including private sector data.[38]

---

36 OECD (n. 5) 64–71.

37 See for instance Office of Fair Trading, which surveyed more than 400 public sector information holders (PSIHs) and 300 businesses buying or licencing data from PSIHs. It estimates that the producer surplus of the PSIHs (of around GBP 66 million p.a.) would have vanished with open access in favour of an increase of the indirect impact (including the consumer surplus of PSI re-use) by GBP 585 million p.a.

38 OECD (n. 5) 59–64.

### C. Major data governance challenges of data access and sharing

Data access and sharing through non-discriminatory regimes not only come with social and economic benefits. They also come with risks to individuals and organisations. These may include the risks of confidentiality and privacy breaches, but also the violation of other legitimate private interests such as commercial interests. The pursuit of the benefits of data access and sharing therefore needs to be balanced against the costs and the legitimate national, public and private interests, in compliance with legislations concerning relevant rights and obligations of the stakeholders involved (such as on the protection of privacy and IPRs). This is in particular the case where sensitive data are involved. Otherwise incentives to contribute data and to invest in data-driven innovation may be undermined, in addition to the risks of direct and indirect harm to right holders (including data subjects).

Evidence confirms that risks of confidentiality breach, for instance, have led users to be more reluctant to share their data, including providing personal data and in some cases even in using digital services such as cloud computing.[39] Furthermore, inappropriate sharing of data can lead to significant costs to the organisation, including fines due to privacy violations as well as opportunity costs due to a lower ability to innovate. For example, it has been noted that sharing data prematurely can undermine the ability to obtain IPRs (e.g. on patents and trade secrets).[40]

This section discusses some major data governance challenges that need to be considered to facilitate data access and sharing. These include: (i) risks of violating private (commercial) interests including the interest in the protection of privacy and IPRs, which go hand in hand with the increasing loss of control of individuals and organisations over their data, which in turn is rooted in the partial excludability of data highlighted in Section B.I; (ii) the need to incentivise data sharing and data-related investment in light of the externalities associated with data sharing and the risk of 'free riding'; and (iii) the need to address uncertainties related to 'data ownership', a concept used to re-establish control over data.

---

39 OECD, *OECD Digital Economy Outlook 2017* (OECD 2017) 252–54.

40 Jorge L, Contreras, 'Data Sharing, Latency Variables, and Science Commons' (2010) 25 Berkeley Technology Law Journal 1601; Michael W. Carroll, 'Sharing Research Data and Intellectual Property Law: A Primer' (2015) 13 PLOS Biology e1002235 <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002235> accessed 31 August 2020.

## I. The loss of control over data, and the risk of violation of privacy and intellectual property rights

Violations of privacy and to some extent of IPRs are often considered as the biggest risks associated with data access and sharing. These include most notably risks of violating (contractual and socially agreed) terms of data re-use, and thus risks of acting against the (reasonable) expectations of users and the law. This is true with respect to individuals (data subjects), their consent and their privacy expectations, but also with respect to organisations and their contractual agreements with third parties and the protection of their commercial interests.

### 1. Violations of agreed terms and of expectations in data re-use

Even where individuals and organisations agree on (and consent to) specific terms for data sharing and data re-use, including on the purposes for which the data should be re-used, there remains a significant level of risk that the data may end up being used differently by a third party.[41] The violation of these terms may not be always the result of malicious intentions. The contextual change that results from transferring data from one context to another will always make it challenging to ensure that existing rights and obligations are not undermined. This is because assumptions and expectations that were implicit in the initial usage (and context) may no longer apply in subsequent uses, an observation that is coherent with the fact that information derived from data is context-dependent, as highlighted in Section B.III., and so are the risks associated with data sharing.[42]

---

41 The case of Cambridge Analytica illustrates this risk: personal data of Facebook users ended up being used, not for academic purposes as some users had consented to, but for a commercially motivated political campaign, and this although Facebook explicitly prohibits data from being sold or transferred 'to any ad network, data broker or other advertising or monetisation-related service'. Kevin Granville, 'Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens' *The New York Times* (19 March 2018) <www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html> accessed 31 August 2020.

42 This is also in line with Nissenbaum's theory of privacy as 'contextual integrity', according to which adequate privacy protection is tied to the norms of the specific context in which data are used, 'demanding that information gathering and dissemination be appropriate to that context and obey the governing norms of distri-

Some of these concerns and risks have been framed as ethical, to underscore the need to recognise the importance of issues such as fairness, respect for human dignity, autonomy, self-determination, the risk of bias and discrimination as complementary to regulatory actions. Data ethics is highlighted in particular in cases where the collection, processing and sharing of data will be legal under existing law, but may generate moral, cultural and social concerns with potential direct or indirect adverse impacts on individuals or social groups. There are expectations that ethics may thus provide an additional promising venue in particular in light of the loss of control over data and, in particular, the role of consent as discussed below. This, however, raises additional issues such as the risks that some approaches to data ethics might be perceived or (mis-)used as a substitute (i) for full compliance with regulations, or (ii) for a thorough assessment and mitigation of ethical concerns ('ethic washing').

### *2. Loss of control over data and the role of consent*

As highlighted in Section B.I., the cost of excluding others from using data will typically be low enough for the original data holder if the data are kept within a controlled environment. However, once the data are shared, unless specific data stewardship and processing provisions are in place, those data move out of the control of the original data holder.[43] The same can be said to be true for individuals, who provide their data and give their consent for their re-use and sharing. In both situations, data holders and individuals lose their capabilities to control how their data are re-used and to object to or (technically) oppose such uses, and can rely solely on law enforcement and redress. The risks of loss of control are multiplied where the data are further shared downstream across multiple tiers, in particular when these tiers are located across multiple jurisdictions.

Consent has been highlighted as a major, although not always effective, mechanism to allow individuals control the collection and (re-)use of their

---

bution within it'. Helen Nissenbaum, 'Privacy as Contextual Integrity' (2004) 79 Washington Law Review 119.

43 See Smitha Sundareswaran, Anna Squicciarini and Dan Lin, 'Ensuring Distributed Accountability for Data Sharing in the Cloud' (2012) 9 IEEE Transactions on Dependable and Secure Computing 556; Martin Henze, René Hummen, Roman Matzutt and Daniel Catrein, 'Maintaining User Control While Storing and Processing Sensor Data in the Cloud' (2013) 5 International Journal of Grid and High Performance Computing 97.

personal data. It requires clear provision of information to individuals about what personal data are being collected and used, and for what purpose – as specified in the data protection and privacy laws of most countries and in the OECD Privacy Guidelines.[44] To assure the maximum level of flexibility in compliance with privacy legislations, some organisations have come to rely, however, on one-time general or broad consent as the basis for data collection, use and sharing. One-time general consent can be used to achieve an appropriate balance between participant rights to determine the future use of their personal data, but only under the condition that data subjects are given reasonable means to extend or withdraw their consent over time. In addition, general consent models have been criticised for posing ethical challenges as data subjects may not realise the full implications of giving a broad consent, particularly in the context of AI and big data.[45]

### *II. Incentivising data sharing in light of positive externalities and the risk of 'free riding'*

While the marginal costs of transmitting, copying and processing data can be close to zero, substantial investments will often be required to collect data and to enable data sharing and re-use. As highlighted in the introduction (Section A), firms are investing a significant share of their capital in the acquisition of start-ups to secure access to data potentially critical for their business. Investments may also be needed for data cleaning as well as data curation,[46] which is often beyond the scope and time frame of the ac-

---

44 OECD, 'Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data' (1980) OECD/LEGAL/0188 (OECD Privacy Guidelines).

45 New consent models have been proposed in the scientific literature, including 'adaptive' or 'dynamic' forms of consent to address these concerns. These also include time-restricted consent models, where individuals consent to the use of their personal data only for a limited period. These models typically enable participants to consent to new projects or to alter their consent choices in real time as their circumstances change and to have confidence that these changed choices will take effect. See Jane Kaye, Edgar A. Whitley and others, 'Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks' (2015) 23 European Journal of Human Genetics 141.

46 Data curation embodies data management activities necessary to assure long-term data quality across the data life cycle.

tivities for which the data were initially collected and used.[47] Furthermore, the investments required for effective data access and sharing are not limited to data themselves. In many cases, complementary investments are needed in meta-data, data models and algorithms for data storage and processing, and to secure ICT infrastructures needed for (shared) data storage, processing and access. The overall total upfront costs can therefore be very high.

Given these significant investment requirements, data holders may not necessarily have the incentives to share their data, in particular if the costs (risks) of data access and sharing are perceived to be higher than the expected (private) benefits. In other words, where organisations and individuals cannot recuperate a sufficient level of the return on their data-related investments, for instance through revenues arising from granting data access against licence fees, there is a high risk that data sharing will not occur at a sufficient level.

The root cause of this incentive problem can be attributed to a positive externality and the risk of 'free riding': data access and sharing may benefit others more than it may benefit the data holder, who may not be able to privatise all the benefits of data re-use. Empirical evidence of the spillover social and economic benefits of data access and sharing presented in Section B.IV. suggests indeed that non-discriminatory access to data, even though it can help increase the value of data to data holders (direct impact), will tend to create 10 to 20 times more value to data users (indirect impact), and 20 to 50 times more value for the wider economy (induced impact). In some cases, however, it may even reduce the producer surplus of data holders, as in the case of open data.

Since data are only partially excludable goods for which the costs of exclusion will tend to be high *once* the data move out of the control sphere of the original data holders, there is the possibility that some may 'free ride' on the data holders' investments. The argument that follows is that if data are shared, free-riding users can 'consume the resources without paying an adequate contribution to investors, who in turn are unable to recoup their investments'.[48] This would then lead to a disincentive to share and, where

47 OECD, 'Research Ethics and New Forms of Data for Social and Economic Research' (2016) OECD Science, Technology and Industry Policy Papers No. 34 <www.oecd-ilibrary.org/docserver/5jln7vnpxs32-en.pdf?expires=1604226871&id=id&accname=guest&checksum=8CD30C2A2939BD7217A6AA3693115FC> accessed 31 August 2020.

48 Frischmann (n. 12) 9.

data access and sharing would be made mandatory, to a disincentive to invest in data in the first place.

However, the assumption that positive externalities and free riding always diminish incentives to invest cannot be generalised, and needs careful case-by-case scrutiny. In this regard, Frischmann notes:

> There is a mistaken tendency to believe that any grain or loss in profits corresponds to an equal or proportional gain or loss in investment incentives, but this belief greatly oversimplifies the decision-making process and underlying economics and ignores the relevance of alternative opportunities for investment.[49]

In addition, free riding is sometimes the economic and social rationale for providing access to data. Open data initiatives, for example, are motivated by the recognition that users *will* free ride on the data, and in so doing will be able to create a wide range of new goods and services that were not anticipated and would not otherwise be produced. In this sense, 'free riding is … a feature, rather than a bug of our economic, cultural, and social systems'.[50] However, even though the externality and the risk of 'free riding' associated with data access and sharing may not necessarily lead to a loss in investment incentives, it may still lead to a loss in incentives to share data and thus to a dysfunctional 'data sharing economy', where only a few market participants are willing to share their data.

### *III. 'Data ownership' as an attempt to regain control over data*

Granting private property rights is often suggested as a solution to the incentive problems related to free riding and, in the case of intangible assets, to address the risk of loss of control that comes with their non-excludability. The often raised question about who 'owns the data' is therefore essentially motivated by the recognition that ownership rights provide a 'powerful basis for control'[51] as 'to have legal title and full property rights to

49 Ibid 161.
50 Ibid.
51 Teresa Scassa, 'Data Ownership' (2018) CIGI Paper No. 187 <www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf> accessed 31 August 2020.

something'[52] implies 'the right to exclusive use of an asset'[53] and the 'full right to dispose of a thing at will'.[54]

In contrast to the concept of ownership of physical goods, where the owner typically has exclusive rights and control over the good – including for instance the freedom to destroy the good – this is not the case for intangibles such as data. For these types of goods, IPRs are typically suggested as the legal means to establish clear ownership. While some have expressed the opinion that, in principal, data cannot be owned,[55] 'the ability to access, create, modify, package, derive benefit from, sell or remove data, but also the right to assign these access privileges to [and retrieve them from] others'[56] – what could be defined as the main rights associated with 'data ownership' – are affected by different legal frameworks differently. IPRs, in particular copyright and trade secrets, can be applicable under certain conditions. In certain jurisdictions, cyber-criminal law may have the effects of conferring ownership-like rights to data holders, while 'data ownership' related questions emerging between firms can also be regulated by competition law.[57] And in the case of personal data, privacy protection frameworks will be relevant for the question of 'data ownership' as well. Thus, in contrast to other intangibles, data typically involve complex assignments of different rights across different stakeholders and are governed by multiple overlapping legal and regulatory frameworks.

### *1. 'Ownership' of personal data*

The complexity of the overlapping legal and regulatory frameworks is particularly apparent when it comes to personal data, which is also the topic where the debate on 'data ownership' seems the most controversial. There seems to be a general belief among many individuals that they 'own' or should 'own' their personal data. The reality, in many, if not most, juris-

52 Malcolm Chisholm, 'What Is Data Ownership?' (*BeyeNetwork* May 2011) <www.b-eye-network.com/view/15697>.

53 Lothar Determann, 'No One Owns Data' (2018) UC Hastings Research Paper No. 265 <https://ssrn.com/abstract=3123957> accessed 31 August 2020.

54 Ibid.

55 Ibid.

56 See David Loshin, 'Who Owns Data?' (*DM Review* March 2003) <http://knowledge-integrity.com/columns/dmr200303.htm>.

57 Osborne Clarke LLP, *Legal Study on Ownership and Access to Data* (European Commission 2016) <https://publications.europa.eu/en/publication-detail/-/publication/d0bec895-b603-11e6-9e3c-01aa75ed71a1> accessed 31 August 2020.

dictions, is that they do *not* legally 'own' their personal data. Data (including personal data) collected by an organisation will typically be considered the property of that organisation. For example, in Canadian case of *McInerney v. McDonald*,[58] 'one of the theories considered, and ultimately rejected, by the court was that a patient owned their personal medical information'.[59] Instead, the court found that the 'physician, institution or clinic compiling the medical records owns the physical records'.[60]

However, in the case of personal data, the 'ownership' rights of the organisation will hardly be comparable to other (intellectual) property rights. As Scassa concludes in regard to *McInerney v. McDonald*, 'the court also recognised an "interest" on the part of the patient amounting to a degree of control over the information.'[61] In fact, most privacy regulatory frameworks give data subjects particular control rights over their personal data, which may interfere with 'the full right to dispose of a thing at will'[62] typically associated with ownership. So in the particular case of personal data, no single stakeholder will have exclusive access and use rights. Different stakeholders will typically have different powers depending on their role.[63] Some authors have therefore stressed that privacy protection frameworks may have some characteristics like those of a property right.[64] As the EU General Data Protection Regulation (GDPR) has extended the control

58 *McInerney v. MacDonald*, 1992 CanLII 57 (SCC), [1992] 2 SCR 138, <http://canlii.ca/t/1fsbl> accessed 10 May 2020.

59 Scassa (n. 51).

60 Ibid.

61 Ibid.

62 Determann (n. 53).

63 See Fred Trotter, 'Who Owns Patient Data? Look inside Health Data Access and You'll See Why "Ownership" Is Inadequate for Patient Information' (*O'Reilly Radar* 2012) <http://radar.oreilly.com/2012/06/patient-data-ownership-access.html> accessed 31 August 2020. The author highlights that in the case of health patient data all stakeholders (including patient, doctor and programmer) 'have a unique set of privileges that do not line up exactly with any traditional notion of 'ownership'. Ironically, it is neither the patient nor the [doctor] who is closest to 'owning' the data. The programmer [or platform] has the most complete access and the only role with the ability to avoid rules that are enforced automatically by electronic health record (EHR) software.'.

64 See Nadezhda Purtova, 'Do Property Rights in Personal Data Make Sense after the Big Data Turn? Individual Control and Transparency' (2017) 10(2) Journal of Law and Economic Regulation 64; Josef Drexl, 'Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy' in Alberto Di Franceschi and Reiner Schulze (eds), *Digital Revolution – New Challenges for Law* (2019) 19; Francesco Banterle, 'The Interface Between Data Protection and IP Law: The Case of Trade Secrets and the Database Sui Generis Right in Marketing

rights of individual to the right of data portability (Article 20), the similarities have become even stronger.

### 2. *Contractual arrangements and the role of contract guidelines and model contracts for data sharing*

The discussion above could suggest, and some have suggested, that multiple 'owners' (with co-ownership rights) will have to be assumed, 'as neither the "data producer" nor the "data gatherer" can claim an exclusive right over the data'.[65] As will be further discussed in Section D.II., multiple stakeholders are often involved in the contribution, collection and control of data, including in particular the data subject himself or herself when it comes to personal data. These stakeholders therefore typically have some interests in the data, and the challenge is how to disentangle and address the multiple (potential) interests of the various stakeholders – including the public interest in data – in a way that is compliant with law and aligned with societal values and objectives (see Section D.II.1.). This, combined with the 'intricate net of existing legal frameworks',[66] may explain current controversies and uncertainties related to 'data ownership', a challenge which is exacerbated where data are created, accessed and shared across jurisdictions.

As a response to this situation, businesses have come to rely on contract law as the primary legal vehicle for determining rights related to data control, access and re-use, in particular in business-to-business (B2B) contexts. These contractual arrangements often can better suit the individual context of data access, sharing and use (freedom of contract). While freedom of contract may give stakeholders the ability to construct well-suited contractual arrangements, existing uncertainties may also increase transaction costs, and expose particularly those that are in a weaker position to negotiate fair terms and conditions. These are typically individuals (consumers)

---

Operations, and the Ownership of Raw Data in Big Data Analysis' in Mor Bakhoum and others (eds), *Personal Data in Competition, Consumer Protection and Intellectual Property Law* (2018) 411.

65 Paul Hofheinz and David Osimo, 'Making Europe a Data Economy: A New Framework for Free Movement of Data in the Digital Age' (Lisbon Council Policy Brief 2017) <https://lisboncouncil.net/wp-content/uploads/2020/08/LISBON-COUNCIL-Making-Europe-A-Data-Economy.pdf> accessed 31 August 2020.

66 Determann (n. 53).

and SMEs.[67] As a result, incentives to share data, including with third parties, remain low and, where data-sharing arrangements are negotiated, they may be perceived as potentially unfair. In addition, the high transaction costs of negotiating fair terms and conditions may prevent the commercialisation of data as a commodity (via data marketplaces).

To address the issues highlighted above, some governments[68] and private sector actors[69] are providing guidance and/or model contracts for data-sharing agreements, including contractual clauses that are based on commonly agreed principles. These clauses constitute the default position that parties can consider when negotiating their data-sharing agreements.[70] Because they are voluntary, parties are free to deviate from the proposed contractual clauses at their will.[71] However, such deviation would have to be justifiable, which is why contract guidelines and model contracts are seen as promising means to assure fair terms and conditions for data access, sharing and re-use, in particular where there are significant power and in-

67 See the conflict between farmers and agriculture technology providers that led in the United States to AG Data Transparent, 'Ag Data's Core Principles: The Privacy and Security Principles for Farm Data' (2016) <www.agdatatransparent.com/principles/> accessed 31 August 2020. Similarly in Europe: COPA-COGECA and CEMA, 'EU Code of Conduct on Agricultural Data Sharing by Contractual Agreement' (2018) <https://copa-cogeca.eu/img/user/files/EU%20CODE/EU_Code_2018_web_version.pdf> accessed 31 August 2020.

68 For example, Japan's Ministry of Economy, Trade and Industry has formulated the 'Contract Guidance on Utilisation of AI and Data', which summarises the issues and factors to be considered when drafting a contract on the utilisation of AI or data. It is intended to be used as a reference when private businesses conclude contracts related to data re-use or development and use of AI-based software. See METI, 'METI Formulates "Contract Guidance on Utilization of AI and Data"' (2018) <www.meti.go.jp/english/press/2018/0615_002.html> accessed 10 August 2020.

69 In July 2019, Microsoft published three data-sharing agreements to be used as a template: (i) the Open Use of Data Agreement (O-UDA), (ii) the Computational Use of Data Agreement (C-UDA) and (iii) the Data Use Agreement for Open AI Model Development (DUA-OAI). These were complemented by a fourth one in November 2019: (iv) the Data Use Agreement for Data Commons (DUA-DC). See Microsoft, 'Removing Barriers to Data Innovation: Empowering People and Organizations to Share and Use Data More Effectively' (2019) <https://3er1viui9wo30pkxh1v2nh4w-wpengine.netdna-ssl.com/wp-content/uploads/prod/sites/560/2019/12/Backgrounder-FAQ-Sheet_FINAL.pdf> accessed 31 August 2020.

70 See European Commission, 'Guidance on Sharing Private Sector Data in the European data economy' SWD(2018) 125 final.

71 It is expected that parties would do so if such deviation better reflected their common interests and the specific context of their data-sharing agreements.

formation asymmetries between parties. Because they are based on agreed principles and refer to applicable national and international laws, contract guidelines and model contracts are expected also to reduce (legal) transaction costs.

However, while these guiding documents can help address the legal uncertainties and also to some extent the power and information asymmetries that may exist between business partners, they may fall short in addressing other important data governance issues where data access and use rights are concerned. This is most notably the case in respect to third parties with an interest in the data but who do not take part in the negotiations of the data-sharing agreements (e.g. consumers, social groups and the public). Therefore, guidelines and model contracts are not a silver bullet solution to maximise the economic and social benefit of data, although they are a promising additional tool in the data governance toolbox to facilitate data sharing.

### *D. Towards a more differentiated data governance approach for data access and sharing*

This section presents possible solutions to the data governance challenges highlighted in Section C., or at least contributions to such solutions. A common cause of these challenges is the loss of control over data, which is rooted in the partial excludability of data, as mentioned several times already. The following section therefore looks at (i) technological means for re-establishing control over data and information, which are promising complementary solutions to legal measures (including IPRs) to help address the risks and challenges discussed in Sections C.I. and C.II. The next section then presents (ii) a data taxonomy for disentangling the various interests in data that can help address some of the challenges related to 'data ownership' discussed in Section C.III. This includes in particular the need to clarify the respective contributions of the potential 'multiple owners' in the data-enabled value creation process, as well as the potential interests of all relevant stakeholders, including the public. The last section briefly looks at (iii) data commons as data-sharing arrangements with variable degrees of openness and control and as a framework solution through which the other solutions discussed before could be implemented.

## I. Technological means for re-establishing control over access to data and information

The following sections provide an overview of technological means for re-establishing control over data, many of which are known as privacy-enhancing technologies (PETs). PETs are typically used to prevent and mitigate the risk of privacy and confidentiality breaches and to enable organisations to better manage data responsibly. These technologies thus make it possible to balance the respective interests of stakeholders, namely by enabling data access and use, while protecting the respective rights of stakeholders, including the privacy rights of data subjects.[72] They are also promising means to deliberately adjust the level of data quality (e.g. level of aggregation and timeliness) and thus to control the information and value of data that are shared. These technological means are clustered in two groups: (i) technologies used for data access control and (ii) those traditionally known as PETs, used to protect privacy and confidentiality, and thus to control access to the information. The latter are referred to as 'confidentiality-enhancing technologies'.

### 1. Data access control mechanisms

There are a wide range of different mechanisms for accessing and sharing data within and across organisational borders. The following three can be considered the most commonly used:[73] data access via (i) (ad hoc) downloads, (ii) application programming interfaces (APIs) and (iii) data sandboxes. These mechanisms are typically implemented in cloud-based solutions, which give data holders an additional level of control, to the extent that the data remain within the same cloud service platform.

#### a) (Ad hoc) downloads

In the case of data access via downloads, the data are stored, ideally in a commonly used format, and made available online (e.g. via a web site). Da-

72 Alessandro Acquisti, 'The Economics of Personal Data and the Economics of Privacy' (2010) <www.oecd.org/sti/ieconomy/46968784.pdf> accessed 31 August 2020.

73 OECD (n. 5) 32–34.

ta access via downloads however raises several issues: Interoperability, for instance, is a major one when it comes to data re-use across applications (e.g. data portability), and one which is not guaranteed even when commonly used machine-readable formats are used.[74] Furthermore, ad hoc downloads fail to address the risk of loss of control over data. Therefore, the provision of data via ad hoc downloads typically comes with significant digital security and privacy risks, as data holders lose their capabilities to enforce any data policies including in respect to the protection of privacy and IPRs.

b) Application programming interfaces (APIs)

APIs enable service providers to make their digital resources (e.g. data and software) available over the Internet. They facilitate the interoperability of the different actors and their technologies and services. A key advantage of APIs (compared to an ad hoc data download) is that APIs enable a software application to directly use the data it needs. Data holders can also implement several restrictions via APIs to better control the use of their data including means to assure data syntactic and synthetic portability.[75] Furthermore, they can help control the identity of API users, the scale and scope of the data used (including over time), and even the extent to which the information derived from the data could reveal sensitive or personal information. APIs thus provide moderate, although in many cases sufficient, control over data.

c) Data sandboxes for trusted access and re-use of sensitive and proprietary data

The term 'data sandbox' is used to describe any isolated environment through which data are accessed and analysed, and analytic results are only

74 These formats may enable *data syntactic portability*, i.e. the transfer of 'data from a source system to a target system using data formats that can be decoded on the target system'. But they do not guarantee *data semantic interoperability*, 'defined as transferring data to a target such that the meaning of the data model is understood within the context of a subject area by the target', ibid. 32.

75 See text at n. 74.

exported, if at all, when they are non-sensitive.[76] These sandboxes can be realised through technical means (e.g. isolated virtual machines that cannot be connected to an external network) and/or through physical on-site presence within the facilities of the data holder (where the data can be accessed). Data sandboxes would typically require that the data processing code is executed at the same location as where the data are stored. Compared to the other data access mechanisms presented above, data sandboxes offer the strongest level of control. Data sandboxes are therefore promising for providing access to very sensitive/personal and proprietary data including across borders.[77]

### 2. *Confidentiality-enhancing technologies for information access control*

#### a) Cryptography

Cryptography is a practice that 'embodies principles, means, and methods for the transformation of data in order to hide its information content, establish its authenticity, prevent its undetected modification, prevent its repudiation, and/or prevent its unauthorised use'.[78] It is increasingly used by businesses and for consumer goods and services to protect the confidentiality of data, such as financial or personal data, whether those data are in storage or in transit.[79] A number of innovative cryptography-based applications for data access and sharing are emerging. These include for instance:

- *Homomorphic encryption*, which makes it possible to run computations on encrypted data whilst protecting privacy and confidentiality. As a re-

76 See Royal Society, 'Privacy Enhancing Technologies: The Current Use, Development and Limits of Privacy Enhancing Technologies in Data Analysis' (2019) 25–28 <https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/privacy-enhancing-technologies-report.pdf> accessed 31 August 2020, where they are referred to as 'trusted execution environments.'.

77 See for example the Virtual Research Data Center (VRDC) of the the Centers for Medicare and Medicaid Services (CMS). ResDAC, 'CMS Virtual Research Data Center (VRDC)' <www.resdac.org/cms-virtual-research-data-center-vrdc> accessed 12 August 2020. See also Henze and others (n. 43); Yves-Alexandre de Montjoye, Sébastien Gambs, Vincent Blondel and others, 'On the Privacy-Conscientious Use of Mobile Phone Data' (2018) 5 Scientific Data No. 180886 <www.nature.com/articles/sdata2018286> accessed 31 August 2020.

78 OECD, 'Recommendation of the Council concerning Guidelines for Cryptography Policy' (1997) OECD/LEGAL/0289 (OECD Crypto Guidelines).

79 OECD (n. 39) 270–73.

sult, a user can for example encrypt data, send it to the cloud for processing and have the results of the computation sent back to him or her or to third parties, all the while maintaining the privacy of the individual concerned and confidentiality of the data.[80]

- *Blockchain or distributed ledger technologies (DLTs)*, which enable decentralised solutions for the storage and management of data to address some of the challenges related to data control and trust. Instead of relying on a centralised operator, a blockchain relies on a distributed network of peers to maintain and secure a decentralised database. What is significant for trust in data access and sharing is not only the fact that blockchains are highly resilient and tamper-resistant,[81] which enables, for instance, robust auditing of the actual usage of data.[82] In addition, blockchains, via e.g. 'smart contracts',[83] can facilitate the management of access control permissions, including for the licensing of data access and use rights.[84]

b) De-identification: from anonymisation to pseudonymisation and aggregation

De-identification covers a range of practices ranging from anonymisation to pseudonymisation and aggregation. These practices share a common aim of preventing the extraction of identifying attributes (i.e. re-identification), or at least significantly increasing the costs of re-identification. Anonymisation is a process in which an individual's identifying information is excluded or masked so that the individual's identity cannot be, or

80 See Royal Society (n. 76) 21–25.

81 Once data have been recorded on the decentralised data store, they cannot be deleted subsequently or modified by any single party.

82 See Sundareswaran, Squicciarini and Lin (n. 43).

83 'Smart contracts' are self-executing decentralised applications based on blockchain that can initiate trackable and irreversible transactions based on predefined conditions. See Mayukh Mukhopadhyay, *Ethereum Smart Contract Development: Build Blockchain-Based Decentralized Applications Using Solidity* (Packt Publishing 2018).

84 See Andreas Muelder, 'Model-Driven Smart Contract Development for Everyone' (*Hacker Noon* 2019) <https://hackernoon.com/model-driven-smart-contract-development-for-everyone-jiu32p0> accessed 31 August 2020; Yongkai Fan, Jinghan Wang, Zhenting Hong and others, 'A Blockchain-Based Data-Sharing Architecture' in Zibin Zheng and others (eds), *Blockchain and Trustworthy Systems* (Springer Singapore 2020) 636.

becomes too costly to be, reconstructed.[85] Some research has shown however that when linked with other data, most anonymised data can be de-anonymised – that is, the identifying information can be reconstructed.[86]

Where stronger protection is required or desired, additional means – including 'noise' addition, functional separation and distribution (decentralisation) and administrative and legal safeguards – are needed. The addition of 'noise' to a data set, for instance, allows analysis based on the complete data set to remain significant while masking sensitive data attributes.[87] Work on 'differential privacy' is one prominent example in which noise is added to the data so that when a statistic is released, information about an individual is not revealed with it.[88] For many applications, however, identifiers are needed because complete anonymity would be useless, preventing for instance two-way communication and transactions. In these cases, pseudonymisation offers a solution whereby the most identifying attributes (i.e. identifiers) within a data record are replaced by unique artificial identifiers (i.e. pseudonyms). Finding the right balance that protects privacy and confidentiality, while minimising the costs to data utility, thus remains a challenge for all these means.

### *II. A data taxonomy for disentangling the various interests in data*

Data are often treated as a monolithic entity, although evidence shows that data are heterogeneous goods whose value depends on the context of their use. A more differentiated approach to the governance of data is therefore needed, and this requires identifying the different types of data and their characteristics.

The following sections present two major dimensions that are considered critical for the governance of data access and sharing. These include:

---

85 Andreas Pfitzmann and Marit Hansen, 'A Terminology for Talking about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management' (2010) <https://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf> accessed 31 August 2020; Kato Mivule, 'Utilizing Noise Addition for Data Privacy, an Overview' (2013) <https://arxiv.org/ftp/arxiv/papers/1309/1309.3958.pdf> accessed 31 August 2020.

86 See de Montjoye and others (n. 77). See also Arvind Narayanan and Vitaly. Shmatikov, 'How to Break Anonymity of the Netflix Prize Dataset' (2006) <https://arxiv.org/pdf/cs/0610105.pdf> accessed 31 August 2020.

87 See Mivule (n. 85).

88 Cynthia Dwork and Aaron Roth, 'The Algorythmic Foundations of Differential Privacy' (2014) 9 Foundations and Trends in Theoretical Computer Science 211.

(i) the *domain of the data*, which describes whether there are potentially personal (individual), private (business) and/or public (societal) interests associated with a particular dataset, and the relevant legal and regulatory regimes; and (ii) *the manner in which data originate*, which reflects the level of awareness and control that various data stakeholders, including data subjects, data holders and data users, can have, and therefore, most importantly, the type of contribution of the various potential 'multiple owners' in the data-enabled value creation process.

### *1. The overlapping domains of data – reflecting the various stakeholder interests*

Besides the dichotomy between personal and non-personal data, the most frequent distinction made in policy debates is between private and public sector data. It is generally accepted and expected, for example, that public sector data in contrast to private sector data should be made available through open data, free of charge and free of any restrictions from IPRs – where there are no conflicting national security or private interests. However, the private–public data distinction often made is not only blurred since public and private sector data cannot always be distinguished,[89] but more importantly, because it risks distracting attention from the related contentious issues highlighted in Section C and further explained below. A differentiation between the following three domains of data is suggested therefore instead (Figure 3):

- *the personal domain*, which covers all data 'relating to an identified or identifiable individual'[90] (personal data) for which data subjects have an interest in privacy,
- *the private domain*, which covers all proprietary data that are typically protected by IPRs or by other access and control rights (provided by

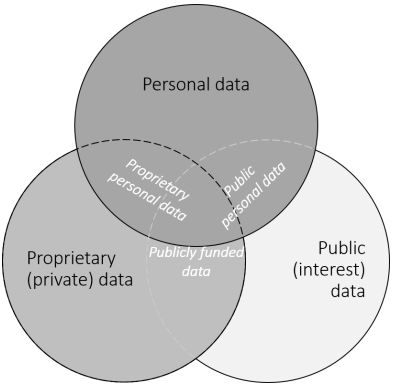
89 Data can often qualify as both public sector *and* private sector data, and this irrespective of any joint activities between public and private sector entities (e.g. public-private partnerships). For instance, data generated, created, collected, processed, preserved, maintained or disseminated by the private sector that are however funded by or for the public sector would qualify as both public sector *and* private sector data. Compare this with the definition of public sector information in OECD, 'Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information' (2008) OECD/LEGAL/0362 (OECD PSI Recommendation).

90 OECD Privacy Guidelines (n. 44).

e.g. contract, cyber-criminal or competition law), and for which there is typically an economic interest to exclude others, and

- *the public domain*, which covers all data that are not protected by IPRs or any other rights with similar effects, and therefore lie in the 'public domain' (understood more broadly than to be free from copyright protection), thus free to access and re-use, as well as data for which there is a public interest.

*Figure 3. The personal, private and public domain of data*




*Source:* OECD, *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies* (OECD 2019) 29.

These three domains not only overlap as illustrated in Figure 3, but they are also typically subject to different data governance and legal frameworks that can affect each domain differently. For instance, privacy regulatory frameworks typically govern the personal domain, while the private domain is typically governed through e.g. contractual, IPR and/or competition regulatory frameworks. The overlaps may partly explain the potential conflicting views and interests of some stakeholder groups, as reflected for instance in issues related to 'personal data ownership' discussed in Section C.III.[91]

Furthermore, these overlaps can explain why data governance is often perceived as complex from a legal and regulatory perspective, in particular when cross-border data flows are concerned. Depending on the jurisdic-

91 See also the call for collaboration between competition, privacy and consumer protection authorities as discussed, for instance, in OECD (n. 1) 109.

tion, some domains may be prioritised differently over others, and this difference in prioritisation seem to reflect differences in culture and legal systems. This is for example reflected in current privacy and data portability rights, which vary significantly across countries, reflecting various approaches to balancing the conflicting interests of individuals and organisations over 'proprietary personal data' (Figure 3). In the case of data portability, which aims to empower individuals and give them more control rights over their personal data, what type of data falls within the scope of data portability varies across initiatives, partly reflecting the (implicit) priorities of the personal v the proprietary domain.[92]

Another area where data governance has proven to be particularly challenging, and where the 'domains of data' could help make more explicit some of the prevailing tensions, is when public interest in data is concerned, in particular when such interest 'overlaps' with the interests in proprietary and personal data (the centre of the Venn diagram in Figure 3). A few countries have started to specify a new class of data, which is often referred to as *data of public or general interest*.[93] The scope of this class varies significantly across countries however. In some countries, data of public interest explicitly refers to private sector data (of public interest), while in others it refers to public sector data. Sometimes both private and public sector data, as well as personal and non-personal data, are included. What they have in common though is that they seem to refer to data needed to fulfil more or less well-defined societal objectives that otherwise would be impossible or too costly to fulfil. These objectives can include the development of national statistics, the development and monitoring of public policies, the tackling of health care and scientific challenges of societal importance and in some cases the provision of public services.[94]

92 See Louisa Specht-Riemenschneider, 'Data access rights – A comparative perspective', in this volume.

93 See Loi n° 2016–1321 du 07 octobre 2016 pour une République numérique 2016 (Law for a digital Republic). It defines 'data of general interest' (données d'intérêt général) as including: (i) private sector data from delegated public services such as utility or transportation services, (ii) private sector data that are essential for granting subsidies and (iii) private sector data needed for national statistics.

94 See Heiko Richter, 'The law and policy of government access to private sector data ("B2G data sharing")', in this volume.

*2. The manner data originate – reflecting the contribution to data creation*

Multiple stakeholders are often involved in the contribution, collection and control of data, including the data subject in the case of personal data. The data categories discussed above – in particular the distinction between the personal domain and the proprietary domain – however do not help differentiate how different stakeholders contribute to data co-creation. The following data categories that differentiate according to the way data are collected or created can provide further clarity in this respect.[95]

- *Volunteered (or surrendered or contributed or provided) data* are data provided by individuals when they explicitly share information about themselves or others. Examples include creating a social network profile and entering credit card information for online purchases.
- *Observed data* are created where activities are captured and recorded. In contrast to volunteered data, where the data subject is actively and purposefully sharing its data, the role of the data subject in the case of observed data is rather passive and it is the data controller that plays the active role. Examples of observed data include location data of cellular mobile phones and data on web usage behaviour.
- *Derived (or inferred or imputed) data* are created based on data analytics, including data 'created in a fairly "mechanical" fashion using simple reasoning and basic mathematics to detect patterns'.[96] In this case, it is (only) the data processor that plays the active role in the creation of data. The data subject typically has little awareness of what is inferred about her or him, given in particular that personal information can be derived from several pieces of seemingly anonymous or non-personal

95 They are based on Bruce Schneier, 'A Taxonomy of Social Networking Data' (*Schneier on Security*, 2009) <www.schneier.com/blog/archives/2009/11/a_taxonomy_of_s.html> accessed 31 August 2020; Martin Abrams, 'The Origins of Personal Data and Its Implications for Governance' (2014) <https://ssrn.com/abstract=2510927> accessed 31 August 2020; Productivity Commission of Australia, 'Productivity Commission Inquiry Report: Data Availability and Use' (2017) <www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf> accessed 31 August 2020.

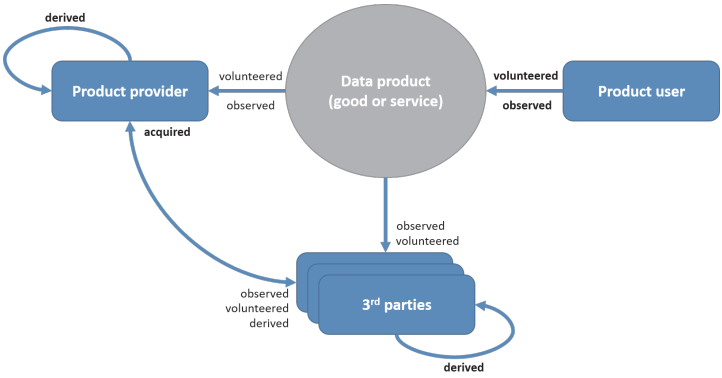
96 OECD, 'Summary of OECD Expert Roundtable Discussion on "Protecting Privacy in a Data-Driven Economy: Taking Stock of Current Thinking"' (2014) DSTI/ICCP/REG(2014)3, 5 <www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/iccp/reg(2014)3&doclanguage=en> accessed 31 August 2020.

data.[97] Examples of derived data include credit scores calculated based on an individual's financial history.

- *Acquired (purchased or licenced) data* are obtained from third parties based on commercial (licencing) contracts (e.g. when data are acquired from data brokers) or other non-commercial means (e.g. when acquired via open government initiatives). As a result, contractual and other legal obligations may affect the re-use and sharing of these data.

To illustrate the moment of creation of the different types of data, Figure 4 presents a stylised process in which a product user (e.g. a consumer) would interact with a data product (e.g. an online service or portable smart device) that is provided by a product provider (e.g. a business). The data product would typically (i) observe the activities of its users, in which case *observed data* are created; and/or (ii) be used to input data volunteered by its users (*volunteered data*). The data could then be accessed for further processing (and the creation of *derived data*) by the product provider as well as by any third parties who may have been granted direct or indirect access to the original (volunteered and observed) data – or in a less identifiable form. The creation of derived data can also be enriched when combined with *acquired data* from (other) third parties.

*Figure 4. Data products and the different manners data originate*




*Note*: Arrows represent potential data flows between the different actors and a data product (good or service). The type of data is highlighted in bold to indicate the moment at which the data are created.

*Source:* Ibid. 31.

97 See Narayanan and Shmatikov (n. 86).

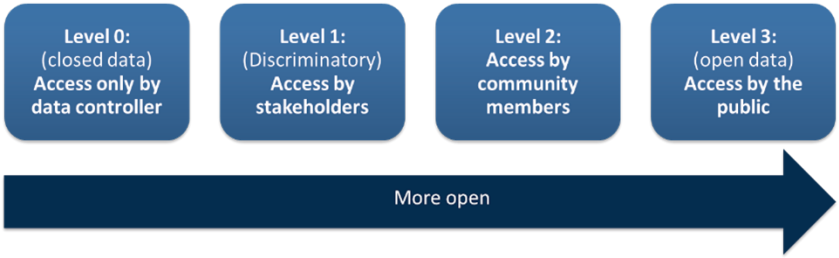
This differentiation is relevant for the governance of data for several reasons: (i) it helps determine the level of awareness that product users (including data subjects) can have about the scale and impact of data collection, which is critical, for example, when assessing the privacy risks associated with data collection and the level of control product users can be expected to have. (ii) It also reflects the contribution of various stakeholders to data creation and therefore their potential rights and interests in accessing and re-using the data. (iii) Last, but not least, this differentiation can help identify the geographic location and jurisdiction based on data generation and collection, and it can therefore help determine the applicable legal and regulatory frameworks.

### *III. Data commons as arrangements with variable degrees of openness and control*

The infrastructural nature of data as non-rivalrous, general-purpose productive capital, combined with the spillover benefits of data re-use and the demand manifestation problem, suggest that maximising access to data, for instance through open data, will in theory maximise social welfare if every additional private benefit comes at no additional cost. The latter condition is not always true, however, given the risks and challenges highlighted in Section C and the resulting need for more controlled data access. This calls for more differentiated approaches to data access and sharing along the data openness continuum illustrated in Figure 5.[98]

98 This continuum suggests that data access and sharing should not be seen as a 'binary concept' (opposing closed to open data), but rather a continuum of different degrees of data openness, ranging from internal access and re-use (only by the data holder), to restricted (unilateral and multilateral) external access and sharing, and open data as the most extreme form of data openness.

*Figure 5. The degrees of data openness and access*




*Source:* OECD, *Data-Driven Innovation: Big Data for Growth and Well-Being* (OECD 2015) 185.

The most prominent approach to non-discriminatory data access and sharing discussed in the literature and by policy makers is still *open data*. Besides open data, a wide range of other approaches exist, with different degrees of data openness that respond to the various interests of stakeholders and the risks they face in data sharing. Many of these approaches are based on voluntary and mutually agreed terms between organisations, while others are mandatory such as the right to data portability under Article 20 GDPR or Australia's recently proposed Consumer Data Right (CDR),[99] which are both non-discriminatory in respect of the data subject's intent of data re-use.

This section first introduces the concept of 'data commons' and argues that many of the approaches to data access and sharing highlighted above can be seen as a special form of data commons, whereby the relevant community varies, ranging from the public at large such as in the case of open data (level 3 in Figure 5), to the members of a community such as in the case of data partnerships (level 2), or the data subject and data controller such as in the case of data portability (level 1). Due to space constraints, the section then only focusses on restricted data-sharing arrangements (level 2), given that the other approaches to data sharing are well documented in the literature, if not in this publication.

99 See ACCC, 'Consumer Data Right (CDR)' <www.accc.gov.au/focus-areas/consumer-data-right-cdr-0> accessed 31 August 2020.

## 1. Data commons for the governance of shared resources of common interests

The concept of 'commons' has been used to describe natural resources that are managed and used for collective benefits, as well as the governance mechanisms (including informal norms and values) affecting their consumption.[100] Commons are therefore defined as collective goods in which stakeholders have common interests, and which are characterised by the governance mechanisms surrounding their production and consumption.

Applied to data, commons imply formal or informal governance institutions to enable the sustainable shared production and/or use of data. Data commons can therefore be defined as the institutionalised sharing of data among members of a community. Although the concept of data commons has been used most prominently for public sector open data – which may explain why data commons sometimes is misunderstood, and used as a synonym for 'open data' – the community within which data sharing is institutionalised may, but does not necessarily need to, include the public at large (as in the case of open data). Commons will for instance emerge where data are not, or no longer, privately 'owned' by a single entity, but rather considered a collective resource of common interest within a community that requires common management and governance institutions, such as in data partnerships discussed below.

In data commons, data access by community members is often granted, independent of their identity and their intent of use (non-discriminatory access). This does not imply that access must be free, unregulated or without any terms and conditions. The important point here is that non-discriminatory access can maximise the value of data, while the scope of access is essentially defined by the scope of the community. The positive spillovers in combination with non-discriminatory access can lead to Rose's 'comedy of the commons',[101] where greater social value is created with greater use of data, in contrast to Hardin's 'tragedy of the commons',[102] where free riding on common (natural) resources leads to the degradation and the depletion of the resources.

---

100 See Charlotte Hess and Elinor Ostrom (eds), *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press 2007); Michael J. Madison, 'Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo' in Brett M. Frischmann, Michael J. Madison and Kathrine J. Strandburg (eds), *Governing Knowledge Commons* (Oxford University Press 2014) 209.

101 Carole Rose, 'The Comedy of the Commons: Custom, Commerce, and Inherently Public Property' (1986) 53 University of Chicago Law Review 711.

102 Garret Hardin, 'The Tragedy of the Commons' (1968) 162 Science 1243.

That said, the establishment of data commons can be quite complex as it involves a number of considerations such as on community definition, institutional design, the relevant regulatory framework, boundaries and exclusion of non-members, pricing and congestion management to assure the sustainability of the commons, and in some cases even considerations on exceptions from the non-discrimination rule.[103] The complexity of the governance is exacerbated by the fact that commons, and knowledge commons in particular, are often clustered in multiple ways as well as nested within each other.[104] Many knowledge communities (e.g. scientific communities) define, and in turn are defined by, the knowledge commons. Patient-related commons, for instance, are often nested together with scientific knowledge commons, and infrastructure and digital-tools-related commons (e.g. software and data). The understanding, establishment and support of commons therefore require systematic analysis of all relevant contextual factors.[105]

### *2. Restricted data-sharing arrangements*

In cases where data are considered too confidential to be shared openly with the public (as open data) or where there are legitimate (commercial and non-commercial) interests opposing such open sharing, restricted data-sharing arrangements can be more appropriate. This is for instance the case when there may be privacy, IPR (e.g. copyright and trade secrets) and organisational or national security concerns legitimately preventing open

---

103 See Frischmann (n. 12) 92, who highlights that 'exceptions [to non-discrimination] arise in many contexts – for reasons of emergency […] or securing the commons itself. In some cases, sustaining a resource as a commons requires narrowly tailored exceptions to address specific, identifiable uses that degrade, deplete, or otherwise harm the resource itself or risk harm to the community of users. That such exceptions exist in some contexts for certain types of infrastructure resources does not undermine the basic nondiscrimination rule, as long as the exceptions do not swallow the rule.'.

104 See Brett M. Frischmann, Michael J. Madison and Kathrine J. Strandburg (eds), *Governing Knowledge Commons* (Oxford University Press 2014).

105 See Elinor Ostrom and Charlotte Hess, 'A Framework for Analyzing the Knowledge Commons' in Charlotte Hess and Elinor Ostrom (eds) *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press 2007) 41. The institutional analysis and development (IAD) framework has been used for the systematic analysis of knowledge commons. See also Frischmann, Madison and Strandburg (n. 104).

sharing of data. In these cases, however, there can still be a strong economic and/or social rationale for sharing data among data users within a restricted community, under voluntary and mutually agreed non-discriminatory terms.

It is, for example, common to find restricted data-sharing agreements in areas such as digital security (e.g. for vulnerability disclosure), science and research (e.g. for health care research) and as part of business arrangements for shared resources (e.g. within joint ventures). These voluntary data-sharing arrangements can be based on commercial or non-commercial terms depending on the context. Two types of data-sharing arrangements are highlighted in the following sections in more detail: (i) *data partnerships*, which are based on the recognition that data sharing can provide not only significant economic benefit to data users, but also to data holders; and (ii) *data for societal objectives* initiatives, where data are shared to support societal objectives.

### a) Data partnerships

In data partnerships, organisations agree to share and mutually enrich their data sets, including through cross-licensing agreements. One big advantage is the facilitation of joint production or co-operation with suppliers, customers (consumers) or even potential competitors (co-opetition).[106] This also enables data holders to create additional value that a single organisation would not be able to create and provides opportunities 'to join forces *without* merging'.[107] Examples include:[108]

106 It is worth noting that the concept of data partnerships offers some similarities with other concepts known in the world of IPRs, most notably patent pools, which are essentially agreements between two or more patent holders to license one or more of their patents to one another or third parties. See WIPO, 'Patent Pools and Antitrust – a Comparative Analysis' (2014) <www.wipo.int/export/sites/www/ip-competition/en/studies/patent_pools_report.pdf> accessed 31 August 2020.

107 Benn R. Konsynski and Warren F McFarlan, 'Information Partnerships – Shared Data, Shared Scale' (1990) 68 Harvard Business Review 114.

108 Other benefits include the ability to: (i) maximise the *option value* of data (i.e. value of keeping the options for irreversible investments open); and (ii) (cross-)subsidise public and social goods, which otherwise would require picking winners (users or applications). See OECD (n. 1) 177.

- The pooling of aggregated data between Take Nectar, a UK-based programme for loyalty cards, and collaborating firms such as Sainsbury (groceries), BP (gasoline) and Hertz (car rentals) to 'allow … the three companies to gain a broader, more complete perspective on consumer behaviour, while safeguarding their competitive positions'.[109]
- The joint venture between DuPont Pioneer and John Deere, which was initiated in 2014 with the aim to develop a joint agricultural data tool.[110]

Similar partnerships also exist in the form of public and private partnerships (data PPPs). For example, the sharing of data (including through open data) with major Internet platform providers such as Google, Waze, Twitter and Apple enabled Transport for London (TfL), a local government body responsible for the transport system in Greater London (United Kingdom), to gain access to new data and that was used to improve its business operation and services.[111]

Data partnerships (including data PPPs) however raise several challenges, some of which are similar to those highlighted in Section C.III. For instance, ensuring a fair data-sharing agreement between the partners can sometimes be challenging, in particular where partners have different levels of market power. Privacy and IPR considerations may also limit the potential of data partnerships by making it harder to sustain data sharing in some cases. Where data partnerships involve competing businesses, data sharing may increase the risk of (implicit) collusion including the formation of cartels and fixing of price. In the case of data PPPs, there may also be some challenges due to the double role of governments, namely as an authority on one hand and service (data) provider on the other.

---

109 Michael Chui, James Manyika and Steve Van Kuiken, 'What Executives Should Know about Open Data' (*McKinsey&Company* 2014) <www.mckinsey.com/industries/high-tech/our-insights/what-executives-should-know-about-open-data> accessed 31 August 2020.

110 See Russ Banham, 'Who Owns Farmers' Big Data?' Forbes (*Forbes*, 8 July 2014) <www.forbes.com/sites/emc/2014/07/08/who-owns-farmers-big-data/> accessed 31 August 2020.

111 See Deloitte, 'Assessing the Value of TfL's Open Data and Digital Partnerships' (2017) <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf> accessed 31 August 2020.

b) Data for societal objectives

Data-sharing arrangements can also be found where private sector data are provided (donated) to support societal objectives, ranging from science and health care research to policy making. For instance, in an era of declining responses to national surveys, the re-use of private sector data can significantly improve the power and quality of statistics, in particular in developing economies.[112] The re-use of private sector data also provides new opportunities to better inform public policy making, for instance, when close to real-time evidence is made available to 'nowcast' policy-relevant trends.[113] In some initiatives, private sector data have been provided, for instance, to address urgent societal objective including during disasters and health crises including pandemics such as COVID-19.[114]

For example, mobile telecommunications services providers in a few countries have started to share geolocation data based on CDRs with governments in an aggregated, anonymised format. As these network operators serve substantial portions of the population across entire nations, they can measure movements of millions of people at fine spatial and temporal scales in near-real time. The resulting information is used by governments seeking to track the COVID-19 outbreak, warn vulnerable communities and understand the impact of policies such as social distancing and confinement. The European Commission, for instance, has been liaising with

112 See Christian Reimsbach-Kounatze, 'The Proliferation of "Big Data" and Implications for Official Statistics and Statistical Agencies: A Preliminary Analysis' (2015) OECD Digital Economy Paper No. 245 <https://www.oecd-ilibrary.org/science-and-technology/oecd-digital-economy-papers_20716826> accessed 31 August 2020.

113 See Hyonyoung Choi and Hal Varian, 'Predicting the Present with Google Trends' (2009) <https://static.googleusercontent.com/media/www.google.com/de//googleblogs/pdfs/google_predicting_the_present.pdf> accessed 31 August 2020; Derrick Harris, 'Hadoop Kills Zombies Too! Is There Anything It Can't Solve?' (*Gigaom* 18 April 2011) <https://gigaom.com/2011/04/18/hadoop-kills-zombies-too-is-there-anything-it-cant-solve/> accessed 31 August 2020; Yan Carrière-Swallow and Felipe Labbé, 'Nowcasting with Google Trends in an Emerging Market' (2013) 32 Journal of Forecasting 289.

114 OECD, 'Ensuring Data Privacy as We Battle COVID-19' (14 April 2020) <www.oecd.org/coronavirus/policy-responses/ensuring-data-privacy-as-we-battle-covid-19-36c2f31e/> accessed 31 August 2020; OECD, 'Tracking and Tracing COVID: Protecting Privacy and Data While Using Apps and Biometrics' (23 April 2020) <www.oecd.org/coronavirus/policy-responses/tracking-and-tracing-covid-protecting-privacy-and-data-while-using-apps-and-biometrics-8f394636/> accessed 31 August 2020.

European telecommunications operators to obtain from them anonymised aggregate mobile geolocation data, and to coordinate measures tracking the spread of COVID-19 at EU level.[115]

### *E. Conclusion*

This chapter highlighted one of the major tensions that policy makers and practitioners face when dealing with data governance issues, namely the tension between the social benefits of 'data openness' on one hand, and individuals' and organisations' risks and legitimate concerns over such openness on the other hand. This tension is rooted in the economic properties of data presented in Section B, most notably (i) their non-rivalrous nature, which calls for maximum openness (data sharing) to leverage the potential spillover benefits of data as a productive capital (Section B.IV) and (ii) their partial excludability, which comes with the risk of loss of control (Section C.I), which in turn can disincentivise data sharing (Section C.II). The discussion in this chapter suggested that granting private 'data ownership' rights was *not* the silver bullet solution. The fact that data are partly excludable however opens the possibility for a wide range of alternative or complementary solutions presented in Section D. Their combination through data commons arrangements, such as for instance data partnerships, promises to address many, if not most, of the challenges highlighted in Section C.

Besides the need for more research and development (R&D) in technological means for enhancing controlled data sharing (including PETs), a major policy recommendation that results from this chapter is the need for guidance on data commons. As highlighted in Section D.III.1., the establishment of data commons can be quite complex as it requires a careful analysis of all the contextual factors relevant for data sharing. While the institutional analysis and development (IAD) framework developed by Nobel Prize laureate Elinor Ostrom is promising from a research perspective, practitioners, in particular SMEs, may require practical guidelines to be able to establish and take advantage of data commons. In view of countries' experiences on contract guidelines and model contracts for data shar-

115 See Commission Recommendation (EU) 2020/518 of 8 April 2020 on a common Union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data [2020] OJ L114/7.

ing, it would seem desirable that similar guiding documents with a focus on data commons be provided. That the private sector, most notably Microsoft,[116] is moving in this direction is therefore a promising development.

116 See the Data Use Agreement for Data Commons (DUA-DC) (n. 69).