

1. Introduction

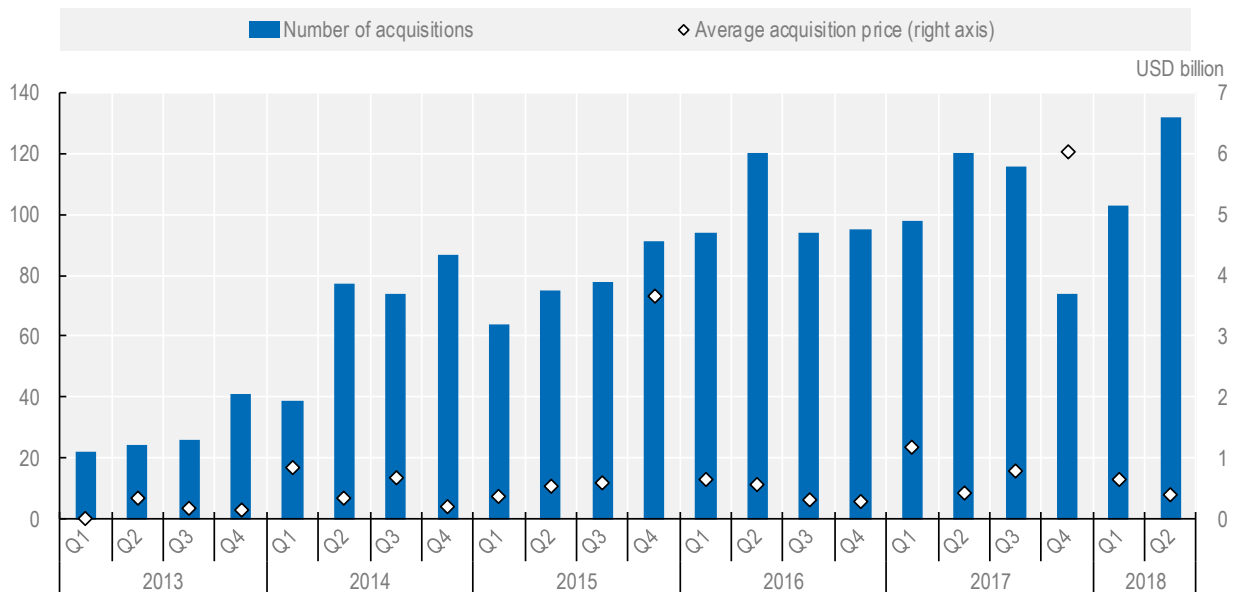
This introductory chapter explains the importance of data access and sharing in the context of current technological developments. It points to barriers to data access, sharing and re-use and some of the issues further discussed in the following chapters. It concludes by presenting the objective of the report and an overview of its structure.

The effective use of data,¹ in combination with data analytics (software), generates information of social and economic value. It can help boost productivity and improve or foster new products, processes, organisational methods and markets. This is referred to as “data-driven innovation” (DDI) (OECD, 2015^[1]). With the increasing use of artificial intelligence (AI) and the Internet of Things (IoT) the supply of, and demand for, data will increase even in traditionally less data-intensive fields, and this to a level that very few organisations will be able to meet alone. A single self-driving car, for example, can generate between 1 terabyte (TB) and 5 TB of data per hour according to some estimates (Grzywaczewski, 2017^[2]; Nelson, 2016^[3]).² Yet, even more third-party data are required for these systems to operate securely irrespective of weather conditions, visibility or road-surface quality.³

Access to data is thus crucial for competition and innovation in the digital economy. Not only for businesses, but also for governments and individuals, including researchers. Access to data can, for instance, help enhance public-service delivery and facilitate the identification of emerging governmental and societal needs. It can help improve forecasting and the reliability of infrastructures (such as in transportation and utilities) and increase their efficiency. In science and research access to data can help review and replicate scientific results, and foster new instruments and methods of data-intensive exploration and scientific experimentation.

The economic importance of data access is reflected in the growing number of mergers and acquisitions (M&As) of data-intensive firms. These M&As are meant to assure access to business-critical data. Some of the largest M&As motivated by access to big data in the last five years include: Monsanto’s acquisition of the Climate Corporation, an agriculture analytic firm, for USD 1.1 billion in 2013; IBM’s acquisition of a majority share of the Weather Company, a weather forecasting and analytic company, for over USD 2 billion in 2015 (Waters, 2015^[4]); and Alibaba’s total investment of USD 4 billion between 2016 and 2018 to acquire Lazada, a leading e-commerce platform founded in 2012 in Singapore. Start-ups specialised in big data are also increasingly the target of acquisitions. The annual number of these acquisitions increased from more than 100 acquisitions in 2013 to more than 400 acquisitions in 2017, with the average price paid exceeding USD 1 billion in some quarters (Figure 1.1).

Figure 1.1. Trends in the acquisition of big data and analytics firms



Source: OECD based on Crunchbase data.

These developments underline the growing social and economic value of data, and the need for access to and sharing of data. The intangible and non-rivalrous nature of data allows a wide range of means of access and sharing, including the commercialisation of data under non-discriminatory conditions and open data. In contrast to rivalrous goods such as oil, which is depleted once extracted, transformed and consumed during production processes, the use of data does not exhaust the supply of data and (therefore) *in principle* its potential to meet the demands of others (OECD, 2015^[1]). Where the value of (secondary) data re-use for society (i.e. social and market value) is larger than the value (of primary data use) for the individual member of society (private value), access to and sharing of data can maximise the re-use and therefore the value of data across organisations, sectors and economies.⁴

Yet, despite the growing need for data and the evidence presented in this report of the economic and social benefits of data re-use and sharing, data access and sharing remains below its potential.

Barriers to data access, sharing and re-use

There are still significant barriers to data sharing and re-use, even within public organisations. The social and economic risks associated with the possible revelation of confidential information (e.g. personal data and trade secrets)⁵ are often the main rationale for individuals and organisations not sharing their data. Identifying which data to share and defining the scope and conditions for access and re-use is perceived as a major challenge, in particular for individuals and small- and medium-sized enterprises. This remains true even in cases where commercial and other private interests would not oppose data sharing and re-use (AIG, 2016^[5]).

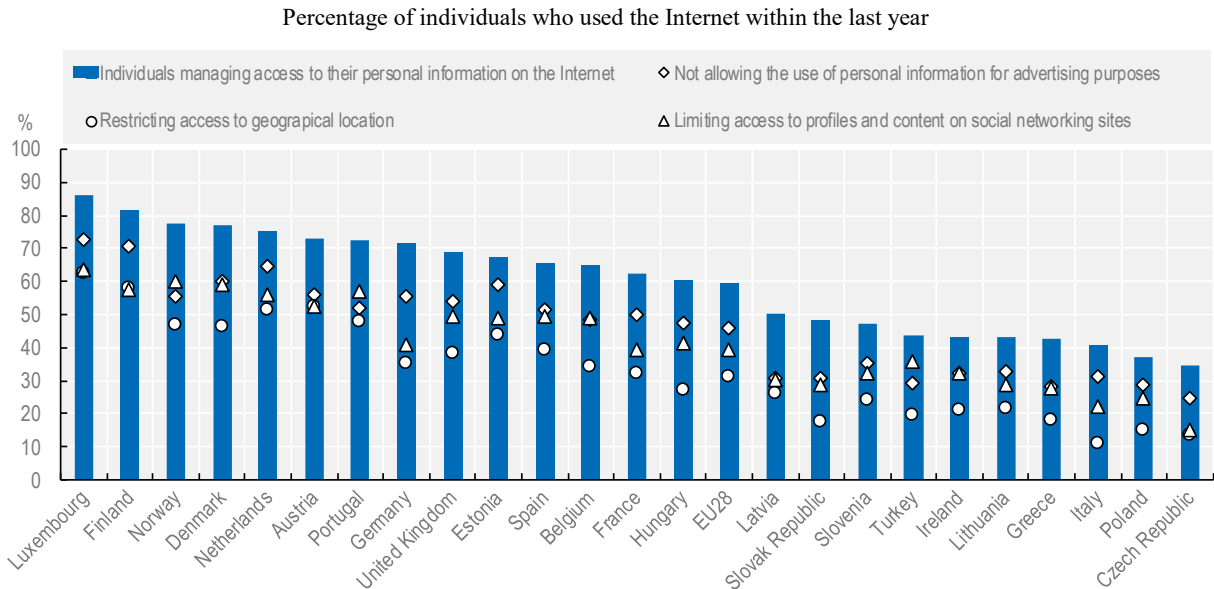
Furthermore, a survey by the Economist Intelligence Unit (2012^[6]) reports that almost 60% of companies stated that “organisational silos” are the biggest impediment to using “big data” for effective decision-making. Individuals are also increasingly wary of the re-use of their personal data. In Europe, for example, individuals i) limit the use of their personal data⁶ for advertising purposes (40% of all the surveyed population in 2016); ii) limit access to their social networking profiles (35%); and iii) restrict access to their geographic location (30%) (Figure 1.2).

As the provision of high-quality data can require significant up-front and follow-up investments, incentives to share data are often too low, in particular when individuals and organisations cannot sufficiently appropriate the returns on their investments. This is in particular true as complementary resources (e.g. additional metadata, data models and algorithms for data storage and processing and even secured information technology infrastructures) have to be made available before the data can be re-used effectively.⁷ These concerns are sometimes exacerbated by the legal complexities and uncertainties related to privacy regulation (e.g. consent), but also cross-border data access and sharing and the question of data “ownership” (see subsection “Uncertainties about ‘data ownership’” in Chapter 4).

For example, some individuals may object to the re-use of their health-related data for research purposes because of confidentiality concerns, though they may be aware of the social benefits that can be derived for themselves and for society. And though there is real market demand, some organisations may be reluctant to share or even sell or license their proprietary data because they cannot perceive their market value or because the cost of making that data available appear higher than the expected benefits. In some cases, organisations may be willing to share the data only under the condition that other organisations do the same or that there are clear benefits for them. According to an AIG-commissioned survey of 400 employees and 250 business executives across nine countries,⁸ more than two-thirds of respondents said they would “engage in the safe sharing of data if they received some

benefits from doing so” (AIG, 2016^[5]). Lack of reciprocity and concerns of “free riding” may thus require co-ordination across industries (i.e. a collective action problem) and strong leadership to establish a culture of trust for data sharing across society.

Figure 1.2. European individuals restricting the use of their personal information over the Internet, 2016



Source: OECD (2017^[7]), *OECD Digital Economy Outlook*, <https://dx.doi.org/10.1787/9789264276284-en>, based on Eurostat Digital Economy and Society Statistics, <http://ec.europa.eu/eurostat/web/digital-economy-and-society/data/comprehensive-database> (data accessed March 2017).

In addition, the lack of dedicated funding for data-sharing infrastructures and the limited pathways for their sustainment even in critical areas like science and health care research, combined with the misalignment of incentives to invest in, curate and share data have increased the risk of data erosion over time. According to Vines et al. (2014^[8]), for instance, the probability of finding the data associated with most scientific papers declines by 17% every year.

All this may lead to significant (social and economic) opportunity costs. As Rufus Pollock, Founder and President of Open Knowledge International, stated at the OECD Technology Foresight Forum in October 2012: “The best thing to do with your data will be thought of by someone else.” This is particularly the case where the spill-overs of data cannot be easily observed or quantified (e.g. socialisation and behavioural change, cultural and scientific exchange, or greater levels of trust induced by transparency). As a result, countries’ capacity to innovate may risk being undermined if less data can be used as input to innovation, in particular in the current age of AI.

Objectives and structure

This report examines the opportunities and challenges of enhancing access to and sharing of data (EASD). It discusses in particular how EASD can be an effective means for maximising the social and economic benefits of data and data re-use, while, at the same time, addressing related risks and challenges and protecting the private interests of individuals and organisations. It provides examples of some approaches to EASD that can enable the free flow of data across nations, sectors and organisations, and at the same time address the legitimate concerns

of individuals and organisations, including governments, while assuring DDI, growth and well-being across societies.

The report builds on the findings of the OECD Expert Workshop on “Enhanced Access to Data: Reconciling Risks and Benefits of Data Re-use”, held in Copenhagen (Denmark) on 2-3 October 2017 (Copenhagen Expert Workshop) and the findings of the Joint CSTP-GSF Workshop “Towards New Principles for Enhanced Access to Public Data for Science, Technology and Innovation”, held at the OECD on 13 March 2018 (CSTP-GSF Workshop) and the Open Government Workshop held in Stockholm (Sweden) on 16 March 2018 and organised by OECD Directorate for Public Governance (GOV) (Stockholm Open Government Workshop).

The rest of the report is structured as follows: Chapter 2 presents data typologies, key data access mechanisms, and the main types of actors and their roles, and then examines the different policy approaches and degrees of openness of EASD. Chapter 3 highlights the benefits of data access and sharing. Chapter 4 discusses the risks and challenges. Chapter 5 presents recent government policy initiatives that promote data access and sharing.

Implications for public policies and business strategies on data access and sharing

The policy issues identified in this report (Chapter 4) require differentiated approaches to data access and sharing. Open data, the most extreme approach to data access and sharing and the most commonly used by policy makers, remains highly relevant, in particular for public-sector and research data. But other approaches and strategies are available to policy makers and business leaders, and they are being adopted across application areas (Chapter 2).

Overall, this report shows that there is no one-size-fits-all optimal level of data “openness”. The optimal level depends on the context of data access, sharing and re-use, including the social, economic, and cultural environment in which these activities take place. At the same time, the report recognises the need for data-governance frameworks that incorporate a whole-of-government approach and are coherent across application areas, sectors and ideally countries. It highlights a number of differentiating factors that data-governance frameworks need to reflect to enhance access to and sharing of data for the benefits of all.

The findings of this report are not only relevant for public policies and business strategies related to data access and sharing. They also have implications on the legal instruments related to data access and sharing developed by the OECD. To this date, the OECD has developed seven Council Recommendations that address data access and sharing directly or indirectly. These include:

- the OECD (2006^[9]) *Recommendation of the Council concerning Access to Research Data from Public Funding*
- the OECD (2008^[10]) *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*
- the OECD (2009^[11]) *Recommendation of the Council on Human Biobanks and Genetic Research Databases*
- the OECD (2013^[12]) *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*
- the OECD (2014^[13]) *Recommendation of the Council on Digital Government Strategies*
- the OECD (2016^[14]) *Recommendation of the Council on Health Data Governance*.

Member countries are currently reviewing these legal instruments to ensure their coherence and continued relevance, in light of the findings of this report.

References

- AIG (2016), *The Data Sharing Economy: Quantifying Tradeoffs that Power New Business Models*, <http://www.aig.com/content/dam/aig/america-canada/us/documents/brochure/the-data-sharing-economy-report.pdf>. [5]
- Economist Intelligence Unit (2012), *The deciding factor: big data & decision making*, <http://www.capgemini.com/insights-and-resources/by-publication/the-deciding-factor-big-data-decision-making/>. [6]
- Grzywaczewski, A. (2017), *Training AI for Self-Driving Vehicles: the Challenge of Scale*, <https://devblogs.nvidia.com/training-self-driving-vehicles-challenge-scale/>. [2]
- Nelson, P. (2016), “Just one autonomous car will use 4,000 GB of data/day”, *NetworkWorld*, <http://www.networkworld.com/article/3147892/internet/one-autonomous-car-will-use-4000-gb-of-dataday.html>. [3]
- OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/eedfee77-en>. [15]
- OECD (2017), *OECD Digital Economy Outlook 2017*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264276284-en>. [7]
- OECD (2016), “Health Data Governance Recommendation”, in *Recommendation of the Council on Health Data Governance*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0433>. [14]
- OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264229358-en>. [1]
- OECD (2014), *Recommendation of the Council on Digital Government Strategies*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0406>. [13]
- OECD (2013), *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, amended on 11 July 2013, OECD, Paris, <https://legalinstruments.oecd.org/public/doc/114/114.en.pdf>. [12]
- OECD (2009), *Recommendation of the Council on Human Biobanks and Genetic Research Databases*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0375>. [11]
- OECD (2008), *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*, OECD, Paris, <https://legalinstruments.oecd.org/public/doc/122/122.en.pdf>. [10]
- OECD (2006), *Recommendation of the Council Concerning Access to Research Data from Public Funding*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>. [9]

- Simonite, T. (2018), *Some Startups Use Fake Data to Train AI*, [16]
<http://www.wired.com/story/some-startups-use-fake-data-to-train-ai/>.
- Vines, T. et al. (2014), “The Availability of Research Data Declines Rapidly with Article Age”, [8]
Current Biology, Vol. 24/1, pp. 94-97, <http://dx.doi.org/10.1016/j.cub.2013.11.014>.
- Waters (2015), “IBM’s latest deal is a new test case for the big data economy”, *Financial Times*, [4]
<http://www.ft.com/content/0fe3ac2e-7e22-11e5-a1fe-567b37f80b64>.

Notes

¹ The term “data” can have multiple meanings. Depending on the context and jurisdiction, for example, “data” can be used to refer to: raw or unprocessed data, whether in analogue or electronic format; personal information; information in electronic form (including reports, maps, and photographs, which are sometimes broadly referred to as “digital content”); or all recorded information. “Data” can refer to “factual records” including single, or a large collection of, items such as numerical scores, textual records, images and sounds used as primary sources for scientific research (see e.g. “research data” in OECD (2006_[9])). In other words, in some contexts, “data” is used interchangeably with “information”, and in other contexts is distinguished from “information”, where the latter is understood as “the meaning resulting from the interpretation of data” (OECD, 2015_[1]). For the purposes of this report, the term “data” covers only electronic versions of data (digital data) and distinguishes between raw data and information as defined in OECD (2015_[1]). Subcategories [which create important distinctions for policy makers] including personal data are further defined and explained in Chapter 2. These and other distinctions are provided in more detail also in endnotes 3 and 7 of this chapter, and endnotes 4, 10, 12, 24, 30 and 31 in Chapter 2.

² As a comparison, an average person is estimated to generate up to 1.5 gigabytes per day by 2020 (Nelson, 2016_[3]).

³ Although “synthetic data” (i.e. data generated via computer simulations) can be used to increase the volume of training data, many real-life problems such as driving are too complex to be simulated realistically and therefore still necessitate access to real life data (OECD, 2019_[15]; Simonite, 2018_[16]). As Intel chief executive officer Bria Krzanich explained, besides the technical (sensor) data, there will also be a need for “societal data, also called crowd-sourced data”, which include anonymised data from online platforms such as e.g. Waze, as well as personal data on individual driving patterns (Nelson, 2016_[3]).

⁴ In this report, the term “sharing” refers to the joint use of a resource such as data, be it in exchange of other resources (money or other goods or services) or for free. It thus includes the re-use of data based on both, data commercialisation, and open access to data (free of costs) and other non-commercial provisions and uses of data.

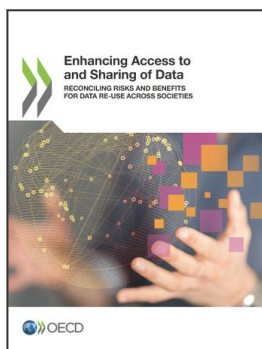
⁵ This includes “undisclosed information” according to Art. 39 of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), which provides some conditions for the information to be considered a trade secret: i) The information must be “secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question”. ii) It must have “commercial value because it is a secret”. And iii) it must have been “subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep

it secret”. Art. 39 (3) TRIPS also includes provisions for undisclosed test or other data submitted to obtain regulatory approval for the marketing of pharmaceutical or of agricultural chemical products.

⁶ Personal data is defined by the OECD (2013^[12]) *Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* as “any information relating to an identified or identifiable individual (data subject)”.

⁷ For example, data from the distributed array telescope may create large data sets, which however require additional meta-data on the direction of the telescopes to be interpreted correctly.

⁸ The countries included Australia, France, Germany, Italy, Japan, United Kingdom, United States, China and Singapore.



From:

Enhancing Access to and Sharing of Data Reconciling Risks and Benefits for Data Re-use across Societies

Access the complete publication at:

<https://doi.org/10.1787/276aaca8-en>

Please cite this chapter as:

OECD (2019), "Introduction", in *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/2e0473fb-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.