This paper presents the summary of the discussions at the OECD Expert workshop on “Enhanced Access to Data: Reconciling Risks and Benefits of Data Reuse”, hosted by the Danish Business Authority on 2-3 October 2017 in Copenhagen (Denmark). The workshop contributed to the first of two workstreams of the OECD project on Enhanced Access to Data [DSTI/CDEP/SPDE(2017)8], which aims to examine the benefits and challenges of enhanced access to data, address knowledge gaps, and help move the policy agenda further by identifying best practices on the reuse and sharing of data in the public and private sectors.

Action requested: Delegates are invited to note this document. The paper will be made available on the website of the Expert Workshop (http://www.oecd.org/internet/ieconomy/enhanced-data-access.htm).

This document is a contribution to IOR 1.3.1.2.1 of the 2017-2018 Programme of Work of the CDEP.

Christian Reimsbach-Kounatze; christian.reimsbach-kounatze@oecd.org, Tel +33145247616;
Elettra Ronchi; elettra.ronchi@oecd.org; Tel +33145241828

JT03430447
# Table of contents

**Note by the Secretariat**

Background information on the document status .......................... 4  
Action requested ........................................................................... 4

**Executive Summary** ...................................................................... 5

**Summary of the discussions at the OECD Expert Workshop on Enhanced Access to Data:**  
Reconciling Risks and Benefits of Data Re-use ................................................. 8

- Introduction ................................................................................. 8
- Approaches to enhance access to data discussed at the Expert Workshop ........................................... 9
  - Open access to data (open data) ........................................................... 9
  - (Restricted) data sharing communities .................................................. 11
  - Data markets (and platforms) ............................................................... 12
  - Data portability (and interoperability) .................................................. 13
- The costs and benefits of enhanced access to data: Evidence from the Expert Workshop ......................... 15
  - Transparency and empowerment of users ............................................. 18
  - New business opportunities including for the creation of start-ups .......... 21
  - Integration of value chains across sectors and countries ....................... 22
  - Strategic partnerships and user-driven innovation (bringing new data back) ....................................... 23
- Addressing cross-cutting issues to data access, sharing and re-use ......................................................... 24
  - Cultivating trust for data access, sharing and re-use ................................ 25
  - Engaging communities and strategic partners ......................................... 33
  - Promoting interoperability and standard mechanisms for controlled access ....................................... 36
  - Acknowledging context dependencies and the heterogeneity of data ................................................. 40
  - Fostering investments for the sustainable provision and re-use of data .............................................. 46
- The role of governments and possible priorities for further work by the OECD ........................................... 48
  - Governments leading by example by opening publicly funded data .............. 49
  - Strengthening data-related skills and competences .................................. 49
  - Promoting data policy frameworks for enhanced access to data ................. 49
  - Striking the right balance between private and public interests .................... 50
  - Focussing on strategic societal objectives .............................................. 51
  - Encouraging a whole-of-society engagement and multi-stakeholder dialogues .......... 52

**References** .................................................................................... 54

**Annex A. Agenda** ........................................................................ 56
Figures

Figure 1. The data spectrum ................................................................. 9
Figure 2. The economic benefits and savings of open data by TfL .................. 17
Figure 3. Comparing value creation of data sharing ........................................ 20
Figure 4. Selected use cases in logistics of the Industrial Data Space (IDS) .......... 22
Figure 5. Data ‘ownership’ involves overlapping rights, responsibilities and opportunities .... 28
Figure 6. Overview of the components and roles of the IDS ......................... 29
Figure 7. High definition 3D map data platform ........................................ 31
Figure 8. Certification system for data sharing platforms ................................ 31
Figure 9. How data is made open by TfL via a single API ............................ 39
Figure 10. Flowminder.org’s data processing flow ..................................... 40
Figure 11. ISO/IEC 19941 Data taxonomies Section ................................. 45
Figure 12. Data category taxonomy in ISO/IEC 19941 ............................... 45
Figure 13. Mapping data identification qualifiers with de-identification techniques .... 46
Figure 14. The problem of data erosion ................................................ 47

Boxes

Box 1. The Ostrom approach to the analyse of knowledge commons ................ 43
Note by the Secretariat

Background information on the document status

On 2-3 October 2017, the OECD held an Expert Workshop on “Enhanced Access to Data: Reconciling Risks and Benefits of Data Re-use” (Expert Workshop) which was hosted by the Danish Business Authority in Copenhagen (Denmark). The revised draft agenda [DSTI/CDEP/SPDE(2017)11/REV1] and the background report [DSTI/CDEP/SPDE(2017)13] of the Expert Workshop were made available to Delegates via O.N.E. Document in fall 2017.

The Expert Workshop contributed to kick-start the first phase of the OECD project on Enhanced Access to Data: Reconciling Risks and Benefits of Data Reuse, whose work plan is described in [COM/DSTI/CDEP/STP/GOV/PGC(2017)1] and which is co-sponsored by Denmark, Finland, Italy, Japan, Norway, and Sweden.

This paper presents the summary of the discussions at the Expert Workshop, which informed the revisions of the background report [DSTI/CDEP/SPDE(2017)13/REV1] and the draft gap analysis of principles on enhanced access to data [DSTI/CDEP/SPDE(2018)5].

The paper was circulated on 11 April 2017 for comments to the experts that participated at the Expert Workshop and to the members of the Joint Steering Group (JSG) of the project of Enhanced Access to Data.

The paper will be made available on the website of the Expert Workshop (http://www.oecd.org/internet/ieconomy/enhanced-data-access.htm).

The paper is a contribution to IOR 1.3.1.2.1 of the CDEP’s 2017-18 PWB. It may also contribute to OECD’s Going Digital Horizontal Project.

Action requested

This draft report is submitted for information and discussion. Delegates are invited to note this document, which will be presented at the 43rd meeting (15-16 May 2018) of the Working Party on Security and Privacy in the Digital Economy under agenda Item 8.
Executive Summary

1. On 2-3 October 2017, the OECD held an Expert Workshop on “Enhanced Access to Data: Reconciling Risks and Benefits of Data Re-use”, in Copenhagen (Denmark). The purpose of the Workshop was to bring together experts from the private and public sector (including policy makers), as well as from civil society and academia to learn about and discuss the benefits and challenges of four different approaches to enhanced access to data: (i) open data, (ii) (restricted) community-based data sharing, (iv) data markets, and (iv) data portability.

2. Workshop participants discussed a number of cases over the course of the workshop. Among the key benefits, experts showed that enhanced access to data enabled:

   1. *Greater transparency and empowerment of users*: Experts illustrated ways in which the different approaches to enhanced access to data improved transparency and empowered users. Open data was highlighted as critical for the validation and replication of scientific results. Discussions on data portability indicated that it can enable consumers and SMEs to more easily move data across service providers while fostering competition on prices as well as other features, including privacy.

   2. *New business opportunities (including for the creation of start-ups)*: Enhanced access to data creates business opportunities, in particular for new and established data intermediaries and mobile application (app) developers. Estimates presented at the Expert Workshop suggested that the re-use of open data of Transport for London (TfL), for instance, generated a gross value added of GBP 12-15 million per year for businesses and directly created more than 500 jobs. This is a significant share of the total estimated annual economic benefits and savings of up to GBP 130 million a year for TfL customers, road users, London and TfL itself.

   3. *Integration of value chains within and across sectors and nations*: Experts showed that enhanced access enabled the linkage and integration of data from multiple sources for the creation of “smart services”. Promising examples included logistics, transport and navigation services, where data needs to be linked across different means of transport (i.e. *multimodal transport and navigation information*) for the deployment of smart transportation and smart city services.

   4. *Strategic partnerships and user-driven innovation*: Strategic partnerships, including through public-private partnerships (PPPs), were highlighted during the Expert Workshop as critical and, in some cases even as the most important enabler, for private sector data sharing and re-use. Within these data partnerships, organisations would share and mutually enrich their data sets, thereby creating added value and in particular the opportunity to bring back new data and insights that a single organisation would not be able to generate.
3. Workshop participants also discussed some of the risks and costs associated with enhanced access to data. Experts agreed that the provision of data required upfront and continuous investments. These included in particular facilitation and maintenance costs, which in the case of TfL, for instance, were estimated to be around GBP 1 million per year. Some experts also highlighted co-ordination costs due to frictions and poor interoperability of data and data processing systems.

4. In addition, experts demonstrated that enhanced access could cause unintended detrimental effects on consumers and SMEs. The large majority of end users (consumers and businesses included) often did not directly use the (raw) data as typically made available. They rather relied on data intermediaries (e.g. data brokers, apps and personal information management systems) that would access the data to extract and present the embedded information in more user-friendly ways, sometimes enriched through additional (inferred) data. While this was highlighted as a significant opportunity for data intermediaries, concerns were raised about possible increased risks of privacy violation and anti-competitive behaviour.

5. Experts discussed the following cross-cutting issues that need to be addressed by all stakeholders to promote a data sharing culture:

- **Cultivating trust for data sharing and re-use:** Trust was stressed as the most critical enabler for data access, sharing and re-use across organisations, sectors and countries. In this context privacy protection (in the case of personal data re-use) is a key component of trust. Experts acknowledged, however, the need to strike a balance between the legitimate interests of the individuals concerned and those controlling and/or using their data. Legal complexity and uncertainties related to privacy regulation (e.g., consent), but also to the free flow of information and the question of data “ownership” were raised as requiring further discussion. Experts highlighted the importance of trusted third parties and the need for data governance frameworks that provide common definitions and principles for guiding data sharing and re-use.

- **Engaging communities and strategic partners** (including in particular PPPs): Trust was also a reason for the critical role of communities and data partnerships. The Expert Workshop revealed that communities were essential for every approach to data access. This suggested that the governance of communities (e.g. how membership is ruled, decisions made and disputes resolved) was a critical part of the governance issues to be addressed. The ability to engage communities and partners required strong leadership with a strategic focus on the value-added to be created by data re-use. Experts strongly advised against “simply throwing data into the public domain”, but to focus on the societal and business challenges to be addressed in order to effectively engage communities and data partners. Data partnerships also raised a number of challenges, in particular in the context of PPPs, where ensuring a fair exchange between (private and public sector) partners was challenging.

- **Promoting interoperability and standard mechanisms for controlled access:** Lack of standards (or the proliferation of incompatible standards) was viewed by experts as a major cause of data silos, which some highlighted as barrier to data sharing and re-use. Experts argued that standards were needed to enable interoperability, which, in turn, would help foster competition by lowering barriers to switching service providers. However, the Expert Workshop also
revealed that data silos were often reflecting organisational structures and business processes, and used as a strategy to protect privacy and intellectual property rights (IPRs). In these cases, application programming interfaces (APIs) were introduced as a possible solution to control how data would be accessed and used. In cases where the data was very sensitive, experts highlighted data sandboxes as a promising approach, where data is accessed and analysed in an isolated environment (sandbox), and results only exported, if at all, when non-sensitive.

- **Acknowledging context dependencies and the heterogeneity of data:** Experts stressed the commercial, legal and cultural environment as major contextual factors to be taken into account by data governance frameworks together with the different types of data and stakeholders. The Expert Workshop suggested that a better differentiation of these factors (e.g. by adopting common taxonomies), and the examination of their relative weights would lead to a more accurate picture of the risks and benefits of data sharing and re-use. It would also help address challenges raised during the re-use of data where the change of context could undermine the rights of data subjects and IPR holders.

- **Fostering investments for the sustainable provision and re-use of data:** Experts highlighted the following measures as essential for the sustainable provision of data: (i) providing dedicated funding for data infrastructures (in particular when data are of public interests such as in science); (ii) better aligning existing incentives structures for data curation and sharing (e.g. by taking data citation into account when rewarding scientists); and (iii) strengthening skills and competences for data analysis and sharing.

6. Workshop participants discussed the role of governments and where the OECD could provide further policy insights and expertise. Experts stressed the leadership role of governments in defining action areas and providing formal governance frameworks that address the issues highlighted above. Some experts stressed that these issues could, but did not have to, require a regulatory response. Overall, the roles of governments as discussed by experts can be summarised as follows: (i) leading by example by opening publicly funded data, (ii) strengthening skills and competences, and (iii) promoting data governance frameworks that (iv) strike the right balance between the private interests of individuals and organisations, and the public interests for data sharing and re-use, while (v) focussing on strategic societal objectives and (vi) encouraging a whole-of-society engagement through PPPs and multi-stakeholder dialogues.

7. Policy experts agreed that international co-operation was needed to promote enhanced access to data across organisations, sectors and countries, and they invited the OECD to: (i) continue to provide evidence on the costs and benefits of enhanced access to data, including through case study analysis of private and public sector initiatives; (ii) monitor policy developments in member countries and showcase innovative practices; (iii) provide clear definitions and taxonomies to assure a common language on data governance issues; and (iv) develop guidelines on common data governance principles in a multi-stakeholder process that would give all countries a tool to move forward their national policies on enhanced access to data.
Summary of the discussions at the OECD Expert Workshop on Enhanced Access to Data: Reconciling Risks and Benefits of Data Re-use

Introduction

8. On 2-3 October 2017, the OECD held an Expert Workshop on “Enhanced Access to Data: Reconciling Risks and Benefits of Data Re-use” (Expert Workshop) which was hosted in Copenhagen (Denmark) by the Danish Business Authority (see the Annex for the Agenda). The Expert Workshop brought together experts from the private and public sector (including policy makers), as well as from civil society and academia. It was attended by more than 90 participants representing 22 delegations.

9. The objective of the Expert Workshop was to examine through concrete case studies how enhanced access to data could maximise social and economic benefits, while addressing legitimate concerns of individuals and organisations (including governments). Four approaches for enhancing access to data were discussed in dedicated sessions: (i) open data, (ii) community-based data sharing agreements, (iii) data markets, and (iv) data portability. The general questions to be addressed were:

- What are the conditions under which specific approaches are more appropriate?
- What promising and sustainable business models are emerging for data re-use?
- What are common and specific data governance challenges and solutions?
- What possible role could governments and the private sector play, respectively?
- Where is international cooperation needed and what would be the role of the OECD?

10. This paper presents the main findings and common (cross-cutting) issues emerging from experts’ presentations and workshop discussions. It is structured as follows: After introducing examples for each of the aforementioned four approaches, the paper highlights the common social and economic costs and benefits. It then presents the cross-cutting issues that were identified, before concluding with a section on the potential role of governments and possible priorities areas for the OECD as discussed in particular during the closing session.

---

1 For further details, please visit the website of the Expert Workshop (http://www.oecd.org/internet/ieconomy/enhanced-data-access.htm).
Approaches to enhance access to data discussed at the Expert Workshop

“Open access works for some resources. But for most, governance is complex, contextual, resource-specific and community-specific.” (Brett Frischmann, The Charles Widger Endowed University Professor in Law, Business and Economics, Villanova University)

11. As highlighted in the (OECD, 2017[1]) background report, enhanced access to data should not be considered a “binary concept” opposing close to open access to data, rather a continuum of different degrees of data openness, ranging from limited access (only by the data controller) to open access to the public. The data spectrum presented by Jeni Tennison (CEO, The Open Data Institute, ODI) illustrates well this continuum and the different types of access to data (Figure 1). It provides in particular some examples showing to what degree different types of data can be shared.

**Figure 1. The data spectrum**

![Data Spectrum Diagram](attachment:image)

Source: presented by Jeni Tennison (CEO, The Open Data Institute).

12. The Expert Workshop had been structured around four substantive sessions aimed at discussing approaches and policy challenges along this continuum, which are further discussed below.

**Open access to data (open data)**

13. Open access is the most prominent approach used to enhance access to data in particular in the public sector and in science (OECD, 2017[1]). The meaning as well as the impact and value of open access to data are, however, still subject to controversy. The aim of the first session was to bring some clarity on these issues. Niall Brennan (President and Executive Director, Health Care Cost Institute, HCCI) in his keynote presentation provided the definitions by the ODI and the Open Data Handbook. The former defines open data as “data that anyone can access, use or share” while the latter defines it as “data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike”. Niall Brennan concluded by proposing that the
term “open data” should be understood as referring to “publicly available data, structured in a way that enables the data to be fully discoverable and usable by end users”. The following factors were highlighted as critical for open data:

- **Accessible**: Open data are made available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched. Formats should be machine-readable (i.e., data are reasonably structured to allow automated processing).
- **Described**: Open data are described fully so that consumers of the data have sufficient information to understand their strengths, weaknesses, analytical limitations, security requirements, as well as how to process them.
- **Reusable**: Open data are made available under an open license that places no restrictions on their use.
- **Complete**: Open data are published in primary forms.
- **Timely**: Open data are made available as quickly as necessary to preserve the value of the data. Frequency of release should account for key audiences and downstream needs.
- **Managed Post-Release**: A point of contact must be designated to assist with data use and to respond to complaints about adherence to these open data requirements.

14. According to Rikesh Shah (Lead Digital Partnerships Manager, Transport for London, TfL), a consensus around the definition of open data is still missing. The expert however agreed that open data shared a number of characteristics such as: (i) access and use at no charge, (ii) a level playing field enabling everyone to access the data, (iii) the easiness with which to access the data, and (iv) the support of multiple data formats.

15. Catriona MacCallum (Director of Open Science, Hindawi; Member of the Boards OASPA & OpenAire) made the comparison between open data and open access to (scientific) publications. She stressed that open data also required opening the required resources along the entire research cycle. Therefore access to open data needed to be complemented with the use of open access to publications, open data infrastructures and open standards (for discoverability) as well as open licenses, arguing that this was also crucial for text and data mining (TDM) (see section on “Addressing cross-cutting issues to data access, sharing and re-use” for more details).

16. Federica Rosetta (Director Global Strategic Networks, Europe, Elsevier) reported (on data markets) that the technical conditions for open data sharing and re-use were summarised comprehensively under the Findable, Accessible, Interoperable, and Reuse (FAIR) principles and related to Elsevier’s 10 components of “highly effective research data” (de Waard, Cousijn and Aalbersberg,(n.d.)):

1. **Stored**: Data that have been acquired need to be stored, so that it is accessible;
2. **Preserved**: Data need to be preserved for the long term;
3. **Accessible**: Enabling both researchers and machines to access the data;
4. **Discoverable**: Data need to be discoverable so it can be accessed;
5. **Citable**: A condition for tracking and incentivising data submissions and reuse;
6. **Comprehensible**: Proper metadata are needed to enable proper reuse of data;
7. **Trusted**: Review and data quality control to assure trustworthiness of data;
8. **Reproducible**: Validate key experimental results via independent replication;
9. **Re-usable**: Data re-use is the key benefit for the wider research community;
10. **Integrated**: Integrate these nine aspects for “highly effective research data”.

Unclassified
(Restricted) data sharing communities

17. The (OECD, 2017[1]) background report highlighted some cases, where data was considered confidential and therefore could not be openly shared with the public (as open data) because of economic (e.g. trade secrets), privacy, organisational or national security reasons. In these cases there could still be a strong rationale for sharing data, but only between members of a predefined community.

18. Experts presented, however, a number of cases of “data sharing communities” which, differently to what was assumed prior to the workshop, supported the notion that communities are key for every approach to data access: Communities of data users (such as application developers) were, for example, highlighted by some experts as success factor for open data projects. And in the case of data markets, access could also be restricted to (registered) members that had to be accredited to join the community of market participants [see section further below on “Engaging communities and strategic partners”].

19. Most of the communities discussed in session 2, however, differed from other communities in that access to data required a much higher level of access control and protection. These communities, which after the workshop are referred to as “restricted data sharing communities, could include data users as in the case of MasterCard’s data philanthropy programs presented by JoAnn Steiner (EVP and Global Privacy & Data Protection Officer, MasterCard), or only include data providers, who could also act as data users, such as in the case of the Industrial Data Space (IDS), a platform for B2B data sharing presented by Jakob Rehof (Director, Fraunhofer Institute for Software and Systems Engineering, ISST, Germany) on behalf of Jan Jürjens (Director Research Projects, Fraunhofer ISST).

20. JoAnn Steiner explained how MasterCard pursued its data research engagements with universities, government organisations, and non-government/non-profit organisations as part of its “data philanthropy” agenda and its commitment to expand financial inclusion around the world. She highlighted that many researchers are interested in using MasterCard data in their studies. To enable collaboration, MasterCard has established two programs for selected universities and research institutions:

- **For the Data Grant Recipients program**, universities or other research institutions were selected to conduct their research off-site. Datasets were reviewed by MasterCard’s Privacy and Data Protection counsel. Control mechanisms were required for the grant – including a procedure that guaranteed the destruction of the data at the end of the project. MasterCard’s data was then transferred via a secure file transfer mechanism so that Data Grant Recipients could work in their own secure, isolated data processing environment with controls approved by MasterCard Corporate Security. Data Grant Recipients had to check in with MasterCard on a quarterly basis and MasterCard reserved the right to review the research results before publication. At the end of the projects, destruction of data had to be verified.

- **For the Data Fellows program**, MasterCard selected institutions for longer-term research relationships. The institutions were then invited to select Data Fellows based on their project. MasterCard’s Data Philanthropy Board reviewed the projects and individuals, which were then allowed to conduct research on-site at MasterCard. Data fellows worked in a segregated, isolated data processing environment monitored by MasterCard’s corporate security team with computing
resources and software tools provided by MasterCard. Projects, datasets and any combinations of data were reviewed by the MasterCard’s privacy counsel and the data remained inside MasterCard’s information systems.

21. According to Jakob Rehof the main benefits and strengths of the IDS platform were its ability to link and integrate data from multiple sources and sectors, and of different types (e.g. product data and environment data of production). This was still a major challenge for Industry 4.0, where different stakeholders across sectors such as automobile manufacturing, electronics and IT services, logistics, machinery and plant engineering and pharma and medicine wants to exchange data. The IDS enables participants to exchange data in a two-steps process: Data generated by digital services and smart devices in the Internet of Things (IoT) are first mapped, combined and aggregated, and then anonymised and assessed according to specific data quality criteria. The data are then shared only with certified IDS members.

Data markets (and platforms)

22. The increasing interest of organisations to act as data providers and to meet existing demand for their data has accelerated the emergence of data markets, where data intermediaries also play a major role (OECD, 2017\(^1\)). Presentations by some experts appeared to suggest that the economic potential of data markets was significant and increasing. According to estimates provided by Fabrice Tocco (co-founder of Dawex, France), and based on Gartner, Forrester and McKinsey, the current potential market size for the commercialisation of data was around USD 500 billion. Fabrice Tocco described Dawex as a platform, which aims to become “a global data exchange where companies meet, buy and sell any kind of data, any industry, anywhere in the world”. The platform acts as an enabler by providing potential sellers and buyers with standard licence schemes, a payment and a data exchange infrastructure. He then presented two typical use cases of Dawex:

- In the *data buyer use case*, a hedge fund would buy different data sets ranging from environmental data, logistic data, mobile telecommunication data, and consumer transaction data, in order to better predict macroeconomic trends effecting its potential and current investments. The data would be used primarily for the hedge fund’s algorithmic trading system.

- In the *data seller use case*, a mobile telecommunication service provider would offer anonymized and aggregated customer data (including geolocation data and call logs) to potential buyers. These could include e.g. geolocation data for: public administrations for improving quality of public services; financial services and insurance companies for improving risk models; and retailers and real estate companies for identifying the geographic areas of high-potential customers.

23. In his key note presentation, Paul Hofheinz (President and Co-Founder, The Lisbon Council), on the other hand, expressed some doubts about the potential of data markets. He argued that one reason why data markets remained rare could be because markets may not be the most appropriate exchange mechanism for data. There was something fundamentally new and unique about data, which made pricing a challenge and required looking beyond current models of intellectual property rights (IPRs) and intangible assets. As an illustration, Paul Hofheinz highlighted Monsanto’s acquisition of Climate Corporation, an agriculture analytic firm, for USD 1.1 billion. He argued that this represented to a large extent the value of the data that Climate Corporation had collected
through sensors used by major farmers in the United States. Paul Hofheinz then explained that the price for which each of these farmers would be willing to sell their data (on average around 2 USD per acres) was, in total, much lower than the acquisition sum payed by Monsanto. This was because the value of data lied in the aggregation of the data, and not in the individual data points.

24. Malte Beyer-Katzenberger (Policy Officer, Unit G3 - Data Value Chain, DG CONNECT, European Commission), in contrast, argued that data were “tradable” goods like any other and data markets therefore a most promising development. He stressed that data could be “traded” without the need to have a right in rem and that marketplaces were typically the way to trade goods as they could lower transaction costs. However, he also acknowledged that not all data could be traded, referring in particular to personal data, which would need to involve the consent of the individual. Whether data could be traded also depended on the ‘data market or data marketplace’ which could be of different types and classified according to:

- Whether direct trading is possible or requires the involvement of data intermediaries;
- Whether individual “micro data” are provided (such as is the case of data marketplace) or whether derived information is provided (such as is the case of data brokers);
- The degree to which the platform is open or closed (i.e. the extent to which the data need to stay in a closed environment such as in the case of the IDS; and
- Whether individuals (data subject) are involved such as is the case with personal information management system (PIMS, or sometimes also referred as personal data store, PDS).

25. While most markets were targeting business-to-business (B2B) data sharing and re-use, Personal Information Management System (PIMS) (as cited by Malte Beyer-Katzenberger) or personal data store (PDS) (as cited by Naoto Ikegai, Interfaculty Initiative in Information Studies, University of Tokyo, Japan) were highlighted as promising platforms to empower and give more control to data subjects (consumers) over their personal data. Naoto Ikegai explained that in Japan the concept of PDS was receiving a lot of attention as a driver for data portability. The PDS functions as a centralised data infrastructure allowing individuals to manage their personal data. By assessing and confirming the reliability and trustworthiness of data users, the PDS aims to increase trust in data re-use and could thus function as “Information Trust Bank”. Naoto Ikegai presented an example of a PDS application in tourism, where an App (Omotenashi) would collect existing personal information from social network services, which would be shared with local businesses (provided user consent is given). The App would also give recommendations for places and businesses to visit based on the data the user would consent to include.

26. PIMS was further discussed during session 4 (on data portability see below), by Robin Wilton (Technical Outreach for Identity and Privacy, Internet Society) highlighting that it could act to restore user agency in the context of the IoT (Urquhart, Sailaja and Mcauley, 2017[3]).

**Data portability (and interoperability)**

27. Data portability was presented at the Expert Workshop as a promising means for promoting cross-sectoral re-use of data, while strengthening the control rights of
individuals over their personal data. In his keynote Mark MacCarthy (Communication, Culture and Technology Program, Georgetown University, United States) defined data portability as a means to providing customer data in a commonly-used, machine-readable structured format either to the customer or to a third party. He further referred to Article 20 of (European Union, 2016) General Data Protection Regulation (GDPR), which states that “the data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance.”

28. The regulation of data portability was perceived by experts as controversial. Mark MacCarthy stressed that data portability may, but did not need to, result from normal market forces. And other experts questioned the economic and social benefits of data portability. In particular, some experts expressed concerns that data portability as regulated by Article 20 of the GDPR may not actually generate the promised benefits to consumers and that it may be too burdensome to some businesses, in particular SMEs and start-ups (see next section on the “Social and economic costs and benefits of enhanced access to data: Evidence from the Expert Workshop”).

29. Ruth Boardman (Partner, International Privacy and Data Protection Group) for instance highlighted that the GDPR only required the data to be made available to the individual within no more than a month, which she criticised as being too long a period for being useful to individuals in most cases. Other experts highlighted that data portability may increase the level of security and privacy risks. As Mark MacCarthy explained, the more accessible an individual’s personal information, the greater the likelihood that information can be accessed and shared inappropriately by a third party. A lot depended on the identification, authentication and other security measures companies would put into place to respond to data portability requests, the expert argued. Ruth Boardman (Partner, International Privacy and Data Protection Group) later noted that there was no obligation to maintain, acquire or process additional information to identify the data subject for sole purpose of compliance; the data subject would have to be informed only if possible.

30. Mark MacCarthy explained that some of the expressed concerns were related to the lack of clarity about data portability and its relationship to concepts such as interoperability, compatibility and “common functions and features”. Data portability – in the case of the GDPR – gives data subjects the right to receive the data provided in a structured, commonly used and machine-readable format and to transmit those data to another controller. As some experts stressed, this however did not mean that the data would be transferred in a format that could be re-used by another system. Although interoperability was the goal of the right to data portability of the GDPR – according to the Article 29 Working Party (Art. 29 WP) – Mark MacCarthy further stressed that there was “no obligation for the controllers to adopt or maintain processing systems which are technically compatible” (see section on “standards and interoperability”). This was later

2 The right to data portability in the GDPR is composed of three different rights: i) the right to receive (without hindrance from the data controller) data concerning data subject which he/she has provided; ii) the right to transmit (without hindrance from the data controller) those data to another controller; and iii) the right to have the personal data transmitted directly from one controller to another.
confirmed by Randi Flesland (Managing Director, Norwegian Consumer Council) who explained that even though many existing digital services may enable consumers to download their personal data, these services still would rarely allow consumers to upload data from other services.

31. An additional issue raised by experts was on the type of data that falls under data portability. The GDPR attributes the right for the data subject to make his personal data portable. Hence, the importance of understanding the term “personal data” correctly (see section on “Acknowledging context dependencies and the heterogeneity of data” for more details).

32. Overall, the dominant opinion seemed to be that data portability regulation needed to be undertaken cautiously and that further studies were required, also to better assess the economic and social effects of the right to data portability of the GDPR. Naoto Ikegai stated that Japan’s Ministry of Economy, Trade and Industry (METI) and other government agencies were therefore conducting studies and experiments to better assess the effectiveness of data portability mechanisms. Japan was considering creating “new rights” to data portability, if its studies would reveal that data portability was underutilised in Japan. Other governments have adopted a similar approach, where data portability is expected to be implemented by businesses on a voluntary basis, but governments reserve the right to intervene in case of insufficient developments. Tanja Cvijanovic (First Assistant Secretary, Policy Innovation and Projects, Department of Prime Minister and Cabinet, Australia) highlighted the Australian (Productivity Commission, 2017[5]) Data Availability and Use Inquiry Report, which also proposes to introduce “a new right that enables both opportunities for active data use by consumers and fundamental reform to Australia’s competition policy e.g. banking, energy, telecommunications.”

The costs and benefits of enhanced access to data: Evidence from the Expert Workshop

“Data is a valuable national resource and a strategic asset to governments, their partners and the public” (Niall Brennan, President and Executive Director, Health Care Cost Institute, United States)

33. The generation of large volumes and a wide range of different types of data – now often referred to as “big data” – is driving, and driven by, the digitalisation of the economy and society. Malte Katzenberger (Policy Officer, Unit G3 - Data Value Chain, DG CONNECT, European Commission) highlighted, for example, the ever-growing number of goods and services that were based on data (including for their design and production). Artificial intelligence (AI) and the Internet of Things (IoT) also presented major developments that were accelerating this global trend, thereby raising the supply of, and demand for, data to a level that, with very few exceptions, no single organisation would be able to meet alone.

Naoto Ikegai (Tokyo University) highlighted as example a report (Japan’s Ministry of Economy, Trade and Industry (METI), 2017[9]), which explores the significance of data portability by 2020 through the implementation of multi-field projects (health care, retail, tourism industry). Questions addressed by the report include: How can data portability be used? What effect will it produce? And what are the technical challenges to be addressed?
34. Overall, the cases presented and discussed at the Expert Workshop underlined that the re-use of data as a (public, private or public-private) platform for innovation supported a wide range of upstream social and economic activities. This led some experts to consider data as an infrastructure as also suggested by the OECD (OECD, 2015[6]). Jeni Tennison, for example noted that as “roads help us navigate to a location, data helps us navigate to a decision”, while Brett Frischmann (The Charles Widger Endowed University Professor in Law, Business and Economics, Villanova University, United States) highlighted that many shared knowledge resources used as inputs and outputs of complex resource systems (such as data in science), depending on context, could also be considered infrastructural resources.

35. Experts shared evidence on direct and indirect economic and social benefits from data re-use across society. Recent estimates by (Deloitte, 2017[7]) presented by Rikesh Shah (Lead Digital Partnerships Manager, Transport for London, TfL, United Kingdom), for instance, showed that the re-use of TfL’s open data was generating annual economic benefits and savings of up to GBP 130 million a year for TfL customers, road users, London and TfL itself (Error! Reference source not found.). TfL’s data also contributed significantly to improving societal outcomes, facilitated innovation and contributed to improving the wider environment (e.g. air quality and lower emissions) as shown in (Deloitte, 2017[7]).

36. Overall, presentations by experts indicated that there are strong grounds to believe that data access and re-use enabled: (i) transparency and the empowerment of users including in particular consumers and small- and medium-sized enterprises (SMEs); (ii) the creation of new business opportunities including the creation of start-ups; (iii) the further integration of value chains across sectors and nations; and (iv) the support of strategic partnerships and user-driven innovation to bring back new data. Experts’ opinions and examples on each of these benefits, but also on related challenges, are further discussed below.
Figure 2. The economic benefits and savings of open data by TfL.

<table>
<thead>
<tr>
<th>TFL Passengers and Other Road Users</th>
<th>London</th>
<th>Transport for London</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saved time for network passengers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Passengers are able to plan their journeys better with TfL’s open data to provide them real-time information and advice on how to adjust their routes.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| - This provides greater certainty on when the next bus/tube will arrive and save time – estimated between 10% and 30%.
| | Gross Value Added | |
| - A number of companies use and reuse TfL data commercially, generating revenue, many of whom are based in London. |
| We estimate that the total Gross Value Added from using TfL data by these companies directly and access to the supply chain and wider economy is between £1bn and £3bn per year.
| | High-value job creation | |
| - TfL open data is intended to directly support around 500 jobs that would not have existed otherwise. |
| - Many of these jobs are in sectors associated with high productivity. |
| | Savings from not having to produce apps in-house | |
| - TfL now oversees over 3,000 registered third-party apps, currently, TfL is allowing the market to develop innovative new transport apps and services. |
| - This creates potential cost savings for TfL of not having to build apps itself or through co-developing with third-party developers. |
| | Savings from not having to invest in campaigns and systems | |
| - The publication of open data gives passengers information directly affecting the pressure on the Contact Centre. |
| - Undertaking an equivalent campaign to make available this information could cost £1m – open data allows TfL to make available the same data at a much reduced cost, increasing customer reach and improving transparency. |
| | Leverage value and savings from partnerships | |
| - Through partnerships with major data and software organisations, TfL receives significant data on travel flows that does not collect itself (e.g., crowdsourced traffic data). |
| - This allows TfL to undertake new analyses and improve its operations. |
| Better information to plan journeys; travel more easily and take more journeys | |
| - Passengers are able to switch to using free apps or free web services for real-time data when they use TfL’s open data. |
| - This creates a cost saving for those who previously subscribed to fee-based SMS alerts, estimated to work up to £2.3m pa. The use value of real-time alert services is estimated to be up to £3.8m pa. |
| Plus improved customer satisfaction from having accurate and reliable information available instantly | |
| Plus supporting the wider UK Digital Economy in London and other cities | |
| | Plus new commercial opportunities arising from open data |

Source: (Deloitte, 2017[7]) presented by Rikesh Shah (Lead Digital Partnerships Manager, TfL)
**Transparency and empowerment of users**

37. Cases presented by experts strongly suggested that access to data was a key means for transparency and for empowering users including in particular SMEs and consumers. Jeni Tennison (ODI), for example, presented the Open Banking initiative, which explores how data can be used to help people transact, save, borrow, lend and invest their money. By increasing transparency in the financial market, the initiative can empower consumers so that they become able to better compare existing offerings. This in turns contributed to a higher level of competition in the market. It is estimated that in the United Kingdom alone, consumers could save up to GBP 70 a year by switching to a bank account that would better fit their needs (Staff, 2017[8]).

38. In the area of science, Catriona MacCallum (Director of Open Science, Hindawi; Member of the Boards OASPA & OpenAire) stressed that access to data was critical for scrutinising and replicating scientific results giving as an example the existing challenges in verifying test results of drug interventions. The quality of scientific research depended on the extent to which the underlying data could be accessed by other scientists, which however was often not the case. In this context, she cited a paper by Ioannidis (Ioannidis, 2005[9]), according to which, for most study designs and settings, it was more likely for a research claim to be false than true. Catriona MacCallum explained that the poor availability of data required to scrutinise and replicate research results was one of the main causes for the significant share of false scientific results. There was a risk that trust in science could erode over time as a consequence.

39. On data portability, Mark MacCarthy highlighted the following benefits: (i) allowing consumers to retrieve data and move more easily to an alternative supplier; and (ii) putting competitive pressure on suppliers to keep prices low and to compete on features, including privacy enhancing features. Overall, experts agreed that data portability could lead to: (i) more vigorous competition among vendors, (ii) greater consumer choice, (iii) reduced switching costs, (iv) lower lock-in effects on consumers, and (v) higher user control of their own data.

**Most end users do not use the data as provided and rely on data intermediaries**

40. Many experts clearly demonstrated that the large majority of end users (consumers and businesses included) often did not use the (raw) data as provided. They rather relied on intermediaries such as data brokers, app developers and PIMS that access the data to provide the embedded information in more user-friendly ways, sometimes enhanced in terms of data quality and enriched through additional (often inferred) data. These intermediaries typically provide added-value services including advanced data analytic services. While businesses tended to use data brokers as main data intermediaries, consumers often had access to added-value information services via apps, the Expert Workshop revealed. Overall, this has led to new demand for added-value services and thus to new business opportunities for new and old intermediaries including data brokers, app developers, but also for some incumbents in ICT and non-ICT industries (e.g. telecommunication and financial services firms).

41. Øyvind Grinde (Head of section, Information Security and Data Sharing, Agency for Public Management and eGovernment, Difi, Norway), for example, showed that major real estate portals in Norway relied on data, most of which were provided by the public sector free of charge via open data (such as data on weather conditions, local stores, public transport, and maps). However, these real estate portals did not use public
sector data directly. Instead they used data provided by data brokers, who had tailored the (public sector) data to the specific needs of their customers, the expert explained. These brokers provided higher quality data in terms of completeness and timeliness. They often also enriched the data with information derived from, but not directly provided by, public sector data (such as on travel time). Øyvind Grinde concluded that data brokers were perceived as very appealing by users that needed high quality and complete data sets and preferred to focus on their core business activities instead of dealing with data collection and data management issues. Øyvind Grinde also confirmed that data brokers were increasingly using public sector open data across Europe.

42. As another example, Rikesh Shah (TfL) showed that a major part of the benefits of TfL’s open data where realised thanks to the development of apps that used TfL open data to provide real-time traffic information for more accurate navigation systems. He showed recent estimates by (Deloitte, 2017), indicating that users were able to save the equivalent of GBP 70-90 million in time savings and GBP 2-3 million compared to other means of information (such as SMS). Rikesh Shah also showed that more than 80 data feeds of TfL were now available for developers through a free unified application programming interface (API), which ensured accurate real-time data for over 13 000 registered developers and more than 600 apps [see section on “From ad-hoc downloads to application programming interfaces (APIs)].

43. John Foster (Director, 4th Platform - Data Strategy, Telefónica) also reported moderate to little value in providing data directly to consumers even in the case of data portability, although he recognised that data portability enabled users to re-use their data for the purpose of archiving, self-analytics and for switching services (subject to the availability of interoperable services). According to his observation, the data were often downloaded on an ad-hoc basis by consumers and rarely re-used. Even in the case, where customer data were re-used to switch providers (such as in the case of cloud and transaction data), there was little evidence of benefits, given that such data re-use was very infrequent (once in five years) and in addition introduced new co-ordination costs with possible negative impacts on innovation (Figure 3). Instead, the expert saw more promising value being created through smart services that would empower individuals, some of which provided by intermediaries including PIMS. Telefónica’s personalised service Aura, was presented as an example of “a pioneering way in the industry to interact with customers based on cognitive intelligence”. With Aura, each customer would have a personal data space (PDS) including data they generate when using Telefónica’s products and services (for example; location, payment history, etc). This

4 Øyvind Grinde explained that the open data sets provided by some Norwegian municipalities, for example, were often not complete (in terms of coverage and attributes which were missing) or were provided in a multitude of formats from one municipality to the other.

5 An open application programming interface (API) can provide both easy access to openly available data (such as a bank’s product offerings) and secure shared access to private data (such as a third party’s access to a user’s transaction history). These APIs would be established by banks and could be integrated with third-party technologies to carry out specific functions related to the banking data. For instance, apps could allow customers to compare banking services to choose what products best suits their needs (Staff, 2017).

PDS would allow Aura not only to offer a personalized experience for each user but also to show customers the data they generate and give them control over how it is used.

**Figure 3. Comparing value creation of data sharing**

![Diagram showing the comparison of value creation of data sharing]

*Source: Slides presented by John Foster (Director, 4th Platform - Data Strategy, Telefónica).*

**Data portability could cause unintended detrimental effects on users**

44. Later during the discussion, some participants argued that data portability provided businesses with new means to gain access to consumers’ personal data collected by competitors and other third parties, and that this data could even be used against the interest of consumers. Robin Wilton (Internet Society), for instance, stressed the risk of a massive over-share with new service providers (including comparison services), while also raising concerns about the tendency towards a re-intermediation of personal data controllers. He criticised in particular the propensity of providers to ask for data as a condition for switching providers, and raised the question: “Do we want the first step in a new IoT purchase to be ‘give us all the accumulated data from your previous use’?”. One participant gave as another example the possible request of an insurance company to a consumer to transmit its social network data as a condition for contracting.

45. The issue above was discussed again later, when the question was raised, whether business customers needed data portability rights. In this context, Mark MacCarthy explained that, in the case of business-to-business (B2B) relationships, the data in question were often non-personal and therefore the GDPR would not apply. Furthermore, and more importantly, contractual agreements would typically define the rights and obligations of all business partners, he stressed. Some participants argued however that SMEs (including start-ups) may deserve special treatments, because of their weaker bargaining power during contractual negotiations. Other participants expressed however their concerns for a business right to data portability even for SMEs and referred to alternative means for addressing the issues at stake (see for instance section on “data governance frameworks and models contracts”).

---

7 See also (Urquhart, Sailaja and Mcauley, 2017[3])
46. Some participants argued for instance that large data-intensive businesses could take advantage of the data portability rights of their users (consumers and businesses) to gain access to the data controlled by possible competitors, including SMEs and start-ups. This would, for instance, be the case if the large business would act as a PDS for consumers, participants explained. Lenard Koschwitz (Director European Affairs, Allied for Startups) confirmed that data portability could be more beneficial to conglomerates and large data-intensive businesses as it would facilitate the acquisition of new data from niche markets, the markets typically served by start-ups. For Lenard Koschwitz, introducing a business right to data portability was therefore likely to cause unintended consequences to the possible disadvantage of start-ups.

47. Other participants highlighted unfair competitive practices as a major rationale for introducing data portability rights for business customers. Naoto Ikegai (University of Tokyo), for instance, highlighted that the Japan Fair Trade Commission (JFTC) issued a report entitled “Data Economy and Competition Policy” in June 2017, which stressed that “it may fall under the antitrust violation (abuse of superior position) if the large enterprise forced business partner SMEs to provide data which is gathered independently.” The report concluded that some practices that unjustly denied access to data could be an antitrust violation classified as “unfair enclosure”. Naoto Ikegai gave as example the case where maintenance work of an elevator in an apartment would be switched from an inspection company of a manufacturer group to a new entrant. Where data on how the elevator is operated was considered indispensable for maintenance and inspection, a manufacturer company that would unfairly refuse to disclose operation data to a new inspection company, would fall under the antitrust violation (trade disturbance etc.), the expert explained. Access to competition essential data was perceived as justifiable regulatory action if undertaken on a case-by-case basis.

**New business opportunities including for the creation of start-ups**

48. The cases illustrated above show that the provision and re-use of data enabled the creation of new business opportunities. For smaller and less competitive firms, however, there may be some challenges, in particular when access to data is mandatory. The business opportunities enabled by access to data include the creation of new products (goods and services) as well as the creation of new businesses (start-ups) including data brokers and app developers.

49. According to Niall Brennan (HCCI) the release of open data in the United States enabled the creation of many new businesses. An example was RowdMap, an analytics company using open data to help health care planers, physician groups, and hospital systems identify, quantify, and reduce low-value care. In July 2017, the company was acquired for USD 70 million by Cotiviti, a provider of analytics-driven payment accuracy solutions.

50. In the case of TfL, estimates suggested that the re-use of TfL’s data by businesses generated a gross value added of GBP 12-15 million per year. Rikesh Shah (TfL) also showed estimations according to which more than 500 jobs were directly created thanks to TfL open data and more 230 indirect jobs across the supply chains and the wider London economy.
Integration of value chains across sectors and countries

51. Experts presented a number of cases showing that the re-use of data enabled the integration of value chain across sectors and even across national borders. The data provided by TfL through open data, for example, enabled the integration of transport and navigation services across different means of transport (i.e. multimodal transport and navigation information), a condition for the deployment of smart transportation services (Alissa Walker, 2016[10]). Thanks to access to TfL open data, services such as Google Maps could provide more accurate multimodal navigation information including for the first and last miles, Rikesh Shah (TfL) explained.

52. Another example of the integration of value chains across sectors through enhanced access to data is the Industrial Data Space (IDS), a platform for B2B data sharing presented by Jakob Rehof (Fraunhofer ISST). As highlighted above, the development of the IDS was motivated by the recognition that the value of data was growing when combined to deliver value-added services. The main benefits and strengths of the IDS was its ability to link and integrate data from multiple sources and of different types (e.g. product data and environment data of production) in order to enable the creation of “smart services”. Figure 4 lists a number of IDS use cases in logistics, where data needed to be shared across the supply chain. Another example includes mobility services where different types of data (such as geolocation data of the means of transport, data on traffic flows, and maintenance data) needed to be combined for innovative smart mobility services as well as for added value services such as new insurance models (e.g. pay as you drive) and just-in-time maintenance services.

Figure 4. Selected use cases in logistics of the Industrial Data Space (IDS)

<table>
<thead>
<tr>
<th>Application Partner</th>
<th>Use Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUDI AG</td>
<td>Transparency in supply network</td>
</tr>
<tr>
<td>DB Mobility Logistics AG / DB Schenker</td>
<td>Transparency in supply chain – reinforced structured / automatized exchange of information between all involved parties along the supply chain</td>
</tr>
<tr>
<td>KOMSA AG</td>
<td>From shipment to customer and consumer behaviour</td>
</tr>
<tr>
<td>REWE Systems</td>
<td>Autonomous transparency in the logistics chain</td>
</tr>
<tr>
<td>Robert Bosch GmbH</td>
<td>High performance supply chain – accumulation and exchange of relevant events along the supply chain</td>
</tr>
<tr>
<td>Robert Bosch GmbH</td>
<td>Luggage Control – support from traveling salesman</td>
</tr>
<tr>
<td>SICK AG</td>
<td>Coaster – Assistance system for workers</td>
</tr>
<tr>
<td>ThyssenKrupp AG</td>
<td>Transport logistics – Optimization of efficiency and observability of truck transport processes</td>
</tr>
<tr>
<td>ThyssenKrupp AG</td>
<td>Energy supply for flexible manufacturing plants</td>
</tr>
<tr>
<td>Wacker Chemie AG</td>
<td>Tracing of consignment of goods and alerting in case of deviation</td>
</tr>
</tbody>
</table>

Source: presented by Jakob Rehof (Director, Fraunhofer-Institute for Software and Systems Engineering, ISST, Dortmund, Germany) on behalf of Jan Jürjens (Director Research Projects, ISST).
Strategic partnerships and user-driven innovation (bringing new data back)

53. The provision of data is not only of benefit to users. For data controllers, data sharing can also provide significant economic benefits. It can for instance enable new strategic partnerships, where organisations agree to share and mutually enrich their data sets, or where a community emerges that create additional value that a single organisation would not be able to create. For TfL, for example, the provision of data enabled new strategic partnerships with major data, software and Internet services providers such as Google, Waze, Twitter, and Apple, as showed by Rikesh Shah (TfL). In some cases, this enabled TfL to gain access to new data sources and crowdsource new traffic data (“bringing new data back”), to undertake new analysis and thus to improve TfL’s business operation. In doing so, TfL could gain access to updated navigation information (on road works and traffic incidents) and could enhance the efficiency of its planning and operation, the expert explained. How to ensure a fair exchange of data between the (private and public sector) actors was according to Rikesh Shah an important question to be addressed in these partnerships.

54. Rikesh Shah also highlighted that TfL received valuable feedback on its data and its API thanks to its community engagement. He stressed that user-driven innovations, knowledge sharing as well as the positive reputation were additional major benefits of open data for TfL. The benefits of user-driven innovation were confirmed by participants during the discussion. It was however stressed that user-driven innovation would only be possible where data users had the possibility and the incentives to provide revisions to the data. This was the case, for example, with data portability where data subjects had the interest and possibilities to improve the quality (accuracy) of the data held about them.

55. John Foster (Telefónica) similarly saw new promising value from “data partnerships”, through which existing data is enriched with third party data about common customers. This could enable the creation of new added value services to customers, while enabling also better decisions for businesses and society (see Figure 3). Contrary to data portability rights, which the expert suggested to be less reliable from a business perspective given that access to data would depend on the initiative of each individual customer, data partnerships would enable the re-use of personal data at larger scale and in a more secure and privacy friendly way, the expert argued. John Foster explained that Telefónica was collaborating with organisations such as Facebook, Microsoft, and UNICEF for its personalised service Aura. Thanks to this collaboration, customers would be able to talk to Aura through Telefónica’s own channels and some third party platforms like Facebook Messenger, and in the future through the Google Assistant and Microsoft Cortana. Later during the discussion, John Foster stressed that the selection of trusted partners was critical for data partnerships.

56. The benefits presented above were confirmed by Bob Bailey (Chief Information Architect, Corporate Technology, Thomson Reuters, United States) who underlined that building and engaging a community via open data was in the genuine interest of a

---

8 Telefónica, Deutsche Telekom, KPN and Orange have jointly been leading the Data Portability Cooperation, an initiative that is analysing the implementation of portability of data in a way that adds value to customers while protecting their privacy. This cooperation will be extended during 2018 to new players to lay the foundation for defining a new global data portability ecosystem.
business as with explosion of data, sources and uses, no single organisation could expect to unilaterally meet all customer information needs. He then presented three main reasons why a business would want to engage in data sharing partnerships (via open data):

1. **Encouraging a community response to scale – inclusion, specialisation**: This would include (i) helping others (customers) to add value to business solutions – lower communal barriers to participate, cooperate, share and reuse, (in the case of PermID, helping make data sharing and re-use easy and valuable for others) and (ii) giving others a reason to include the business in their solutions.

2. **Identifying where to collaborate, new ways to compete, how to foster an ecosystem**: This includes (i) technical and commercial experimentation (inside and outside of the business), (ii) learning by converting capabilities into products (e.g. Thomson Reuters Labs), and (iii) using customers as signposts.

3. **Building partnerships and collective action, powering user-driven innovation while maximising the option value of investments**: This includes in particular the development of open standards and, in the case of PermID, understanding the key barriers to data sharing and re-use and in particular the role of identity.

### Addressing cross-cutting issues to data access, sharing and re-use

“[We need to] fuel a data driven entrepreneurial culture inside and outside government.” (Niall Brennan, President and Executive Director, Health Care Cost Institute)

“[A] culture of risk aversion is leading to overly cautious interpretation of the legislations, and approval process complexity.” (Tanja Cvijanovic, First Assistant Secretary, Policy Innovation and Projects, Department of Prime Minister and Cabinet, Australia)

57. Experts agreed that a cultural change towards an entrepreneurial data sharing culture was needed inside and across organisations (including governments) in order to foster data sharing and re-use across society. The different cases presented at the Expert Workshop suggested however that promoting such a culture was complex. As highlighted further below, this is to a significant extent due to the different contextual factors affecting data access and re-use (including, but not limited to, the types of resources and stakeholders involved). Presentations and discussions at the Expert Workshop however also showed that there were a number of common issues related to data access, sharing and re-use that needed to be addressed.

58. The following sections present the common issues that experts discussed at the Expert Workshop. These include: (i) cultivating trust for data sharing and re-use (i.e. privacy risks, legal complexity and uncertainties including about data ownership, and the role of data intermediaries and data governance frameworks), (ii) engaging communities and strategic partners (including in particular through PPPs), (iii) promoting interoperability and mechanisms for controlled access to overcome data silos (i.e. standards and interoperability, APIs and data sandboxes), (iv) acknowledging context dependencies and the heterogeneity of data (i.e. data categories), and (iv) fostering investments for the sustainable provision and re-use of data (i.e. funding for data infrastructure of public interest, aligning incentive structures for the sustainable provision of data, and strengthening skills and competences for data curation, sharing and analysis).
Cultivating trust for data access, sharing and re-use

“Without trust, data value may vanish” (John Forster, Director 4th Platform – Data Strategy, Telefónica)

59. According to the interventions made at the expert workshop the most critical enabler for promoting data access, sharing and re-use across organisations, sectors and countries was trust; trust among the stakeholders that each behave cooperatively as well as trust in the quality of the data and the security of the data and infrastructure enabling data sharing and re-use. The key role of trust may explain why trusted third parties, and why building and supporting a community around data sharing and re-use, are considered major success factor for data sharing and re-use [see section below on “engaging communities and strategic partners”].

60. Tanja Cvijanovic (Department of Prime Minister and Cabinet, Australia) stressed that trust in Government’s data and digital initiatives was essential. The Government of Australia had therefore decided to demonstrate: (i) how and why government used the data of its citizens, and (ii) that the data were handled according to citizens’ rights (transparency) and the safety and security of citizens’ data were guaranteed. It was also crucial to demonstrate the benefits of data and their re-use in order to build citizens’ trust. This point was reiterated by others experts during the Expert Workshop.

61. Fabrice Tocco (Dawex) emphasised that unlocking data re-use through data intermediaries such as Dawex required trust, transparency, complete control, encryption, privacy, and compliance. Mayer-Katzenberger (DG CONNECT, European Commission) later highlighted the role of a number of technological enablers such as self-executing (smart) contracts, the need for traceability mechanisms for logging data access, usage and transaction (such as block chain and watermarks) and encryption, aggregation and anonymization techniques. These technological enablers were key for personal data where “consent management” and the “rationalisation of privacy notices” were needed, the expert explained.

Privacy and data protection

62. Experts agreed that trust was particularly challenging when it came to individuals and personal data. As Robin Wilton (Internet Society) stressed, there was often a risk of disparity and information asymmetry between an individual (data subject) and the data controller in terms of the bargain that was offered to individuals for their personal data. The expert highlighted so-called “free” services, where the individual would pay with its personal data. Beyond that initial transaction between the individual and the data controller, multiple third parties (besides the data controller) such as data aggregators, advertising placement partners, third party product and service vendors, profiling services etc. were involved in the monetisation of the individual’s personal data. All of those further instances of personal data re-use would generate economic surplus, from which however the data subject would not benefit to the same extent.

63. One of the most critical challenge related to personal data re-use discussed at the Expert Workshop was related to the question of how to strike the right balance between the legitimate interests of the individuals concerned (data subjects) and those controlling
The following main points emerged from discussions:

- **Permission-based access mechanisms**: Niall Brennan (HCCI), for example, stressed that sharing personal data required permission-based access mechanisms, but he also admitted that in some cases the public interest could outweigh the private interest of privacy protection, such as when open data on health care related activities would reveal the identity of physicians.

- **Increasing difficulty and costs of anonymisation**: Privacy protection was also challenged by the power of re-identification, many experts stressed. As Robin Wilton (Internet Society) explained, references to “data” gave a sense of neutrality which was less and less appropriate; the more supposedly anonymous data would become identifiable. The world of “big data”, data mining, inference, and machine learning was one in which data easily became information, the expert explained. A permanent one-way transformation in the opposite direction (anonymization) was therefore increasingly hard to achieve, making personal data sharing and re-use more risky (for more see section on “personal data and data identification, processing and use categories”). The expert also stressed that the privacy challenge was exacerbated due to the fact that the privacy of individuals no providing data could easily be affected by those that provided their personal data.

- **The role of consent**: Some participants also raised the questions of the role of consent and whether new consent models would be required. In his presentation on the collaboration between the tax authority and the financial sector to implement automatic delivery of tax relevant income information to financial institution, Øyvind Grinde (Difi, Norway) stressed that the exchange of personal data was based on users’ time-restricted consent. Katarina de Brisis (Deputy Director General, Ministry of Local Government and Modernisation, Norway) later in her closing remarks stressed that consent models would remain an important element of the legal ground for data sharing and re-use.

**Legal complexity and uncertainties**

Some experts highlighted the complexity of the legal environment as a barrier preventing data sharing and re-use at the national as well as international level. Laws and regulations were often inconsistent and complex and therefore a limiting factor in data sharing and data interoperability. JoAnn Steiner (MasterCard), for instance, stated that the use of private sector data required careful consideration to ensure compliance with various laws and regulations in addition to business, reputational and ethical concerns which needed to be taken into account for each and every project. Paul Hofheinz (The Lisbon Council) expressed similar concerns and argued for the establishment of “safe

---

9. During the Closing Session of the Expert Workshop, the initial findings of the OECD report on “Open access to data in science, technology and innovation” [DSTI/STP(2017)25] were presented. The report highlights a number of common issues, which so far received insufficient policy attention, among which “balancing between the benefits of sharing against privacy and confidentiality concerns” was highlighted one of the most critical issues.
havens”\textsuperscript{10} for data sharing. Citing from the Australian (Productivity Commission, 2017\textsuperscript{(5)}) Data Availability and Use Inquiry Report, Tanja Cvijanovic noted that: “there was no overarching legislation at the Commonwealth level or for most States and Territories that addressed, in a whole-of-government way, how data is made available and used”. “Data sharing involved multiple approval processes by agencies that adhered to different legislation and adopted different policies on data access” with detrimental effects on sharing.

65. Naoto Ikegai (University of Tokyo, Japan) observed that legal barriers may also act as trade and competition barriers: Although the Cross Border Privacy Rules (CBPR) of the Asia-Pacific Economic Cooperation (APEC)\textsuperscript{11} were created to solve this problem, their dissemination was still in the process and slow however. He noted that data localization laws (including for data other than personal data) were expanding, especially in developing economies, forcing companies to install servers within national borders for providing domestic ICT services. As a result, these laws were likely to illegitimately hinder the free flow of information, the expert explained. Naoto Ikegai then stressed the G7 ICT Ministerial Meeting in Takamatsu, Japan, April 29-30, 2016, where ministers agreed “except for cases with legitimate public policy objectives, … to oppose data localization requirements (domestic installation such as server equipment for providing ICT services) that are likely to hinder the free flow of information.”\textsuperscript{12}

66. Tanja Cvijanovic (Department of Prime Minister and Cabinet, Australia) highlighted that a culture of risk aversion was often leading to an overly cautious interpretation of the legislations, which in turn increased approval process complexity for data sharing and re-use. Experts acknowledged that what may be considered sometimes legal barriers may be necessary safeguards to protect sensitive and confidential (personal and non-personal) data and that digital security risks were perceived by many as a real threat to data sharing and the free flow of data in particular.

\textit{Uncertainties about “ownership” and intellectual property rights}

67. Another set of legal uncertainties arises from data ownership and intellectual property rights (IPRs). Catriona MacCallum (Hindawi; OASPA & OpenAire) stressed that most legal uncertainties were not related to safeguards for sensitive data, such as patient confidentiality and endangered, but related to IPRs and the question of data ownership. Experts agreed that the concept of “data ownership” remained a fuzzy concept and therefore challenging to understand, even for some experts. Multiple factors contribute to this challenge:

\textsuperscript{10} In the area of health research data, ‘safe haven’ refer to both, a defined physical location and an administrative set of policies and procedures relating to the secure handling of confidential patient information (Burton et al., 2015\textsuperscript{[21]}; Lea et al., 2016\textsuperscript{[22]}).

\textsuperscript{11} As of 2017, Japan, Singapore and Korea were member countries from Asia, Naoto Ikegai (Interfaculty Initiative in Information Studies, University of Tokyo, Japan) explained.

\textsuperscript{12} See \url{http://www.soumu.go.jp/joho_kokusai/g7ict/english/about.html}. 

Unclassified
1. Current legal IPR regimes such as copyright and trade secrets rights are only applicable to a limited extent, and do not protect single data points or non-substantial parts of databases.\\footnote{Databases are protected by copyright under certain circumstances, but in some countries – namely in the European Union, Japan and Korea – they are also protected by a so-called \textit{sui generis} database right (SGDR) aimed at protecting the investment.}

2. The data ownership concept is often used to describe possible control rights over data, and thus equates with data portability, which many perceived as giving in particular individuals a sense of “ownership over their personal data. As Randi Flesland (Managing Director, Norwegian Consumer Council) stressed: “If you can’t move it, you don’t own it”.

3. Data ownership was often diffuse, involving a cascading array of overlapping rights, responsibilities and opportunities as noted by Paul Hofheinz. Multiple stakeholders may be involved in the contribution, collection and control of personal data in addition to the data subject. Hence the same data may have multiple “owners”. In such case, according to Paul Hofheinz, it is more a case of “co-ownership”, as neither the “data producer” nor the “data gatherer” can claim an exclusive right over the data. Paul Hofheinz concluded that the concept of “commons”\footnote{According to (Hess and Ostrom, 2007\textsuperscript{[18]}), commons are defined as “resource management / governance institutions that enable sustainable shared use of certain resources within a community”\textsuperscript{[14]}.} may be more appropriated in some cases, although to be distinguished from “co-ownership” (Figure 5).\footnote{Co-ownership was essentially above a limited number of stakeholders having limited responsibilities and rights vis-à-vis each other, the expert explained. To help understand the implications of the proposed new concept of data ownership, Paul Hofheinz advised to compare data “ownership” to the rights parents have over their children (Hofheinz and Ratas, 2017\textsuperscript{[11]}).}

\textbf{Figure 5. Data ‘ownership’ involves overlapping rights, responsibilities and opportunities}

\begin{center}
\includegraphics[width=\textwidth]{data_ownership.png}
\end{center}

\textit{Source: Slide presented by Paul Hofheinz, President and Co-Founder, The Lisbon Council.}
68. Some experts such as Naoto Ikegia (University of Tokyo) confirmed that the clarification of data ownership was needed to promote the functioning of data market. Although this would not necessary mean that there would be a need for a right in rem for data, as noted later by Malte Beyer-Katzenberger (Policy Officer, Unit G3 - Data Value Chain, DG CONNECT, European Commission), many experts highlighted that standard data governance frameworks and model contracts were needed to facilitate data sharing and re-use (see section below on data governance frameworks).

*The role of data intermediaries and trusted third parties*

69. The existence of trusted third parties was often stressed by experts as a major enabler of trust and a facilitator of data sharing, re-use and analysis. The expert workshop presented several models of trusted third parties or platforms.

- **Data intermediary acting as a third party certification authority**: Some cases showed that data intermediaries would act as a “certification authority” as in the case of the industrial data space (IDS). The “certification authorities” of the IDS certifies all participants against standards defined by the IDS (regarding e.g. security, privacy and terms of use). “Data owners” (data providers) would define terms of use and the fees of data use, which “data brokers”, would use to match “data owners” and “data users” (Figure 6).

![figure 6](source: Slides provided as backup by Jakob Rehof (Director, Fraunhofer-Institute for Software and Systems Engineering, ISST, Dortmund, Germany) on behalf of Jan Jürjens (Director Research Projects, ISST))
• **Private sector designation of a trusted data sharing platform:** In many, if not most, cases presented at the Expert Workshop, major data providers would come together and either designate an existing trusted organisation or create a new trusted organisation and platform. As an illustration, Naoto Ikegai (Tokyo University) presented the *Ship Big Data Platform*, which was established by the Japanese Sea Association and combines and provides different datasets from weather data to ship data including information of ship ownership, operators, and trajectory. The *High Definition 3D Map Data Platform*, founded by 15 Japanese companies provides another example of a platform for co-operative research and development (R&D) in the area of AI enabled driving technologies (Figure 7). Some experts presented alternative models, whereby an existing trusted organisation would engage with potential data providers, who would agree to collaborate by sharing their data. An example is the Health Care Cost Institute (HCCI), a non-profit organisation, which provides information about health care utilisation and costs in the United States based on data are contributed by health care and health insurance companies in the United States (e.g. Aetna, Humana, Kaiser Permanente, and UnitedHealthcare). The data provided to HCCI are shared with selected researcher institutions after the information about the data providers (the providing health care and health insurance companies) have been removed.

• **Public sector designation of a trusted data sharing platform:** Some experts also presented cases where the government would act as, or create, a trusted third party. In Japan, the government has established a certificate system to support the efforts of companies that wanted to share their data (Figure 8). In addition, the government also established a data request system, which is a system that allows data sharing companies to request data that have been provided to relevant ministries and agencies. The government would provide support, in particular through tax incentives and administrative guidance, but the revocation of the accreditation was also possible in some cases, Naoto Ikegai explained. Similarly in Australia, where the government initiated the *Data Integration Partnership for Australia (DIPA)*, “an investment to maximise the use and value of the Government’s data assets.” While agencies in social services, health, education, finance, and other government agencies would provide data for linking and integration, *Accredited Integrating Authorities* would enable the integration of longitudinal data assets – “housed in a secure environment, using privacy preserving linking methods and best practice statistics to link social policy and business data”.

---


17 Tanja Cvijanovic also presented the recommendation of the Australian (Productivity Commission, 2017) Data Availability and Use Inquiry Report to the (i) “government [to establish] a new Office of the National Data Custodian”, (ii) “(predominantly existing) public bodies [to] be accredited as sector-based, national data release authorities” and that (iii) “a small number of nationally beneficial data sets be designated as National Interest Datasets”. The latter was related to the selection of strategic datasets which some experts discussed as presented further below (see section on “engaging communities and strategic partners”).

Unclassified
Figure 7. High definition 3D map data platform

Source: Dynamic Map Platform Co., Ltd (2017), slide presented by Naoto Ikegai (Interfaculty Initiative in Information Studies, University of Tokyo).

Figure 8. Certification system for data sharing platforms

Source: slide presented by Naoto Ikegai (Interfaculty Initiative in Information Studies, University of Tokyo).

Data governance frameworks

70. Data governance frameworks can act as “complementary ‘legal’ enablers” (Malte Mayer-Katzenberger, EC) for data sharing and re-use by reducing legal uncertainties on major issues ranging from privacy (e.g. the role consent) to ownership and accountability. Some of the frameworks presented at the Expert Workshop have been established by data
intermediaries, and trusted third parties, others were led by government. Examples brought to the workshop included:

- **The draft data governance framework of the Swiss personalized health network (SPHN)** presented by Effy Vayena (Professor of Bioethics at the Swiss Institute of Technology, ETH Zurich): After assessing 49 documents with guidance on sharing health related data issued between 1996 and 2017 from 32 organisations, the Department of Health Sciences and Technology at ETH Zürich identified a set of the most frequently recalled principles relevant for SPHN (Respect for persons; Privacy; Data Fairness and Accountability).

- **The Public Library of Science (PLOS) ’ Data Policy** presented by Catriona MacCallum: The policy requires all authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exceptions. (Data Availability Statement). “Refusal to share data and related metadata and methods in accordance with this policy was grounds for rejection”\(^\text{18}\). PLOS Data Policy provides guidance on acceptable data-sharing methods (such as a list of acceptable public data repositories) and unacceptable data access restrictions (e.g. conclusions depend solely on the analysis of proprietary data). Catriona MacCallum highlighted that more than 65,000 papers were now published with a data statement at PLOS with around 20% based on data repositories. She noted that requiring a Data Availability Statement had increased the number of papers with the accessible underlying data in the area of genetics from 12% to nearly 70%.

- **The 2017 Contract Guidelines on Data Utilization Rights**\(^\text{19}\) by METI and the IoT Acceleration Consortium presented by Naoto Ikegia. The guidelines provide: (i) model definitions and terms for data use in contracts between business partners for cases that multiple companies jointly carry out business using IoT and big data; and (ii) model clauses for setting data utilization rights (ownership) according to the level of business’s contribution to data creation.

- **The draft “model contract terms” and “default contractual rules” for a possible future EU framework for data access** presented by Malte Mayer-Katzenberger. The Commission Staff Working Document on the Free Flow of Data and Emerging Issues of the European Data Economy (European Commission, 2017\(^\text{[11]}\)) describes model contract terms as a “set of recommended contract terms” that could be used by market participants on a voluntary basis.\(^\text{20}\) Default contract rules, in contrast, would be used for data licences and “could be laid down in legislation which could constitute a balanced solution for business-

---

18 See [http://journals.plos.org/plosone/s/data-availability](http://journals.plos.org/plosone/s/data-availability)


20 (European Commission, 2017\(^\text{[12]}\)) states that: “The voluntary integration of such contract terms could create more balanced terms for small businesses and reduce transaction costs for instance by making available model provisions for data usage licences that cover the most common business needs, while still safeguarding complete contractual freedom. Accordingly, weaker parties could have more and fair opportunities to exploit data.”
business contractual terms in case parties have not foreseen contractual clauses on the specific points.”

- **Efforts towards a new modernized data governance framework in Australia (the Commonwealth Data Sharing and Release Act)** presented by Tanja Cvijanovic, which would include:
  - A risk-based approach to improved data sharing and release.
  - New institutional arrangements to promote trust and confidence in the data sharing and release system are centred on improved capabilities, accreditation procedures for both users and custodians, and most particularly the National Data Custodian.
  - National Interest Datasets would overcome impediments to the effective integration, sharing and use of data of national significance, currently hindered by multiple legislative barriers.
  - The Comprehensive Right for consumers to access their data from government and private data holders alike, for the purposes of improving the services that are offered to them by alternative providers.\(^{21}\)

### Engaging communities and strategic partners

“[The] explosion of data, sources and uses means [Thomson Reuters] cannot expect to unilaterally meet all customer information needs – no-one can do everything.” (Bob Bailey, Chief Information Architect, Corporate Technology, Thomson Reuters)

71. As highlighted above, an important finding was the recognition that data sharing communities always mattered. Two types of communities were highlighted as critical for the success of data sharing projects and are further discussed below: the community of data users and the community of data providers (sometimes referred to in the Expert Workshop as “data partners”, “data holders” or “data owners. Strategic partnerships between the private and the public sector (public private partnerships, PPPs) turned out to be also highly relevant.

#### Data user communities

72. Building data user communities and data partnerships was in the genuine interest of data controllers willing to share their data, as presented above in the section on “Social and economic benefits of enhanced access to data: Evidence from the Expert Workshop”.

73. Many experts highlighted that it was critical to understand and communicate how data was creating value to data users to better engage with the community of users. Rikesh Shah (TfL) stressed, for example, that data sharing projects would have to be seen not as technology projects but rather as customer projects, and that it was important to engage the community as early as during the design stage of (open data) projects.

74. Rikesh Shah also highlighted that releasing data were a necessary, but not sufficient, means to engage the community of users (developers). He explained that TfL had published its data through web pages (as plain text) in the past. These pages were scrapped by users to extract TfL’s data, but many users got the data wrong and blamed

TfL for the resulting errors. By releasing the data through applications programming interfaces (APIs) and by engaging with users for instance through hackathons, TfL was able to assure the appropriate re-use of its data. Complementary measures were thus needed, such as the engagement of the community through social media and events including hackathons, Rikesh Shah highlighted.

75. Niall Brennan (HCCI) also presented a number of initiatives by the United States Government to better engage with users of government data. Examples included: (i) Challenge.gov (“a listing of challenge and prize competitions, all of which are run by more than 100 agencies across federal government” of the United States’22), (ii) the combat feeding hackathon of the United States Department of Defense (DoD), under which combat ration nutritional data sets were released to allow participants to “look for innovative ways that soldiers could interact with this data in a mobile environment”23, and (iii) the Health Datapalooza Conference, an annual event focussing on how to make health data open, accessible, and useful.

Data provider communities (data partnerships)

76. According to experts, not only the community of users deserved special attention but also the community of data providers (data partners), which was considered critical to data sharing. While generally distinct, in some cases the two communities overlapped (see for example the case of the IDS and Dawex). As highlighted above in the section on the “Social and economic benefits of enhanced access to data: Evidence from the Expert Workshop”, data sharing enabled the creation of new strategic partnerships, enabling organisations to share and mutually enrich their data. TfL, for example, reported having entered into new data-sharing partnerships with major data, software and Internet services providers such as Google, Waze, Twitter, and Apple. Another example was Telefónica’s Aura which involved data partnerships with organisations such as Facebook, Microsoft, and UNICEF.

Public private partnerships (data PPPs)

77. The workshop provided insights about effective ways for governments to conduct Public private partnerships for data sharing and reuse (data PPPs). There are many opportunities and different approaches in conducting such data PPPs which, if effectively implemented can provide benefits for government, business as well as the general public.

78. Øyvind Grinde (Difi), for example, presented the case whereby the Norwegian tax authority had agreed with the financial sector to implement automatic exchange of information in the context of loan applications. Instead of having to repeatedly ask users for information they have already provided to public administration, such information would only need to be provided once and would be re-useable based on users’ consent during e.g. the loan applications. Katarina de Brisis reported that the Brønnøysund Register Centre in Norway had already started to look into restricted data access of personal data by the financial and insurance sector. The data was stored by Altinn, a digital infrastructure that links data from public agencies, municipalities and registers of more than 4 million inhabitants and 1 million enterprises in Norway. The benefit for the

---

22 See https://www.challenge.gov/about/
23 See https://combatfeedinghack.devpost.com/
financial institutions and their customers was a more efficient credit risk assessment process. The benefit for the tax authority was, however, less clear and the question was raised, whether and under what circumstances it would be legitimate for a tax authority to ask for access to an individual’s financial data (collected by the financial institutions).

79. In such a situation, privacy and ownership considerations may limit the potential of PPPs. Robin Wilton (Internet Society) further shared the case of “Parking in a Smart City”, where public and private sector actors were involved in the collection and use of personal data related to a municipal asset, which would also raise particular privacy issues (see section on “privacy and data protection”). Catriona MacCallum further noted that IPRs could be an issue as well. As an example the expert highlighted pre-competitive drug discovery, where IPRs would limit the potential of PPPs by making it harder to sustain knowledge sharing even in the pre-competitive phase. Other participants, stressed that the linkage of private and public sector data could raise issues in cases where data sets are provided via open data licenses such as the Open Database License (ODC-ODbL).24

80. A number of questions were raised during the discussion, in particular: What can governments do to facilitate such sharing, and where does the role as authority stop and the role as a service provider begin? What types of rules should apply for this type of data sharing, and should private sector pay a fee for the data?

The need for strong and trusted leadership...

81. There was consensus that the development of an entrepreneurial data sharing culture would require strong leadership to be successful. Niall Brennan (Health Care Cost Institute), for instance, highlighted that the fast development of open data is the United States government was driven thanks to a “young progressive geeky data literate president who surrounded himself with geeky data literate people”. Rikesh Shah (Lead Digital Partnerships Manager, Transport for London) confirmed that having a champion at highest management level was critical for the success of TfL’s open data initiatives. He also advised to set up “champions” across the organisation.

82. Brett Frischmann (The Charles Widger Endowed University Professor in Law, Business and Economics, Villanova University) also confirmed the importance of leadership. He explained that case studies on the governance of knowledge commons revealed that informal governance institutions, and especially trusted leadership, often played a key role in sustaining commons. In this context, the expert highlighted the leadership role of governments in defining action arenas and providing formal governance frameworks. He also explained that “commons seemed to play an important role in early stages of some industries (evolution)”, which suggests that commons would be relevant for industrial policies where the role of governments would be strategic.

... with a strategic focus on the value-added created by data re-use

83. Given the critical role of the community of data users and providers (highlighted above), understanding the needs of the community members was stressed by many

24 ODC-ODbL Share-Alike condition states that, “if you publicly use any adapted version of this database, or works produced from an adapted database, you must also offer that adapted database under the ODbL.”
experts as crucial. Bob Bailey (Thomson Reuters), for example, stressed that it was critical to understand how to maximise the usefulness of the data to the community. He advised against “simply throwing internal data into the public domain” but to focus on data that are useful to the community. In addition, he advised to focus on costly challenges that were common to all major stakeholders. Thomson Reuters decided to focus on identity, the focus of open PermID, the explained.

84. This idea was reiterated by other experts. John Foster (Telefónica) stressed that even in the case of data portability it was critical to understand and communicate how value was created when making customer data available. He explained that Telefónica had made available data from its internal processes (data for operations, data use across product silos and customer insights) in accordance with regulatory requirements of the GDPR. However, the expert saw little value in providing raw data directly to customers for the reasons elaborated above (see section on data portability and Figure 3).

85. In a similar note, Erik Wetter (Professor, Stockholm School of Economics; Chairman, Flowminder.org) criticised the tendency to always ask for more data. He explained that what mattered was not the data, but the insights that could be derived from the data. Very often stakeholders requesting data (including government agencies) had not the sufficient capacity to analyse the data, even if the data would be made available, he argued. In this context, he expressed his scepticism about asking for, and using, data without having a clear question and purpose behind the request and use.

86. Some governments had also recognised the need for initiatives to focus on high-value data assets according to some experts. Tanja Cvijanovic (Department of Prime Minister and Cabinet, Australia), for instance, explained that the Australian Government had designated a small number of nationally beneficial “National Interest Datasets”, including over 28 000 data records published on data.gov.au, including high-value datasets, such as the Geocoded National Address File (G-NAF). 25

87. Naoto Ikegai (Tokyo University) also highlighted that the Japanese government was considering measures to promote further data sharing, especially in areas that strongly contributed to the public interest. In this context he highlighted that it has been commonly recognized in Japan that there was a need to expand “cooperative (data sharing) areas”. The expert was expecting that such efforts would expand internationally and would lead to efficient global data markets.

**Promoting interoperability and standard mechanisms for controlled access**

“Seven out of ten Danish companies have experienced barriers that have prevented them from using data.” (Søren Gaard, Deputy Permanent Secretary, Ministry of Industry, Business and Financial Affairs, Denmark)

88. Some experts presented data silos as major barrier to data sharing and re-use. The discussions at the Expert Workshop revealed that data silos were often reflecting, and an

---

25 G-NAF is Australia’s authoritative, geocoded address database, which contains more than 13 million Australian physical address records including geocodes, which are latitude and longitude map coordinates. As of today G-NAF was downloaded 3 million times.
inevitable consequence of, organisational structures and business processes however. Some data silos were also implemented as measure to protect privacy and intellectual property rights (IPRs).

89. The lack of a whole-of-government approach and of coherence in existing data governance frameworks was recognised by some experts as a major challenge that if addressed would help overcome barriers to data access across silos. Applying a coherent data governance framework across data silos would also do better justice to use of data silos as technical and organisational measures to protect data against illegitimate access.

90. Experts also discussed three (technical) approaches for overcoming data silos and enhancing interoperability, which are presented in more details in the next sections: (i) standards, (ii) APIs, and (iii) data sandboxes.

**Standards and interoperability**

91. One of the most frequently discussed barriers to data sharing and re-use was the lack of standards (or the proliferation of incompatible standards as some participants noted) including at international level. Some participants highlighted, for instance, that inconsistent data formats and other data related standards were a barrier to the construction of longitudinal datasets, where changes in measurement and collection practices would make it hard to compare and aggregate data.

92. Many experts stressed the key role of platforms for the (bottom-up) development and promotion of data related standards (see section on “the role of data intermediaries and trusted third parties”). As Øyvind Grinde (Difi) highlighted, the lack of a common data format across municipalities was one of the reasons why end-users including businesses relied on data brokers, who used a common data format across municipalities and assured the (syntactic and semantic) compatibility of all their data sets. Experts also showed that major platforms have developed standards that were now used as de facto standards in some areas. One example is Google's General Transit Feed Specification (GTFS), a common format for public transportation schedules and associated geographic information.

93. The question of data standards turned out to go beyond the question of data formats. Babak Jahroumi (IT Standards Architect, Microsoft) stressed that, while the use of a commonly-used format (such as csv or json) may enable data syntactic portability, it did not guarantee data semantic portability. As the expert explained: data syntactic portability was about the transfer of “data from a source system to a target system using data formats that can be decoded on the target system”\(^{27}\), while data semantic portability was “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.” In addition to common

\(^{26}\) Many organisational information systems, and legacy systems in particular, were not developed and implemented such to reflect most recent data governance requirements Ruth Boardman (Partner, International Privacy and Data Protection Group) stressed, for example, that many organisational information systems were structured in such a way that would make data portability difficult to implement.

\(^{27}\) As Babak Jahroumi (Microsoft) further explained: In this case “data is structured according to the data model defined by the semantic facet, and is encoded using a particular syntax, such as XML.”
machine readable data formats, data semantic portability required mutually understood ontologies and metadata such as OWL and the Dublin Core Schema\textsuperscript{28} to assure a common meaning of the data, the expert explained. The expert later also introduced the concept of “data policy portability”, which was defined as “the ability to transfer data between a source and a target while complying within the legal, organizational, and policy frameworks applicable to the source and target. This includes regulations on data locality, rights to access, use and share data, and mutual responsibilities with respect to security and privacy between a [cloud service provider] and a [cloud service customer]”.  

94. Participants at the Expert Workshop highlighted interoperability as one of the major benefits of the use of standards. Mark MacCarthy (Georgetown University) highlighted the following benefits of interoperability (in the context of data portability): (i) reduced costs and increased convenience for customers; (ii) lower barriers to switch services; (iii) even greater competitive pressure on services to provide high quality, a variety of features and low prices; and (iv) access to larger installed base of established companies, which would facilitate market entry for new companies.

95. The discussion about interoperability raised a number of controversial issues however. Some participants noted that the lack of obligations for interoperability in the context of Art. 20 GDPR on data portability, could also have detrimental effects on users. Robin Wilton (Internet Society), for instance, stressed that the lack of interoperability and compatibility requirements could lead to a race to the “lowest common denominator” of standard data sets provided by data controllers. Mark MacCarthy, on the other hand, expressed concerns about, and presented some main challenges of, strong interoperability and compatibility requirements. He highlighted that universal requirements to interoperate with all other services would be expensive with uncertain benefits for most users and that burden would fall disproportionately on start-ups and small and medium sized enterprises (SMEs), who would have to enter the market with systems in place to interoperate with all other systems already on the market. In cases where all competing services would have to have common features and functions, this would even result in less variety and feature competition in the marketplace, the expert argued. This in turn would reduce consumer choice and innovation.

*From ad-hoc downloads to application programming interfaces (APIs)*

96. A number of experts presented their infrastructure for data sharing where APIs\textsuperscript{29} played an important role, often together with cloud computing. In the case of TfL, for example, a unified API was used as common gateway for TfL data: TfL powered their own website with the same data and the same APIs as provided to third party developers (Figure 9). The infrastructure was built in such a way to allow new data sets to be easily

\textsuperscript{28} The Dublin Core Metadata Terms were endorsed in IETF (Internet Engineering Task Force) RFC 5013 and ISO (International Organization for Standardization) Standard 15836-2009 [see Box 2.6 in (OECD, 2015[6])]

\textsuperscript{29} As applications increasingly rely on data, accessing data without human intervention becomes essential. Application programming interfaces (APIs) enable service providers to make their digital resources (e.g. data and software) available over the Internet. APIs thus act as interface that enables the smooth interoperability of the different actors, their technologies and services.
included from different systems and sources and to be updated efficiently, and by using the cloud the infrastructure was also scalable.

Figure 9. How data is made open by TfL via a single API


97. A key advantage of an API (compared to an ad-hoc data downloads) is that an API enables a software application (or App) to directly use the data required (the example of TfL enabling third party apps was already described above). Data controllers can also implement a number of restrictions via APIs so to better control the use of their data. They can control the identity of the API user, 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 / personal information.

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

98. The term “data sandbox” is used in this paper to describe any isolated environment presented at the Expert Workshop, through which data is accessed and analysed, and analytic results are only exported, if at all, when non-sensitive. 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 controller (where the data is located). Data sandboxes would typically require that the analytical code is brought to the data (in contrast to bringing the data to the code as is the case with downloads and most API based data access). Compared to the other data access mechanisms presented above, data sandboxes offer the strongest level of control as illustrated by the cases discussed as the Expert Workshop. Data sandboxes are therefore promising for providing access to very sensitive/personal data. MasterCard’s Data Fellows program presented in the section above on “restricted data sharing communities” is an example of a data sandbox realised through physical on-site presence in combination with a segregated, isolated data processing environment, that is in addition monitored by MasterCard’s corporate security team.

99. Another example is the Centers for Medicare and Medicaid (CMS) Virtual Research Data Center (VRDC), which was presented by Niall Brennan (HCCI).30 The

30 See https://www.resdac.org/cms-data/request/cms-virtual-research-data-center
CMS VRDC is a virtual research environment that provides timelier access to Medicare and Medicaid program data (such as beneficiary level protected health information). Researchers working in the CMS VRDC have direct access to approved data files and are able to conduct their analysis within the CMS secure environment. They have the ability to download aggregated reports and results to their own personal workstation. Researchers can in addition upload external data files into their workspace to link and analyse their data with the approved CMS data files. Access is provided over a Virtual Private Network (VPN) and a virtual desktop to satisfy all CMS privacy and security requirements.

100. Another example of data sandboxes involving private sector actors was Flowminder.org, an initiative combining new types of data to support the most vulnerable people in low- and middle-income countries. Erik Wetter (Stockholm School of Economics; Flowminder.org) showed how Flowminder.org relied on a secured access to personal data in particular mobile operators’ Call Detail Records (CDRs) including de-identified low-resolution location data (on nearest tower location) (Figure 10). In order to assure the privacy of their users, mobile operators hosted separate dedicated servers behind their firewalls, on top of which Flowminder.org researchers would conduct analyses under the operators’ supervision, keeping the data always within the operators’ servers. Only non-sensitive aggregated mobility estimates were able to be exported outside the operators’ servers. Erik Wetter stressed that this configuration allowed sensitive data to remain in the mobile operators’ custody and control, which helped to minimize privacy, security and commercial concerns.

Figure 10. Flowminder.org’s data processing flow

Source: Slides presented by Erik Wetter, Professor, Stockholm School of Economics; Chairman, Flowminder.org.

Acknowledging context dependencies and the heterogeneity of data

“‘Enhanced Access’ implies semantics and context.” (Robin Wilton, Technical Outreach for Identity and Privacy, Internet Society)

101. The Expert Workshop revealed that there were still a number of misconceptions or at least a lack of clarity cluttering the policy (and business management) debate around data, and data access and (re-)use. Erik Wetter (Stockholm School of Economics; Flowminder.org), for example, highlighted that “enhanced access” could mean different
things: for example, downloading data or remotely accessing the data via a site or an API, on a one-time or continuous basis. He also stressed that the scale, scope and timeframe of access could differ significantly as well as the level of (de-) identification and (de-) aggregation of the data.

102. Similarly, Robin Wilton (Technical Outreach for Identity and Privacy, Internet Society, United Kingdom) emphasised that, when talking about enhanced access to data, one would need to clarify: “Access by whom, for what, and enhanced in what way, and to what end?” He therefore concluded that “enhanced access” was a term that incorporated a lot of semantic and contexts that needed further differentiation and precision. According to him, the focus on data was misleading, and he advised instead to focus instead on information, which would imply, for example, to also consider the role of meta-data, and the issue of information asymmetry as a manifestation of, and cause for, power asymmetry.

103. When presenting the Swiss Personalised Health Network (SPHN), which aims at enabling nationwide accessibility and exchange of health data of patients and healthy citizens, Effy Vayena (ETH Zurich) also highlighted the heterogeneity of the stakeholders and the data involved. She showed that the SPHN typically involved health service providers, the research and academic community, as well as the healthcare and the ICT industry, and last but not least individuals and the government. In terms of data, she explained that the SPHN would include standard health data (such as health service data, public health data and research data) as well as “expanded sources” (such as environmental data, lifestyle and socioeconomic data and behavioural and social data). This led to a complex data ecosystem with a wide range of issues affecting data sharing and re-use at different levels such as privacy and security, ownership and control, interoperability, and data quality, the expert explained.

104. A number of experts stressed the commercial, legal and cultural environment as a major contextual factor to be considered as well. For example, Mark MacCarthy (Communication, Culture and Technology Program, Georgetown University) stressed that data portability took place in a wide range of contexts including business-to-business (B2B), cloud computing services, government-provided services, consumer web services, mobile apps, and automated data processing systems. Each raised different issues depending on the context, the expert explained. Data portability in the case of B2B, for instance, would need to be considered differently than in a business-to-consumer (B2C) context, the expert argued (see section on data portability).

105. JoAnn Steiner (MasterCard) highlighted that the legal complexity of data sharing made data governance of private sector data very challenging. She explained that the use of private sector data required careful consideration to ensure compliance with various laws and regulations in addition to business, reputational and ethical concerns which needed to be taken into account for each and every project.

106. Erik Wetter (Stockholm School of Economics; Flowminder.org) confirmed that taking the commercial, legal and cultural context into account was necessary, but also a reason why data governance was very complex. He therefore cautioned about the “data policy pitfall” that resulted from (i) a lack of clear definitions, (ii) the tendency to look for one solution (silver bullet) to a multidimensional problem, (iii) the lack of consideration of second and third order effects, and (iv) the tendency to develop policies affecting private data without industry dialogue.
107. Brett Frischmann (Villanova University) also stressed that the governance issues to be addressed were in most cases contextual, resource-specific and community-specific. The complexity of governance was exacerbated by the fact that knowledge commons\(^{31}\) were often clustered in multiple ways as well as nested within each other. Many knowledge communities (e.g. scientific communities) defined, and in turn, were defined by, the knowledge commons. Patient-related commons, for instance, were often nested together with clinician/researcher-related commons, infrastructure and tools-related commons, and data and analytics-related commons. The analysis of commons and their contextual factors would therefore require a more systematic analysis, for which the Ostrom approach was most promising, Brett Frischmann argued (see Box 1).

108. The context dependencies highlighted at the Expert Workshop strongly suggested that data governance frameworks needed to better take into account the different types of resources and institutions involved. These would not only include the different approaches to data access, which initially structured the sessions of the Expert Workshop (open data, data sharing communities, data markets, and data portability), but also the different types of actors/communities as well as the different data types, data processing and data use categories, which are presented further below.

109. A better differentiation of all relevant contextual factors could lead to a more accurate picture of the context affecting data governance, which would allow for more differentiated approaches to data access and re-use and would enable a more effective management of the associated risks. It would also help address the challenges raised by the change of context in some cases of data re-use, in particular in respect to the rights of data subjects and IPR holders. As Robin Wilton (Internet Society) stressed, enhanced access to data (and data portability in particular) were about taking data from one context and transferring it to another context. Referring to (Nissenbaum, 2004\(^{12}\)) on privacy as contextual integrity, Robin Wilton argued that the change of context made it challenging to ensure that existing rights and obligations were not (accidently) undermined, for instance, when privacy assumptions that were implicit in the initial usage, no longer applied in subsequent uses.

31 According to (Hess and Ostrom, 2007\(^{18}\)), “knowledge […] refers to all intelligible ideas, information, and data in whatever form in which it is expressed or obtained”. Daniel Bell defines information as “data processing in the broadest sense” and knowledge as “an organized set of statements of facts or ideas […] communicated to other”, see also (Madison, 2014\(^{17}\)). Commons were defined as “resource management / governance institutions that enable sustainable shared use of certain resources within a community”.

Unclassified
**Box 1. The Ostrom approach to the analysis of knowledge commons**

Brett Frischmann (Villanova University) presented the institutional analysis and development (IAD) framework for structured case study analysis, a framework developed by Elinor Ostrom for her analysis of economic governance, especially the commons, of natural resources, for which she was awarded the 2009 Nobel Prize in Economic Sciences. Brett Frischmann explained that applying the IAD framework to knowledge resources would require addressing the following questions:

<table>
<thead>
<tr>
<th>I. Background environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What is the background context (legal, cultural, etc.) of these particular commons?</td>
</tr>
<tr>
<td>2. What is the “default” status of the resources involved in the commons? Patented? Copyright? Open?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Attributes of the commons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Resources</td>
</tr>
<tr>
<td>a. What resources are pooled and how they are created or obtained?</td>
</tr>
<tr>
<td>b. What are the characteristics of the resources, such as whether they are rival or non-rival, whether they are tangible or intangible,</td>
</tr>
<tr>
<td>c. What technologies and skills are needed to create, obtain, maintain and use them?</td>
</tr>
<tr>
<td>2. Community Members</td>
</tr>
<tr>
<td>a. Who are the community members and what are their roles?</td>
</tr>
<tr>
<td>b. What are the degree and nature of openness of the community with respect to each type of community member and the general public?</td>
</tr>
<tr>
<td>3. Goals and objectives</td>
</tr>
<tr>
<td>a. What are the goals and objectives, including obstacles or dilemmas to overcome?</td>
</tr>
<tr>
<td>b. What are the history and narrative of the commons?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Governance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What are the governance mechanisms of the commons (e.g., membership rules, resource contribution or extraction standards and requirements, conflict resolution mechanisms, sanctions for rule violation)?</td>
</tr>
<tr>
<td>2. Who are the decision-makers and how are they selected?</td>
</tr>
<tr>
<td>3. What are the institutions that govern decision-making?</td>
</tr>
<tr>
<td>4. What informal norms govern the commons?</td>
</tr>
<tr>
<td>5. How do non-members interact with the commons? What institutions govern those interactions?</td>
</tr>
<tr>
<td>6. What legal structures (including intellectual property rules, subsidies, contract and licensing law, antitrust provisions) govern the functioning of the commons?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV. Patterns and outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What benefits are delivered to members and to others (including innovations and creative output, production, sharing, and dissemination of those innovations and output to a broader audience, and social interactions that emerge from the commons?)</td>
</tr>
<tr>
<td>2. What costs and risks are associated with the commons, including, for example, any negative externalities?</td>
</tr>
</tbody>
</table>

*Source:* Slide presented by Brett Frischmann (The Charles Widger Endowed University Professor in Law, Business and Economics, Villanova University, United States).
**Data categories**

110. The presentations given at the Expert Workshop strongly suggested that the type of data significantly determines not only the relevant approach to enhanced access to data, but also the governance issues to be addressed and the possible means to address them. The data spectrum introduced by Jeni Tennison (ODI) (Figure 1), for instance, highlighted how different types of data could be associated to different access mechanisms.

111. Experts agreed that the concept of “data” needed further clarity. Federica Rosetta noted that data may refer to machine and environmental settings, scripts and algorithms, raw data or processed data, as well as protocols, methods and workflows. Data may thus be classified according to a number of different dimensions. Data traded on Dawex, for example, had been clustered as follows:

- industrial data vs personal data;
- raw data vs. aggregated vs. insights data;
- historical vs real-time data;
- structured vs. unstructured data;
- bulk data vs. on-demand data; and
- regulated vs. unregulated data.

112. Some data categories have more policy implications than others. Personal data requires more restrictive access rights than, for example, non-personal administrative data. Experts also acknowledged the increasing uncertainties and complexities with the definition of personal data and the blurring distinctions between personal and non-personal data in the context of big data analytics (see next section on “Data identification, processing and use categories”). As an illustration, Mark MacCarthy highlighted the uncertainties about the interpretation of the right to data portability under the GDPR (Art. 20), which only applies to personal data provided by the data subject to a data controller. This provision, according to the Article 29 Working Party (Art. 29 WP) should be interpreted to include also “observed data” about a data subject, which may be generated passively through interactions with a service. It would however exclude data derived from additional processing (“inferred data”) that are considered often proprietary.

113. In a follow-up presentation, Babak Jahroumi (IT Standards Architect, Microsoft) presented the data taxonomy section of the ISO/IEC 19941 standard, which aims to achieve interoperability and portability in cloud computing by establishing common terminology and concepts (Figure 11). It includes the data category taxonomy (Figure 12), which differentiates between:

- **Customer data**, that is mainly contributed data from a user of a cloud service provider (e.g. credentials, personal health data and medical records, and financial details);
- **Derived data**, that is observed and/or inferred data about user;
- **Cloud service provider data** including mainly operations data and access and authentication data; and
- **Account data**, including mainly account or administration contact information and payment instrument data.
Personal data and data identification, processing and use categories

114. Some presentations at the Expert Workshop suggested that the amount of information conveyed by the data was maybe as important as, if not more important than, the categorisation by data types. In the context of personal data, the information conveyed was determined by the different degrees of anonymization and aggregation, and the related data processing activities. In his presentation, John Foster (Telefónica), for instance, explained that the creation of insights (knowledge) from personal data would require the data to be anonymised and then aggregated.
115. Data would thus typically undergo different states of data identification. Babak Jahroumi highlighted that ISO/IEC 19441 distinguished between five categories or states of data identification. These were aligned with the privacy enhancing de-identification techniques included under the ISO/IEC 20889 standard (Figure 13).

**Figure 13. Mapping data identification qualifiers with de-identification techniques**

<table>
<thead>
<tr>
<th>ISO/IEC DIS 19944 data identification qualifiers describing state of data</th>
<th>Privacy enhancing data de-identification techniques whose application leads to the corresponding state of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified data</td>
<td>original, unprocessed data containing identifiers; in other words, no de-identification techniques are applied yet; for the other qualifiers, the identifiers are removed (masked)</td>
</tr>
<tr>
<td>Pseudonymized data</td>
<td>data processed using pseudonymization techniques with controlled re-identification possible/implemented</td>
</tr>
<tr>
<td>Unlinked pseudonymized data</td>
<td>data processed using pseudonymization techniques with no controlled re-identification allowed</td>
</tr>
<tr>
<td>Anonymized data</td>
<td>data processed using generalization and/or randomization techniques</td>
</tr>
<tr>
<td>Aggregated data</td>
<td>data processed using aggregation techniques</td>
</tr>
</tbody>
</table>

*Source: Slide presented by Babak Jahroumi (IT Standards Architect, Microsoft)*.

116. The ISO/IEC 199441 standard also introduced two more data taxonomies: a data processing taxonomy and a data use taxonomy. The former listed the basic measures for transforming data and included: (i) fragmentation, (ii) integration (linkage), (iii) data fusion (merging), (iv) encryption, (v) replication, (vi) deletion, and (vii) re-identification (Figure 11). The latter, in contrast, reflected the different ways how data could be (re-) used, that is whether data was used to: (i) provide a service, (ii) improve a service, (iii) personalise a service, (iv) market and advertise another product, or whether data was (v) shared with third parties.

**Fostering investments for the sustainable provision and re-use of data**

“The biggest barriers to data sharing are the perverse incentives in the reward and evaluation systems that make authors reluctant to share” (Catriona J. MacCallum, Director of Open Science, Hindawi; Member of the Boards OASPA & OpenAire)

117. Experts agreed that the provision of data required upfront and continuous investments. According to a study presented by Rikesh Shah (TfL), for instance, the annual costs for TfL of publishing open data was estimated to be around GBP 1 million. A significant part of the costs presented by experts was related to maintenance and facilitation (e.g. the engagement of data users). Some experts also stressed the significance of co-ordination costs due to frictions and poor interoperability as well as possible opportunity costs due to unrealised innovation.

118. Given the significant cost for sustaining the provision of data, (i) providing dedicated funding for data infrastructures (such as in science), (ii) better aligning existing incentives structures for data curation and sharing, and (iii) strengthening skills and
competences for data analysis and data sharing were highlighted by experts as essential for the sustainable provision of data.

**Funding for data infrastructures**

119. Experts stressed that sustainable funding was a major concern for public and scientific data repositories. Niall Brennan (HCCI), for instance, highlighted the lack of dedicated funding for data sharing infrastructures and the limited pathways for their sustainment as main challenges, and Catriona J. MacCallum (Hindawi; OASPA & OpenAire) presented strong evidence for the increasing erosion of data over time. She highlighted a study by (Vines et al., 2014\(^{13}\)), according to which the probability of finding the data associated with a paper declined by 17% every year (Figure 14).

**Figure 14. The problem of data erosion**

![Graph showing probability of finding data associated with a paper over time]

*Source:* (Vines et al., 2014\(^{13}\)), presented by Catriona J. MacCallum (Director of Open Science, Hindawi; Member of the Boards OASPA & OpenAire).

120. Catriona MacCallum recommended that institutions (including funding agencies) needed to provide or fund the infrastructure needed for data stewardship and data sharing, as well as the necessary training to support these activities (see following section on "strengthening skills and competences for data curation, sharing and analysis"). She explained during the discussion that PLOS had a pay-to-publish business model, where the institutions, and not necessarily the authors, were expected to pay to make research outputs (including data and publication) open.

**Aligning incentives structures for a sustainable provision and re-use of data**

121. Experts agreed that better aligning existing incentives structures was crucial for data related investments (e.g. data curation) and data sharing, but also complex. As Brett Frischmann (Villanova University) highlighted “knowledge commons governance often did not depend on one strong type or source of individual motivations for cooperation.” Diverse obstacles or social dilemmas had to be overcome, many of which were not well described or reducible to the simple free rider dilemma\(^{32}\), the expert highlighted.

---

\(^{32}\) As (Frischmann, 2012\(^{20}\)) explains: “There is a mistaken tendency to believe that any gain or loss in profits corresponds to an equal or proportional gain or loss in investment incentives,
122. Catriona MacCallum confirmed that there was a need to better aligning incentives at multiple levels, including government, funder and institutional policies, to support data sharing. She stressed that data stewardship and data sharing therefore needed to be enshrined within the grant application process. However, she also noted that the willingness to do so would depend on the demography of authors and the status of journals: Some publishers were reluctant to impose data sharing policies as it could make them less competitive given that some authors could start submitting to journals without such policies in order to bypass any data sharing obligations.

123. Catriona MacCallum also criticised the perverse incentives in current reward and evaluation systems, which she highlighted as the main reason why authors were reluctant to share scientific data. She called for rethinking the way scientific research is evaluated and presented innovative practice of the ASCB (American Society for Cell Biology), which – together with scholarly journals and funders – has put in place an initiative to:

1. improve the way in which the outputs of scientific research are evaluated;
2. eliminate the use of journal based metrics, such as Journal Impact Factors, in funding, appointment, and promotion considerations; and
3. assess research on its own merits rather than on the basis of the journal in which the research is published.

124. Enabling data citation and taking its impact into account for rewarding scientists were highlighted as promising incentives for the scientific community. Federica Rosetta (Director Global Strategic Networks, Europe, Elsevier), for instance, presented how the Mendeley Data Platform could encourage authors to better cite and access data. The Platform was encouraging the implementation of the FORCE 11 data citation principles which are based on the “dual necessity of creating citation practices that are both human understandable and machine-actionable”\(^3\). 

**Strengthening skills and competences for data curation, sharing and analysis**

125. Very few experts discussed the need to strengthen skills and competences for data re-use (analysis), curation and sharing. Among the few, Catriona MacCallum explained that data infrastructure investments needed to be accompanied with training to support data stewardship and data sharing. Brett Frischmann, as another example, showed that the IAD framework for analysing knowledge common (see Box 1) includes the assessment of the technologies and skills needed to create, obtain, maintain and use the relevant resource (data).

**The role of governments and possible priorities for further work by the OECD**

“Reconciling the risks and benefits of data sharing is a global issue, and it requires a global solution. So the conversation needs to be global. As we are but this belief greatly oversimplifies the decision-making process and underlying economics and ignores the relevance of alternative opportunities for investment. The conversion of surplus realised by a free rider into producer surplus may be a wealth transfer with no meaningful impact on producers’ investment incentives or it may be otherwise, but there is no theoretical or empirical basis for assuming that such producer gains are systematically incentive-relevant.”

\(^{33}\) See [https://www.force11.org/datacitationprinciples](https://www.force11.org/datacitationprinciples)
“doing here in Copenhagen, we need to get all of the stakeholders at the table – government, civil society, the private sector, scientists and academics.” (Douglas Frantz, Deputy Secretary General, OECD)

126. During the closing session, policy experts gave their final remarks where they highlighted their key insights from the Expert Workshop on the role of governments and possible priorities for further work by the OECD. These are presented in the following sections, clustered around six themes that can be summarised as: (i) leading by example by opening publicly funded data, and (ii) promoting data policy frameworks that (iii) strike the right balance between the private interests of individuals and organisations, and the public interests for data openness, while (iv) focussing on strategic societal objectives, (v) strengthening data-related skills and competences, and (vi) encouraging a whole-of-society engagement though PPPs and multi-stakeholder dialogues.

Governments leading by example by opening publicly funded data

127. Some policy experts stressed that governments should act as role model in the economy by enabling the re-use of publicly funded data across society. Torsten A. Andersen (Deputy Director General, Danish Business Authority), for instance, highlighted that governments should actively promote the use of data by (i) making more publicly funded data public and (ii) improving the quality of their data. He argued that good examples were needed on the re-use of data from both, the public and private sector, for showing the potential of enhanced access to data, and for encouraging businesses and public authorities to make the investments necessary for opening and sharing data. He highlighted the government initiatives in Australia presented by Tanja Cvijanovic (Department of Prime Minister and Cabinet, Australia) as inspiring examples that should not be reserved to governments alone.

128. Katarina de Brisis (Deputy Director General, Ministry of Local Government and Modernisation, Norway) emphasised that there should be an obligation to publish data in the public sector (including via APIs) and that guidelines therefore would be needed. She explained that access to data in Norway was still based on requests under the freedom of information act. The critical role of open access to data, including via open APIs, was also highlighted by Jenni Nordborg (Head of Health, Swedish Governmental Agency for Innovation Systems), who stressed that open APIs should become a funding criteria in science similarly to open data.

Strengthening data-related skills and competences

129. Policy experts highlighted the importance of skills and competencies for data sharing and re-use. Manuela Siano (Italian Data Protection Authority), for instance, stressed that educational campaigns were crucial to help increase the skills base necessary to leverage big data. Jenni Nordborg further stressed that data related education would have to start early to engage and prepare the young generation for the digital paradigm shift. Katarina de Brisis also highlighted that the increased value creation from data and the need to better handle privacy and digital security risks required new knowledge and skills.

Promoting data policy frameworks for enhanced access to data

130. All policy experts agreed that better and more coherent data governance frameworks would be needed. Jenni Nordborg, for instance, stressed the need for a
strategy and action plan for interoperability within data commons to assure data linkage across different levels of data openness. As example, the policy expert highlighted health data, where open data needed to be combined with (mostly sensitive) data made available through more restricted access mechanisms. She also explained that the interoperability of fragmentized and decentralized systems was of high importance. There was in particular a need for improving national incentive structures and harmonizing data governance policies across institutions within the existing ethical frameworks for responsible data processing. In this context, Jenni Nordborg welcomed the example of the SPHN presented by Effy Vayena (ETH Zurich).

131. Daisuke Nagasaki (International Affairs Office, Commerce and Information Policy Bureau, Ministry of Economy, Trade and Industry [METI], Japan) underlined the development of basic mechanisms for facilitating data sharing and re-use as a condition for effectively promoting data sharing and re-use across society. He explained that METI had formulated a guiding document for contracting the rights to data utilization among concerned businesses to clarify the rights of data holders and users, including in respect to data ownership. He also explained that global guidelines like the (OECD[14]) Council Recommendation concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (Privacy Guidelines) would be needed to promote data sharing as companies’ global supply chains were now driven by data. Daisuke Nagasaki then expressed Japan’s interest in sharing its guiding document and invited the OECD to develop comparable guidelines.

132. Similarly, Torsten A. Andersen argued that policymakers should strive to improve the regulatory framework by reducing unjustified barriers, removing unnecessary burdens and making all existing legislations fit for the digital age. He argued that new regulation was not necessarily the right way forward, and where necessary their development should be based on thorough analyses. Ensuring future-proof ways of dealing with the emerging data governance issues was more critical and non-regulatory approach such as the promotion of data related standards were promising.

133. Anne Kauhanen-Simanainen (Ministerial Advisor, Ministry of Finance, Finland) also stressed that there was a need for more coherent and comprehensive data and information policies, and that international cooperation was needed in developing the conceptual frameworks for data and information policies. She invited the OECD to develop clear definitions and taxonomies to assure a common language and understanding of the main concepts and advised the OECD to monitor the developments in member countries, when developing a data governance framework and common principles and guidelines.

134. Katarina de Brisis advised to also acknowledge the context dependencies and the heterogeneity of data, which mattered in particular for private sector data sharing. As example, she highlighted geographical data and research data where the governance issues to be addressed (such as the type of incentives to share) would be different.

**Striking the right balance between private and public interests**

135. In their closing remarks, policy experts addressed the question of how to strike a right balance between the legitimate interests of the individuals (data subjects) and business concerned and those controlling (data controllers) and/or using the data (data users), from a national and international perspective. Two topics were raised in this context in particular: issues related to (i) data ethics and (ii) the free flow of information.
Data ethics

136. Manuela Siano (Service for EU and International Matters, Italian Data Protection Authority), for instance, argued that policy recommendations should take a balanced view on data-driven innovation to (i) stimulate best practices, (ii) increase consumer awareness and (iii) protect personal data and privacy. She stressed data ethics as an area where policy makers could play a major role by creating the conditions for the promotion of company accountability to data protection and privacy standards articulated in the Privacy Guidelines and the (European Union, 2016[4]) GDPR. Manuela Siano also argued that providing more control rights to the individuals could be disruptive for the data ecosystem as it would lead to the democratization of the control over personal data where changes would be driven by the public, and the younger generation in particular.

137. Torsten A. Andersen stressed that ensuring the protection of individuals and organisations’ rights and interests would require thorough analysis and the development of solutions that promote access to, and reuse of, data, while mitigating the potential risks without unnecessarily burdening businesses and authorities. He also stressed that there was not necessarily a trade-off between enhanced access to data on one side and trust and privacy on the other side. In some cases, data sharing and re-use across states and sectors would contribute to better privacy protection, for instance when used to provide access to more secured cloud solutions. Torsten A. Andersen however also highlighted that there was a need to address issues related to data ethics, in particular in the context of AI, where major issues emerged not only related to the collection of data, but in particular to how the data was used. Given that data ethics could not be addressed only at a national level, further international co-operation would be required.

Barriers to the free flow of information.

138. Daisuke Nagasaki explained, there was a risk that data localization requirements that emerged in some countries could prevent data sharing and re-use. He advised that these barriers should be opposed where they were not justified and primarily used as illegitimate trade barriers. He also emphasized the need for keeping international discussions ongoing on this subject matter, and stressed the importance of ensuring the basic principle of the free flow of information across borders articulated in the Outcomes of the G7 ICT Ministers’ Meeting in Takamatsu, Kagawa. Daisuke Nagasaki then invited the OECD to continue to find a consensus on how to assure that innovation was not hindered while respecting the interests of individuals and organizations concerned.

Focussing on strategic societal objectives

139. Policy experts confirmed the importance of focussing strategically on societal objectives and the data that matter therefore. Anne Kauhanen-Simanainen, for instance, argued that data sharing initiatives should focus on high-priority issues to intensify the value cycle from data to information and knowledge, nationally and internationally to increase the value for individuals, society and the economy.

140. Similarly, Jenni Nordborg urged to move from talking about access to data, to instead elaborating on the problem that needed data collaboration to find better solutions. She argued that the public sector should better formulate and communicate the questions to be addressed so that both, societal challenges could be better tackled and new business opportunities be stimulated.
141. Daisuke Nagasaki also argued for a strategic approach to data sharing and re-use. The policy expert stressed that communication to the public was effective for promoting data sharing and re-use. He highlighted the “Connected Industries” initiative that Minister Seko is advocating since March 2017 as a model for future Japanese industries and underlined that the initiative’s focus on “solutions for social challenges” was a widely acceptable and sharable vision that facilitated data sharing in Japan. Daisuke Nagasaki invited the OECD to continue its analysis and the discussion on data flows and to provide further opportunities for counties to mutually study data governance issues by gathering and analysing cases and other empirical evidence on promising private sector initiatives and innovative policy practices.

**Encouraging a whole-of-society engagement and multi-stakeholder dialogues**

142. All policy experts agreed that whole-of-society engagement and multi-stakeholder dialogues were essential for encouraging data sharing and re-use. Torsten A. Andersen, for instance, highlighted that collaboration between the public and the private sector through public-private partnership (PPPs) would be needed to make sure that major societal goals are attained. Participants agreed that sharing best practices and providing further evidence on the costs and benefits of data access and re-use, including case study analysis, was an important contribution of the OECD.

143. Daisuke Nagasaki also stressed that developments of policies should be undertaken not only by the government, but in a multi-stakeholder approach and should involve PPPs. He explained that METI facilitated multi-stakeholder discussions in five priority working areas, such as (i) autonomous driving and mobility service, (ii) manufacturing and robotics, (iii) biotechnology and material, (iv) industrial plant and infrastructure safety, and (iv) smart life. In each working area, discussions on what kind of data could be shared and how such data could be re-used were ongoing.

144. Although there was a consensus about the importance of PPPs, some policy experts noted that the responsibilities of the different stakeholders were not clear. Anne Kauhanen-Simanainen, for instance, stressed that it was critical to clarify the responsibilities and roles by governments (for example by the ministries), private sector actors (including non-for profit organisations) on issues such as data access, data quality and data related skills and competences. She urged to recognise that there were many different interests involved, which would have to be recognised and made transparent.

145. Jenni Nordborg also stressed the need for a holistic strategy, including a whole-of-government approach and cross-sectoral collaboration. She noted however that having policy dialogues with everyone around the table was not sufficient, in particular where there were strong business interests. She concluded that economic incentives were needed to move from a “one-company philanthropy” model for data sharing to an open data sharing community including competing firms. She invited the OECD to showcase innovative practices including private sector initiatives such as in the European aviation industry.

146. Katarina de Brisis argued that the governments should give directions and develop policy on open data, which would involve data sharing between public and private sectors, data ownership and regulations on open data. Private sector, overall or by industrial sectors, should work to establish data intermediaries (platform) and best practices for open data in the private sector. She highlighted the recent open data legislation in France, which puts in place provisions obliging businesses to open up
“public interest data”\textsuperscript{34}, and invited the OECD to explore the best practices to learn from the experience in France.

147. In her closing remarks, Katarina de Brisis concluded that changing the data culture was urgent to encourage further data sharing, and could require a regulatory response. PPPs were a promising way forward to involve private companies to share data where public goods were concerned, and she highlighted that this was different from “data philanthropy”. Engaging relevant public and private entities in a multi-stakeholder dialogue was in particular needed also to ensure the protection of individuals and organisations’ rights and interests. Katarina de Brisis then concluded that national strategies for the re-use of data that address data governance issues in a comprehensive way would be needed, and that a coordinated voice, possibly by the OECD, would give all countries the tools to argue for, and move forward, their national policies on enhanced access to data.

\textsuperscript{34} “Public interest data” (“données d'intérêt général”) are defined by Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique, JO République Française n°0235 of 7 October 2016. These data include data generated in the context of procurement (article 17), commercial data for the establishment of official statistics (article 19), certain electricity and gas production and consumption data held by transmission and distribution systems operators for re-use by any other party (article 23), and certain data relating to changes in real estate ownership for re-use by certain third parties (article 24).
References


Annex A. Agenda

Monday 2 October 2017

08:30 - 09:00  Registration of participants and coffee

09:00 - 09:45  Opening Session

Welcome remarks by:
- Mr. Søren Gaard, Deputy Permanent Secretary, Ministry of Industry, Business and Financial Affairs
- Mr. Douglas Frantz, Deputy Secretary General, OECD

Chair  Remarks by the Chair and introduction of key note speaker
- Ms. Katarina de Brisis, Deputy Director General, Norwegian Ministry of Local Government and Modernisation, Chair of the OECD Working Party on Digital Security and Privacy

Key Note Speaker  • Mr. Niall Brennan, President and Executive Director, Health Care Cost Institute
<table>
<thead>
<tr>
<th>Time</th>
<th>Session 1: Open data - what is the economic and social rationale?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Questions</strong></td>
</tr>
<tr>
<td></td>
<td>- Where does open data make a difference? What are the conditions under which open data becomes an economic and social necessity? Where are potential unforeseen benefits? And what are the conditions under which data cannot be opened?</td>
</tr>
<tr>
<td></td>
<td>- What are the benefits and costs of open access to data? How can they be better assessed, in particular in light of existing potential (positive) externalities?</td>
</tr>
<tr>
<td></td>
<td>- How can existing initiatives be improved? What promising sustainability models are emerging? What cost recovery and revenue models are more pertinent in the public and private sector?</td>
</tr>
<tr>
<td></td>
<td><strong>Moderator</strong></td>
</tr>
<tr>
<td></td>
<td>Mr. Douglas Frantz, Deputy Secretary General, OECD</td>
</tr>
<tr>
<td></td>
<td><strong>Key Note Speaker</strong></td>
</tr>
<tr>
<td></td>
<td>Ms. Tanja Cvijanovic, First Assistant Secretary, Policy Innovation and Projects, Department of Prime Minister and Cabinet, Australia</td>
</tr>
<tr>
<td></td>
<td><strong>Panelists</strong></td>
</tr>
<tr>
<td></td>
<td>Mr. Øyvind Grinde, Head of section, Information Security and Data Sharing, Agency for Public Management and eGovernment (Difi)</td>
</tr>
<tr>
<td></td>
<td>Mr. Rikesh Shah, Lead Digital Partnerships Manager, Transport for London</td>
</tr>
<tr>
<td>10:45-11:00</td>
<td><strong>Coffee Break</strong></td>
</tr>
<tr>
<td></td>
<td>- Ms. Catriona MacCallum, Director of Open Science, Hindawi; Member of the Boards OASPA &amp; OpenAire</td>
</tr>
<tr>
<td></td>
<td>- Mr. Bob Bailey, Chief Information Architect, Corporate Technology, Thomson Reuters</td>
</tr>
<tr>
<td></td>
<td>- Ms. Jeni Tennison, CEO, The Open Data institute (ODI)</td>
</tr>
<tr>
<td></td>
<td><strong>Followed by discussion [30 min]</strong></td>
</tr>
<tr>
<td>12:15-13:15</td>
<td><strong>Lunch break</strong></td>
</tr>
</tbody>
</table>
### Session 2: Leveraging data sharing communities for cross-sectoral reuse of data

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:15-15:30</td>
<td>Questions</td>
</tr>
</tbody>
</table>
|              | • What data sharing communities are emerging today? And what is the rationale for their creation and for data sharing arrangement including “data donation” and “data philanthropy”?
|              | • How are the rules to access governed, particularly when depending on cross-sectoral arrangements? To what extent is access granted on equal or non-discriminatory terms?
|              | • How do these communities protect the interest of third parties including data users and donors (including e.g. consumers)? How are the interests and risks balanced between the different stakeholders involved? |

**Moderator**
Ms. Elettra Ronchi, Senior Policy Analyst, OECD Directorate for Science, Technology and Innovation; Digital Economy Policy

**Key Note Speaker**
• Mr. Brett Frischmann, The Charles Widger Endowed University Professor in Law, Business and Economics, Villanova University

**Panelists**
• Ms. Effy Vayena, The Swiss personalized health network (SPHN)
• Mr. Erik Wetter, Professor, Stockholm School of Economics; Chairman, Flowminder.org
• Ms. JoAnn Stonier, EVP and Global Privacy & Data Protection Officer, MasterCard
• Mr. Jakob Rehof, Director, Fraunhofer Institute for Software and Systems Engineering ISST
• Mr. Jean-Marc Lazard, Founder & CEO, OpenDataSoft

**Followed by discussion [30 min]**

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:30-15:45</td>
<td>Coffee break</td>
</tr>
</tbody>
</table>
15:45-18:00  Session 3: Towards efficient, global and trustworthy data markets

**Questions**

- How can data markets help determine the economic value of data? How is data priced?
- What are current barriers preventing the upscale of efficient global and trustworthy data market platforms? What are promising sustainable business models?
- What are the limitations of data markets? What are potential trust issues to be addressed?
- How can transparency be enhanced? How can platforms ensure users understand and feel in control of how their data is being used?

**Moderator**

Mr. Christian Reimsbach-Kounatze, Information Economist / Policy Analyst, OECD Directorate for Science, Technology and Innovation; Digital Economy Policy

**Key Note Speaker**

- Mr. Paul Hofheinz, President and Co-Founder, The Lisbon Council

**Panelists**

- Mr. Naoto Ikegai, Interfaculty Initiative in Information Studies, University of Tokyo
- Ms. Federica Rosetta, Director Global Strategic Networks, Europe, Elsevier
- Mr. Fabrice Tocco, Co-founder, Dawex
- Mr. Stephen Deadman, Global Deputy Chief Privacy Officer, Facebook
- Mr. Malte Beyer-Katzenberger, Policy Officer, Unit G3 - Data Value Chain, DG CONNECT, European Commission

Followed by discussion [30 min]

18:00- 
**Social event**

18:00-19:00  
*Copenhagen Canal Tour*

19:00-  
*Buffet dinner with*

**Welcome remarks by:**

- Mr. Torsten A. Andersen, Deputy Director General, Danish Business Authority

**Brief presentation on the Nordic SmartGovernment initiative by:**

- Mr. Carsten Ingerslev, Head of Digital Innovation, Danish Business Authority
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 - 9:00</td>
<td>Coffee</td>
</tr>
<tr>
<td>09:00 - 11:45</td>
<td><strong>Session 4: The potentials and limitations of data portability</strong></td>
</tr>
<tr>
<td>Questions</td>
<td>• What are the benefits and main challenges of data portability?</td>
</tr>
<tr>
<td></td>
<td>• What are the costs for implementing and maintaining data portability and who should bear them?</td>
</tr>
<tr>
<td></td>
<td>• How can consumers be further empowered to take advantage of data portability?</td>
</tr>
<tr>
<td>Moderator</td>
<td>Mr. Jan Krewer, Deputy General Secretary, French Digital Council</td>
</tr>
<tr>
<td>Key Note Speaker</td>
<td>• Mr. Mark MacCarthy, Communication, Culture and Technology Program, Georgetown University</td>
</tr>
<tr>
<td>Panelists</td>
<td>• Ms. Ruth Boardman, Partner, International Privacy and Data Protection Group</td>
</tr>
<tr>
<td></td>
<td>• Mr. Lenard Koschwitz, Director European Affairs, Allied for Startups</td>
</tr>
<tr>
<td></td>
<td>• Ms. Randi Flesland, Managing Director, Norwegian Consumer Council</td>
</tr>
<tr>
<td></td>
<td>• Mr. John Foster, Director 4th Platform (Data Strategy), Telefónica</td>
</tr>
<tr>
<td></td>
<td>• Mr. Babak Jahromi, IT Standards Architect, Microsoft</td>
</tr>
<tr>
<td></td>
<td>• Mr. Robin Wilton, Technical Outreach for Identity and Privacy, Internet Society</td>
</tr>
<tr>
<td>10:45 - 11:00</td>
<td><strong>Coffee Break</strong></td>
</tr>
<tr>
<td></td>
<td>Followed by Discussion [30 min]</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>11:30-13:00</td>
<td><strong>Closing Session</strong> highlighting the key insights from the four substantive sessions and possible priorities for further work.</td>
</tr>
</tbody>
</table>

**Questions**

- How can policy makers address how to promote data sharing and re-use across society and across countries and at the same time ensure the protection of individuals and organisations’ rights and interests?
- What are high-priority issues that policy makers need to address first?
- What possible roles could governments and private sector play, respectively?
- Where is international cooperation needed and what role should the OECD play?

**Moderator**

Ms. Anne Carblanc, Senior Policy Analyst, Head of Division, OECD Directorate for Science, Technology and Innovation; Digital Economy Policy Division

**Highlights of ongoing OECD work**

- Mr. Christian Reimsbach-Kounatze, Information Economist / Policy Analyst, OECD Directorate for Science, Technology and Innovation; Digital Economy Policy
- Ms. Barbara Ubaldi, Senior Policy Analyst, OECD Directorate for Public Governance and Territorial Development; Digital Government Project
- Mr. Alan Paic, Senior Policy Analyst, OECD Directorate for Science, Technology and Innovation; Science and Technology Policy

**Panelists**

Final remarks by country representatives and OECD Secretariat [50 min]:

- Mr. Torsten A. Andersen, Deputy Director General, Danish Business Authority
- Ms. Anne Kauhanen-Simanainen, Ministerial Advisor, Ministry of Finance, Finland
- Ms. Manuela Siano, Service for EU and International Matters, Italian Data Protection Authority
- Mr. Daisuke Nagasaki, International Affairs Office, Commerce and Information Policy Bureau, Ministry of Economy, Trade and Industry, Japan
- Ms. Jenni Nordborg, Head of Health, Swedish Governmental Agency for Innovation Systems
- Ms. Katarina de Brisis, Deputy Director General, Ministry of Local Government and Modernisation, Norway
- Ms. Elettra Ronchi, Senior Policy Analyst, OECD Directorate for Science, Technology and Innovation; Digital Economy Policy

**Followed by discussion [20 min]**

**Closing Remarks**

- Ms. Katarina de Brisis, Deputy Director General, Ministry of Local Government and Modernisation, Norway