A data trust for industry data sharing[†‡]

A data trust is a knowledge commons model for data sharing which can empower and create value for its members. The access provided by the trust does not depend upon the opaque and internal decisions of a firm operating for profit in a competitive market. Under a data trust arrangement, participants that share data create and control the resource, governing access according to mutually agreed principles. The data trust creates a new resource for under-resourced market participants—it is an economic resource that is shared for mutual benefit.

Building a data trust for industry data sharing is a collective action challenge. It requires encouraging a diverse range of actors to share commercially sensitive information about their assets and operations. It also requires the sharing of information in a form that is standardized and comparable so that the analysis of that information is accurate and useful.

A successful data trust must be mutually beneficial for all participants. Enabling collaboration at this scale requires a carefully calibrated system of incentives and investment in providing value back to participants. If carefully designed to meet the needs of participants, a data trust is an institutional model with the potential to facilitate sustainable industry data sharing and to improve opportunities for analytics-led innovation and growth.

How do you design a sustainable industry data sharing trust that will provide members with opportunities for analytics-led innovation and growth?

Knowledge commons research shows that governance design is critical to the sustainability of a commons. This is because a commons requires cooperation by a variety of actors, many of whom will be competitors (Frischmann, Madison, and Strandburg 2014, p.477). Sustainable cooperation among competitors requires trust, a sense of community, and reciprocity (Hess and Ostrom 2007, pp.43-44). The data trust must be designed and governed in a way that engenders these three properties among its members.

In her analysis of voluntary communal management of common pool resources, Elinor Ostrom identified that sustainable commons share particular design features that provide certainty for participants about the use and governance of the resource (Ostrom 1991). Certainty about how a resource is used and governed engenders trust and supports sustainable collective action (Cox, Arnold, and Villamayor Tomás 2010, p. 2). In the context of data sharing, this principle is illustrated by the results of a global 2017 survey in which '25% of respondents from health care industries said they are likely to share data with competitors, compared with 19% of respondents from other industries' (Ransbotham and Kiron 2017, p.12). The authors of the

---

[†] This document is an extract from the report *Building a Trusted Framework for Coordinating OA Monograph Usage Data.* Primary authors of this section of the report are: Nic Suzor and Joanne Gray. The authorship team also included Cameron Neylon, Lucy Montgomery, Alkim Ozaygen. Kevin Hawkins (University of North Texas), Charles Watkinson (University of Michigan) and Brian O'Leary (BISG) contributed edits and improvements.

[‡] This version was made available to the community for comment and feedback in November 2018. It is licensed to the general public under a Creative Commons Attribution 4.0 International license.

survey concluded that in the health care sector, established patient privacy and data regulations encourage industry data sharing. Structured governance is important for engendering trust and supporting resource sharing.

According to Ostrom's analysis, a sustainable commons must have clear boundaries. This requires defining the resource, who may use it, and who is excluded from using it. Defining who is included and who is excluded from the commons provides participants with assurance that the benefits derived from their contribution will not be unfairly enjoyed by those that have not made a similar contribution (Ostrom 1991, p. 91). Once boundaries are defined, establishing a sustainable commons requires clear rules for using and contributing to the commons that are tailored to fit the specific resource. Rule-making and rule-modification should be devised with the participation of users of the commons. The use of the commons should be actively monitored in order to ensure compliance with commons rules and boundaries, and there should be low-cost dispute resolution mechanisms and graduated sanctions applied in response to boundary or rule violations (Ostrom 1991, Cox *et al.* 2010, p. 10). A sustainable commons also requires that the self-governance of the commons is respected by external authorities. Finally, for complex common resources, governance of the commons should occur on multiple levels— for example local, regional and national—in order to successfully manage the different and interconnected components of the resource system (Ostrom 1991, p. 101).

Ultimately, both the governance and technological layers of the data trust must be tailored to the specific requirements of the data resource and its members.

European union guidance on business-to-business data sharing

In 2018, the European Union released a report providing initial guidance on private sector data sharing. This guidance identifies important issues that require consideration when designing an organization that shares industry data and has informed the principles and contractual terms proposed in this document.

The report provided the following five principles for business-to-business data sharing arrangements (European Commission 2018):

> a) *Transparency*: The relevant contractual agreements should identify in a transparent and understandable manner (i) the persons or entities that will have access to the data that the product or service generates, the type of such data, and at which level of detail; and (ii) the purposes for using such data.

> b) *Shared value creation:* The relevant contractual agreements should recognize that, where data is generated as a by-product of using a product or service, several parties have contributed to creating the data.

> c) *Respect for each other's commercial interests:* The relevant contractual agreements should address the need to protect both the commercial interests and secrets of data holders and data users.

> d) *Ensure undistorted competition:* The relevant contractual agreements should address the need to ensure undistorted competition when exchanging commercially sensitive data.

> e) *Minimized data lock-in:* Companies offering a product or service that generates data as a by-product should allow and enable data portability as much as possible. They

should also consider, where possible and in line with the characteristics of the market they operate on, offering the same product or service without or with only limited data transfers alongside products or services that include such data transfers.

The report also specified issues for consideration when preparing or negotiating data sharing contracts (European Commission 2018):

1. What data shall be made available?
2. Who can access and (re-)use the data in question?
3. What can the (re-)user do with the data?
4. Define the technical means for the data access and/or exchange
5. What data do I need to protect and how do I protect it?
6. Include rules on liability provisions for supply of erroneous data, disruptions in the data transmission, low quality interpretative work, if shared with datasets, or for destruction/loss or alteration of data (if it is unlawful or accidental) that may potentially cause damages.
7. Define rights of both parties to perform audits on the respect of the mutual obligations.
8. What is the intended duration of the contract? What rights to terminate the contract? What notice to be given to your partners?
9. Agree on applicable law and dispute settlement mechanisms.

Industry data sharing

Current industry data sharing practices suggest data sharing is emerging as an important driver of innovation. A 2017 data and analytics study by MIT found that highly innovative organizations share data internally and externally at higher rates than other organizations (Ransbotham and Kiron 2017, p. 10). In 2015, German car companies BMW, Daimler, and Volkswagen collectively purchased a business that owns a digital mapping technology that facilitates traffic jam and road condition video sharing between drivers (Ransbotham and Kiron 2017, p. 11). Through the jointly owned business, the car companies share data collected from sensors in their vehicles and are developing services that deliver real-time alerts to drivers. As Sam Ransbotham and David Kiron (2017, p.11) note, the customer base and data that each company brought to the partnership has allowed the companies to collectively create a more valuable service for their customers than if they had each attempted the project individually.

Industry data sharing occurs commonly in the field of medical research. Within the pharmaceutical industry, for example, clinical trial data sharing is recognized as important for supporting scientific discourse, innovation and discoveries (Francer, 2015). In this field, a common model for data sharing is via an independent intermediary. For example, Johnson & Johnson works with the the Yale School of Medicine's Open Data Access (YODA) project (Yale University Open Data Access Project 2018). YODA is responsible for deciding when to share data that is requested by third parties. In this industry setting, organizations seeking to share data must carefully navigate issues relating to patient privacy and commercial incentives. The high rate of data sharing in this sector highlights that data obtained through large financial investments, with a potentially high commercial value and subject to significant regulatory requirements, can be shared successfully.