

## Data intermediaries for health

Lack of available, diverse, well-curated data has long been a rate-limiting step in developing innovative medical therapies. So-called 'data intermediaries' are being hailed as the key to unlocking data across industries and sectors. These data stewardship structures have the potential to mitigate some well-known risks and barriers for data sharing, and whilst their flexibility offers great potential value to sectors like health, it also means there are many outstanding questions on how they should work ethically and legally.

### What are data intermediaries?

While there is no standard definition of a data intermediary, they are best understood as a set of specified relationships between individuals or groups of people and data. The kind of relationship and the law that governs it will be specified by a core document (e.g. if it is trust law then a trust deed). Intermediaries are overseen by specified individual(s) who hold duties to prevent breaches and ensure obligations are carried out under the mutually agreed terms of those relationships. Some intermediaries are designed to give individuals maximum control over data (e.g. Personal Information Management Systems), whereas others place the responsibility to make decisions about data use with those in charge of the intermediary.

The Centre for Data Ethics and Innovation (CDEI) identifies seven types of data intermediaries (see below). Of these, data trusts are generating particular interest and will be explored in a future blog.

The closest the health sector has to intermediaries are trusted research environments (TREs). Not to be confused with data trusts, these are also a relatively new approach to sharing data that has come to prominence following their use in the COVID-19 pandemic and recommendations in the [Goldacre Review](#) for their expansion.

Trusted research environments fulfil many, but not all, of the common criteria for an intermediary. They grant use rights and place limitations on those who are permitted to access its data library. For example, researchers may be permitted to run tests on the data but not alter, manipulate, or download the data. They also have enhanced security features and are overseen by those placed in a position of trust to ensure use rights are not breached.

However, intermediaries are associated with independence from the activities of parent institutions. It is also not clear how laws governing intermediaries will interpret the concept of independence and therefore whether existing TREs will meet its requirements.

**Types of data intermediaries identified by the CDEI<sup>1</sup>**

| Data intermediary                              | Intended purpose   | Existing example  |
|--|--|---|
| Data exchange                                  | Operate as online data platforms where datasets can be advertised and accessed - commercially or on a not-for-profit basis                             | AWS Data Exchange   |
| Data cooperatives                              | Enable shared data spaces controlled by data subjects  | (Likely to be a private intermediary)   |
| PIMS (Personal Information Management Systems) | Seek to give data subjects more control over their personal data   | Digi.me or SOLID  |
| Trusted third party                            | Provide assurance to those looking to access confidential datasets that the data is fit-for-purpose (e.g., in terms of quality or ethical standards)   | Open Banking Implementation Entity (OBIE)   |
| Data custodian                                 | Enable privacy-protecting analysis or attribute checks of confidential data, for example, via the application of Privacy Enhancing Technologies (PETs) | OpenSAFELY, Genomics England, Pensions Dashboard or National Institute of Standards and Technology (NIST) |
| Industrial data platform                       | Provide shared infrastructure to facilitate secure data sharing and analysis between companies   | Advanced Product Concept Analysis Environment (APROCONe), MK Data Hub                                     |
| Data trusts                                    | Provide fiduciary data stewardship on behalf of data subjects  | (Likely to be a private intermediary)   |

## How can several individuals hold interests in the same data?

Intermediaries are founded on the basis that interests in their data libraries, and rights to their use, can be held simultaneously by multiple parties. This can appear complex but it is possible to break it down in simple terms.

There is a common misunderstanding that data is 'property'. In law, it is not necessarily the data itself that amounts to property, but the benefit of legally recognised rights or interests in relation to it that can be considered legal property.

One way to understand this is to think of ownership in terms of a bundle of sticks. The bundle represents the 'property' e.g. the data, and each stick on its own represents an interest in that property. To hold all the sticks will mean you hold exclusive ownership over that property and can use or dispose of it as you choose. Yet exclusive ownership is rare because all property has lawful limitations in its use. Splitting up its interests creates a market for parties to exercise or transfer interests, creating value in resources such as data.

It is because several groups or individuals can hold concurrent interests, reflecting its multiple possible uses that makes the data valuable. These multiple parties can include those the data is about, researchers, governments, or other parties interested in data for innovation.

## How intermediaries could support data sharing in healthcare<sup>2</sup>

| Challenges  | Solution   |
|---|--|
| Sharing data may be giving valuable information to competitors  | Overcome IP risks via fiduciaries controlling what information is shared. APROCONE provides an example from the Aviation industry  |
| Lack of knowledge on value and availability of data   | Intermediaries could provide health institutions with information availability of relevant existing data sets, on the value of their own data sets, and on risks, and their mitigation, of sharing certain types of data in certain contexts |
| Lack of standards across datasets could lead to commercial, reputational, ethical and legal risks or damage | Intermediaries could set standards for data sets shared with and by them and consequently begin trailblazing the standard-setting needed for healthcare data. Such standards are needed to mitigate and prevent harm                         |
| Missed opportunities to use data in the public interest   | Data siloing can hinder healthcare innovation. It can also lead to regulatory and legal risk where an opportunity for a sector to strengthen its processes and safety standards is missed  |

## Data trustees and risks

Data intermediaries promise benefits such as the absorption of risks associated with sharing and collaborating with others on data sets and standard-setting; they could also become experts on what data sets exist and which are needed but missing from certain sectors. However, it is unclear how those running intermediaries will be insured against the vast reputational and legal risks they would subsume as data controllers for various institutions' data sets. Depending on the context, these individuals may have different titles, such as trustees or board members. Here we use trustees broadly to mean those in a legally bound position of trust to oversee the agreement.

Central to the success of data intermediaries, particularly in healthcare, will be their trustworthiness. This places considerable onus on the individuals and organisations taking on such roles and their potential for regulatory and reputational risk. This also raises questions of how such individuals may be remunerated for absorbing such significant risk where they need to remain independent and whether insurance companies would be willing to provide professional indemnity insurance for them. Realising the benefits intermediaries can offer could depend on the answers to these questions.

## Priorities for policy

Policy questions still to be addressed to leverage value from data intermediaries for health include:

- To understand how an intermediary's 'trustees' are to absorb the vast reputational and legal risks associated with such broad scale data stewardship and how they could be insured
- To look at when and what types of data intermediary might be useful in a healthcare context (at both an inter-organisational level like APROCONE or an organisation to individual level like Genomics England's TRE)

Other areas we will be investigating further are:

- Data trusts and the question of data ownership
- The pressing question of ethical and legal incentives for individuals to share genomic data
- The ethical and legal boundaries on the broader types of data that intermediaries could collect from individuals to foster further innovation

## References

1. Adapted from: Centre for Data Ethics and Innovation. Unlocking the Value of Data: Exploring the Role of Data Intermediaries. 2021
2. Adapted from Frontier Economics and the Department of Communication, Media and Sports. Increasing Access to Data Across the Economy. And: Centre for Data Ethics and Innovation. Unlocking the Value of Data: Exploring the Role of Data Intermediaries. 2021

Author: Elizabeth Redrup

Published: July 2022

Contact: [intelligence@phgfoundation.org](mailto:intelligence@phgfoundation.org)

PHG Foundation is a health policy think tank with a special focus on how genomics and other emerging health technologies can provide more effective, personalised healthcare