

# Designing sustainable data institutions

---



# Contents

<b>Executive summary</b>	<b>5</b>
<b>Introduction</b>	<b>7</b>
<b>Example data institutions</b>	<b>8</b>
<b>What do we mean by sustainability?</b>	<b>11</b>
<b>A framework for sustainability</b>	<b>13</b>
Actors in the ecosystem	14
Data contributors	14
Data users	15
Case study: HiLo	15
Decision makers	16
Funders	16
Revenue models versus business models	16
Applying the Business Model Canvas to data institutions	18
Funding versus financing	19
Cost structure	20
Case study: ROR	21
Sources of revenue	22
Earned revenue	22
Grants, donations and other non-earned revenue sources	24
<b>Common patterns in data institution sustainability</b>	<b>26</b>
Data institutions use mixed revenue streams	26
Five approaches to financing data institutions	26
Case study: OpenCorporates	27
Revenue models evolve over time	28
Institutions plan for future investments	29
Data institutions develop contingency plans	30
Good governance and community support are critical to sustainability	31
<b>Issues and tensions with achieving sustainability</b>	<b>32</b>
Difficulty in attracting early financing	32
Aligning revenue models with institutional goals	33
The ever-evolving relationship between a data institution and its community	36
<b>Conclusion and recommendations</b>	<b>39</b>
Areas for further research	39
Suggestions for those scoping, designing and running data institutions	40
<b>Methodology</b>	<b>42</b>
Research questions	42
Criteria to select institutions	42
Desk research	43
Interviews	43

## About

This report has been researched and produced by the Open Data Institute, and was published in April 2020. The lead authors are Leigh Dodds, Diána Szász, Jared Robert Keller, Ben Snaith and Sonia Duarte, with contributions from Jack Hardinges and Jeni Tennison.

If you would like to send us feedback, please get in contact with us at [policy@theodi.org](mailto:policy@theodi.org).

# Executive summary

In spring 2019, the Open Data Institute (ODI) published pilot reports concluding that data trusts – new organisational structures providing independent stewardship of data – could be useful in increasing access to data while retaining trust. Data trusts are one example of a broader range of data institutions.

Data institutions are organisations whose purpose involves stewarding data on behalf of others, often towards public, educational or charitable aims. For these institutions to achieve this purpose, they need to be sustainable. Sustainability will ensure that institutions can steward data over the long term, in ways that help to increase the value that can be created from that data, while minimising potential harms.

This report is the result of an initial investigation into the sustainability of data institutions. We have explored the business and revenue models for a small number of existing data institutions. Some are still in the early stages of their lifecycle, while others have been in operation for 20 or more years.

We propose a framework for thinking about the sustainability of data institutions that involves consideration of three different elements:

- The role that the data institution plays in its data ecosystem, which directly impacts the type and source of its revenue; and the underlying costs that the institution has to cover.
- The stage that the data institution is at in its lifecycle, which informs the type of revenue it needs to acquire, and how those revenue sources may evolve over time as the institution and its ecosystem matures.
- The business model of the data institution, which informs the choice of revenue model that best aligns with its purpose.

We draw on existing work, such as Strategyzer's Business Model Canvas, developed by Alexander Osterwalder, to help explore the differences between business and revenue models.

In exploring the potential revenue streams available to data institutions we differentiate between earned revenue, which is generated through subscription and usage fees attached to delivery of value and services provided by the organisation, and non-earned revenue including grants, donations and in-kind contributions. The latter tend to provide direct support for the institution's core mission which provides more flexibility around how that revenue may be used.

Our research has identified some broad patterns around how data institutions are approaching sustainability.

At an early stage in their lifecycle, data institutions need financing to support the process of scoping and launching the institution. This is typically provided through grants and other non-earned revenue sources.

Later in their lifecycle they need funding to cover costs of service provision while generating sufficient revenue to build reserves to ensure long-term sustainability and invest in necessary infrastructure. At this stage institutions use a mixture of earned

and non-earned revenue. For many institutions this mix evolves over time based on the changing needs of the community they support.

We have identified several issues that data institutions face in achieving sustainability. The difficulty in attracting the financing necessary to cover setup costs is a challenge. Data institutions are using a mixture of grant funding, incubation and hosting by existing organisations to develop to a point where other revenue streams become feasible.

Aligning revenue models with their institutional goals is also difficult. Some goals, for example, to provide strong governance and protection of data, are at odds with revenue streams based on usage of data. Data institutions, funders and other stakeholders need to understand the potential range of revenue sources available. We have mapped out an initial list of these in this paper – see [Sources of revenue](#) section.

Finally, the challenge of responding to the evolving needs of their community means that institutions need to continue to adapt, requiring regular re-evaluation of revenue streams, as well as the ability to support ongoing investment rather than a strict cost-recovery model.

This project has run in parallel to a related project that has focused on designing *trustworthy* data institutions<sup>1</sup>. Both projects have highlighted that trust and sustainability are deeply connected, requiring strong governance and ongoing community engagement.

As exploratory research, this project has also helped to identify areas for further work and analysis that we will continue to explore in the year ahead. For example, we plan to conduct a more comprehensive review of existing institutions to understand how data institutions operating in different sectors, with different goals, have developed their revenue models.

We also believe there is a role for additional guidance and tools to help new and existing data institutions to plan and evolve their revenue models, as well as recommendations for how funders can better support the creation of sustainable data institutions. We will develop and test the latter by directly engaging with and supporting a range of new and existing institutions.

---

<sup>1</sup> Open Data Institute (2020), 'Trustworthy data institutions', <https://theodi.org/project/trustworthy-data-institutions>

# Introduction

Data stewardship involves collecting, maintaining and sharing data, and, in particular, determining who has access to it, for what purpose and to whose benefit.

How data is stewarded is important, as it affects what it can be used for and how it may bring benefit or cause harm. Data institutions are organisations whose purpose involves stewarding data on behalf of others, often towards public, educational or charitable aims. At the ODI, we have been exploring the role of data institutions in increasing access to data.

Data trusts are one type of data institution. They provide independent, fiduciary stewardship of data.<sup>2</sup> With data trusts, one party authorises another to make decisions about data on their behalf, for the benefit of a wider group of stakeholders. The independent person, group or entity stewarding the data takes on a fiduciary duty, which is considered the highest level of obligation that one party can owe to another.

Some data institutions will follow a similar pattern, where an organisation, or group of organisations, entrusts them to share data with others. Other data institutions will play different roles, such as combining or linking data, and providing benchmarks and other insights to the contributing organisations. Regardless of the role they play, designing data institutions to be sustainable in the long term is vital, both to maximise their impact and to provide assurance to the data contributors and users who invest time and effort in dealing with them.

Sustainability can be defined in a number of ways, depending on where and how the term is being used. Broadly, we have focused on sustainability as the ability of an institution to continue operations so that it can deliver on its purpose. More specifically, we have looked at financial sustainability, that is the ability to maintain financial capacity over time.<sup>3</sup>

If data institutions are delivering societal value through their activities, it is imperative that they continue to do so for as long as possible. In this report, we set out the range of revenue models that can be adopted by sustainable data institutions. The revenue model adopted by a data institution reflects the choice it has made in order to balance benefits, costs and risks. For some data institutions, a commercial or profit-driven business model might interfere with the types of value they are trying to deliver. For others, a commercial focus might enable them to invest more in, and maximise, the value generated from the data they steward.

Sustainability is often highlighted as a common challenge across the digital commons, such as open source projects and open standards. While funding of physical infrastructure is well understood but challenging, building a sustainable, well-maintained data infrastructure is something that, in many cases, we are still learning how to do. In her report 'Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure',<sup>4</sup> funded by the Ford Foundation, Nadia Eghbal describes how the digital infrastructure that so many of us rely on every day is often built and maintained by communities of volunteers. This, as Michael Brannen from the Ford

---

<sup>2</sup> Open Data Institute (2020), 'Data trusts in 2020', <https://theodi.org/article/data-trusts-in-2020/>

<sup>3</sup> Bowman W (2011), 'Financial Capacity and Sustainability of Ordinary Nonprofits'.

<sup>4</sup> Ford Foundation (2016), 'Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure', <https://www.fordfoundation.org/work/learning/research-reports/roads-and-bridges-the-unseen-labor-behind-our-digital-infrastructure/>

Foundation explains, leads to “thousands of critical software projects being at risk of being undermaintained, and of collapsing as a result”.<sup>5</sup> Nadia Eghbal found that “while lots of companies and organisations rely on digital infrastructure to build their products and services, few of those projects are well funded or supported”. Nadia notes the public good nature of digital infrastructure, which results in a lack of incentive on behalf of individual companies or organisations to offer financial support.

To explore this topic, the ODI carried out desk research into how existing institutions have become sustainable. This was followed by interviews with representatives from a number of data institutions:

- **Higher Education Statistics Agency (HESA)**<sup>6</sup>: an official body which collects, analyses and publishes data about higher education in the United Kingdom (UK). HESA collaborates with higher education providers, such as universities, to collect and curate higher education data sources. Its products are used by researchers and policymakers for transparency, retaining public trust and decision making.
- **Research Organization Registry (ROR)**<sup>7</sup>: a community-led project working to produce a unique, open, usable and sustainable identifier for every research organisation in the world.
- **HiLo Maritime Risk Management**<sup>8</sup>: a not-for-profit joint industry initiative providing analysis of shipping data to make the industry safer. Shipping companies share safety-related data from vessels and HiLo runs it through a risk algorithm, and shares insights with the companies.
- **OpenCorporates**<sup>9</sup>: the world’s largest open database of information about companies. All of the data on OpenCorporates comes from primary public sources, and is used by individuals, journalists, non-governmental organisations (NGOs) and companies.
- **MusicBrainz**<sup>10</sup>: a project to create a collaborative database about artists, songs and albums. Any user can contribute and release the music metadata under open licences.<sup>11</sup>

These data institutions are comparatively successful and the lessons from their individual journeys towards sustainability are useful. We note, however, that the research presented here is limited as we have not sought to understand what makes a data institution unsustainable.

This report presents the results of our initial exploratory research. We hope this research will help develop further guidance to support the effective adoption of suitable business models by new and existing data institutions, as well as to support other actors (including governments, regulators and funders) that aim to fund and support them.

---

<sup>5</sup> Ford Foundation (2017), ‘Every day, we rely on digital infrastructure built by volunteers. What happens when it fails?’, <https://www.fordfoundation.org/ideas/equals-change-blog/posts/every-day-we-rely-on-digital-infrastructure-built-by-volunteers-what-happens-when-it-fails/>

<sup>6</sup> Higher Education Statistics Agency (n.d.), <https://www.hesa.ac.uk/>

<sup>7</sup> Research Organization Registry (n.d.), <https://ror.org/>

<sup>8</sup> HiLo Maritime Risk Management (n.d.), <https://hilomrm.com/>

<sup>9</sup> OpenCorporates (n.d.), <https://opencorporates.com/>

<sup>10</sup> MusicBrainz (n.d.), <https://musicbrainz.org/>

<sup>11</sup> MusicBrainz (n.d.), ‘About: Data License’, [https://musicbrainz.org/doc/About/Data\\_License](https://musicbrainz.org/doc/About/Data_License)



# Example data institutions

The following table provides some examples of existing data institutions across a range of sectors, at different stages of maturity and with a range of legal forms. We discuss some of these in more detail later in this report.

Data institution	Legal status	Established	Current revenue	Previous revenue	Sector
<b>HESA</b>	Not-for-profit, private limited company owned by its members, operating under a statutory footing	1993	<ul style="list-style-type: none"> <li>• Mandatory subscriptions from higher education providers (~£10m annual)</li> <li>• Service revenue share with a partner (less than ~£0.5m annual)</li> <li>• Grant income (around ~£0.5m annual)</li> </ul>	<ul style="list-style-type: none"> <li>• Subscriptions from higher education providers</li> <li>• Service revenue</li> </ul>	Higher education
<b>MusicBrainz</b>	Operated by the MetaBrainz Foundation, a 501(c)(3) tax-exempt non-profit corporation	2000	<ul style="list-style-type: none"> <li>• Donations from individuals (1% of income)</li> <li>• Donations from Google (\$30k annual)</li> <li>• Sponsorships</li> <li>• Monthly support payments for commercial use of data</li> </ul>	<ul style="list-style-type: none"> <li>• Privately maintained. Licence was Creative Commons non-commercial licence with companies paying for use</li> </ul>	Music
<b>Crossref</b>	Not-for-profit membership organisation, a 501(c)(3) tax-exempt non-profit corporation	2000	<ul style="list-style-type: none"> <li>• Membership fees</li> <li>• Subscription service fees</li> <li>• Content (data) registration fees</li> <li>• Service fees</li> </ul>	<ul style="list-style-type: none"> <li>• Membership fees</li> <li>• Subscription service fees</li> <li>• Content (data) registration fees</li> <li>• Service fees</li> </ul>	Research
<b>UK Biobank</b>	Registered charity	2004	<ul style="list-style-type: none"> <li>• Grants and core funding (total ~£244m committed since establishment)</li> <li>• Application and administration fees</li> </ul>	<ul style="list-style-type: none"> <li>• Grants and core funding</li> </ul>	Health (research)

<b>OpenCorporates</b>	Social enterprise	2010	<ul style="list-style-type: none"> <li>• Optional, charged bulk data and Application Programming Interface (API) access (~75%)</li> <li>• Grant funding (~25%)</li> </ul>	<ul style="list-style-type: none"> <li>• Grant funding for individual projects</li> </ul>	Business information
<b>OpenActive</b>	None	2015	<ul style="list-style-type: none"> <li>• Government grant</li> <li>• In-kind support</li> </ul>		Physical activity
<b>HiLo</b>	Private company	2018	<ul style="list-style-type: none"> <li>• Subscription fees collected from data contributors</li> </ul>	<ul style="list-style-type: none"> <li>• Incubated by Shell (2018)</li> </ul>	Maritime
<b>ROR</b>	Project hosted by a group of organisations	2019	<ul style="list-style-type: none"> <li>• In-kind donations</li> <li>• Fundraising campaign ongoing (community contributions and donations)</li> </ul>	<ul style="list-style-type: none"> <li>• In-kind donations (staff time, financial resources, seed data)</li> </ul>	Research

**Table 1:** Examples of existing data institutions across a range of sectors, at different stages of maturity and with a range of legal forms.

# What do we mean by sustainability?

There are various definitions of ‘sustainability’. In general, sustainability means being able to exist over the long term.

A sustainable business is one that has “a positive effect, on the global or local environment, community, society, or economy”.<sup>12</sup> Organisational sustainability can also be broken down into different elements, including strategic, product, personnel and financial sustainability.<sup>13</sup> Financial sustainability has been defined, in short, as “keep the business going”.<sup>14</sup>

The challenges of sustainable funding for shared resources and infrastructure are not new. For example, Olson (1965)<sup>15</sup> describes the challenges of finding sustainable business models for common resources due to the issue of freeloading, whereby somebody gains access to, and benefits from, infrastructure, but does not contribute to it.

Our initial research draws significantly on a number of papers about maintaining scholarly infrastructures. The infrastructure underpinning online academic research has been developing over the last 20 years, through organisations such as Crossref<sup>16</sup> and ORCID,<sup>17</sup> and more recently through ROR. The process of maintaining this shared infrastructure is analogous to the idea of a data institution; conclusions from the academic sector are relevant to our research here.

The article ‘Principles for Open Scholarly Infrastructures’<sup>18</sup> sets out five methods for maintaining sustainability through ensuring capacity and resources:

1. Time-limited funds are used only for time-limited activities – day-to-day operations should be supported by day-to-day sustainable revenue sources.
2. Goal to generate surplus – it is not enough to merely survive, an organisation has to be able to adapt and change. To weather economic, social and technological volatility, they need financial resources beyond their immediate operating costs.
3. Goal to create a contingency fund to support operations for 12 months – a high priority should be generating a contingency fund that can support a complete, orderly wind down (12 months in most cases). This fund should be separate from those allocated to covering operating risk and investment in development.
4. Mission-consistent revenue generation – potential revenue sources should be considered for consistency with the organisational mission and not run counter to the aims of the organisation.

---

<sup>12</sup> Wikipedia (2020), ‘Sustainable Business’, [https://en.wikipedia.org/wiki/Sustainable\\_business](https://en.wikipedia.org/wiki/Sustainable_business)

<sup>13</sup> Free Management Library (n.d.), ‘Organisational Sustainability’, <https://managementhelp.org/organizational-sustainability/index.htm>

<sup>14</sup> Colbert B, Kurucz E (2007), ‘Three conceptions of triple bottom line business sustainability and the role for HRM’, [https://www.academia.edu/5316345/Three\\_Conceptions\\_of\\_Triple\\_Bottom\\_Line\\_Business\\_Sustainability\\_and\\_the\\_Role\\_for\\_HRM](https://www.academia.edu/5316345/Three_Conceptions_of_Triple_Bottom_Line_Business_Sustainability_and_the_Role_for_HRM)

<sup>15</sup> Olson, M (1965), ‘The Logic of Collective Action’, <http://www.sfu.ca/~poitras/collective-action.PDF>

<sup>16</sup> Crossref (n.d.), <https://www.crossref.org/>

<sup>17</sup> ORCID (n.d.), <https://orcid.org/>

<sup>18</sup> Bilder G, Lin J, Neylon C (2015), ‘Principles for Open Scholarly Infrastructure’, <https://wiki.lib.sun.ac.za/images/f/f6/2015-principle-for-open-scholarly-communication-infrastructure.pdf>

5. Revenue based on services, not data – data related to the running of the research enterprise should be community property. Appropriate revenue sources might include value-added services, consulting, API Service Level Agreements or membership fees.

The article ‘Sustaining Scholarly Infrastructures through Collective Action: The Lessons that Olson can teach us’<sup>19</sup> states that “focusing purely on financial sustainability in the absence of considering governance principles and community is the wrong approach”. Models for sustainability will “need to change over the lifecycle of an infrastructure with the growth (or decline) of the community”.

This article also highlights that the nature of shared digital infrastructure leads to a further issue: institutions and their communities can grow large due to the limited amount of friction generated from additional contributors or users. This suggests that: “Digital resources are not natively excludable; a technical barrier has to be put in place.”<sup>20</sup>

Therefore, when we are thinking about the sustainability of data institutions, we are looking at three things:

1. The existence of a sustainable revenue model, underpinned by a mix of revenue streams, to ensure the ongoing stewardship of data.
2. The continued demand for the data from data users and those that ultimately use the products and services, analyses and insights they create – “infrastructures need to be seen as both sustaining and being sustained by the communities that they serve.”<sup>21</sup>
3. Sufficient governance to ensure that the aims of the organisation can be delivered in the long term.

---

<sup>19</sup> Neylon C (2017), ‘Sustaining Scholarly Infrastructures through Collective Action: The lessons that Olson can teach us’, <https://kula.uvic.ca/articles/10.5334/kula.7/>

<sup>20</sup> Ibid.

<sup>21</sup> Ibid.

# A framework for sustainability

In the previous section we noted that the sustainability of a data institution is linked to the ecosystem that surrounds it and the community that supports it. An ecosystem diagram can be a helpful way of understanding the role that aspects such as financing, funding and value creation play in the ultimate sustainability of a data institution. It can also be useful in clarifying the difference between a data institution's business model and its revenue model.

To that end, [Figure 1](#) below centres on a hypothetical data institution and illustrates common sources of revenue available to it, as well as the types of organisations from which those sources of revenue often flow. It also illustrates the way that data tends to flow within such systems and highlights the flow of insights and other information that informs decision making within those systems. It does not illustrate the more intangible types of value that the actors shown in the diagram, and the wider community, might receive due to the presence of the data institution.

In producing the diagram, we utilised an ODI tool, Data Ecosystem Mapping.<sup>22</sup> The approach draws on ideas from rich picturing,<sup>23</sup> systems thinking and value network analysis to generate a visual map of how data is being accessed, used and shared by a variety of organisations within an ecosystem.<sup>24</sup>

Though data ecosystem mapping can be used to illustrate the ecosystem that exists around a single organisation or institution, [Figure 1](#) is not intended to represent a real-world data institution or describe a complete ecosystem; rather, it is meant to showcase the common actors that appear in such an ecosystem and demonstrate the various *types* of revenue or value that are available to a typical data institution.

---

<sup>22</sup> The Open Data Institute (2018), 'Mapping data ecosystems: methodology', <https://theodi.org/article/mapping-data-ecosystems/>

<sup>23</sup> Wikipedia (2019), 'Rich picture', [https://en.wikipedia.org/wiki/Rich\\_picture](https://en.wikipedia.org/wiki/Rich_picture)

<sup>24</sup> Wikipedia (2020), 'Value network analysis', [https://en.wikipedia.org/wiki/Value\\_network\\_analysis](https://en.wikipedia.org/wiki/Value_network_analysis)

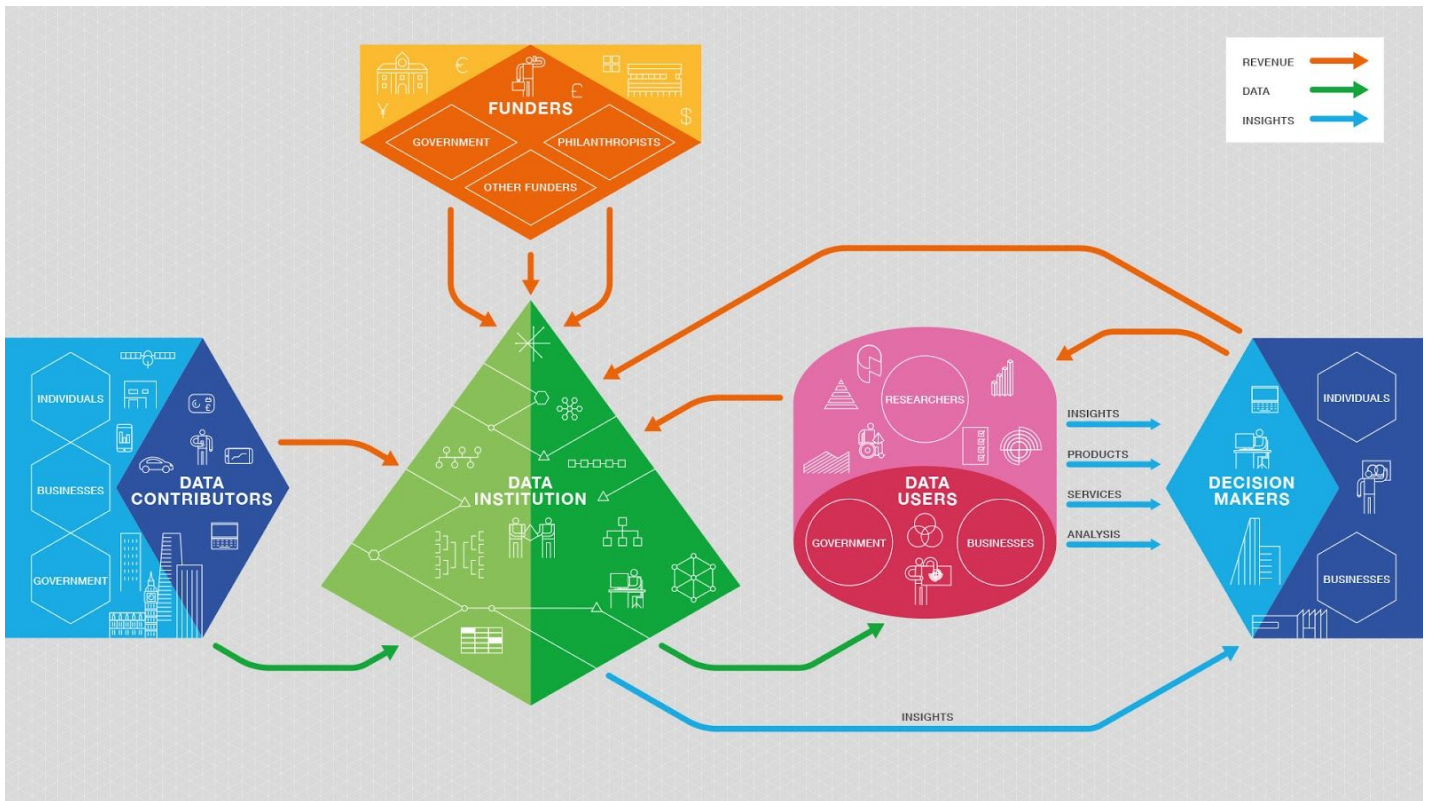


Figure 1: [Common actors and flows of value and data related to data institutions](#)

## Actors in the ecosystem

The ecosystem diagram in [Figure 1](#) is made up of data contributors, the data institution, data users, decision makers and funders.

It is important to note that none of these categories are mutually exclusive. An organisation may occupy multiple roles within the ecosystem that surrounds a data institution and, in doing so, may contribute to the sustainability of a data institution in multiple ways. [Microsoft and AOL](#), for instance, have in the past simultaneously funded the open source map, OpenStreetMap; contributed datasets to improve its accuracy; and used the platform in some of its commercial services.<sup>25</sup>

Throughout this report, we also refer to the actors in the ecosystem that surround a data institution as a ‘community’.

### Data contributors

We define ‘data contributors’ as people or organisations who contribute to, or help curate, a dataset. They may do so explicitly, using tools and frameworks provided by a data steward, or they may do so as a by-product of using a service.

In the context of HiLo, for instance, the data contributors are the shipping companies that have joined the venture and agree to share accident and safety data about their operations with HiLo each month. Whereas for the MusicBrainz project, the contributors are the members of the volunteer community who commit to adding

<sup>25</sup> OpenStreetMap (2010), ‘OpenStreetMap Gets Noticed by Microsoft, AOL’, <https://blog.openstreetmap.org/2010/12/08/openstreetmap-gets-noticed-by-microsoft-aol/>

information on their favourite artists and their works. In contrast, cancer clinics and other organisations that use Flatiron Health’s software to manage their cases and workflows permit Flatiron, a healthcare technology and services company in the United States, to process data generated by their use of those services to share it for vital research.

## Data users

To the right of the ‘data institution’ in [Figure 1](#) are ‘data users’ – people or organisations accessing data stewarded by the data institution and using it to create products and services, analyses and insights, or stories and visualisations.

For instance, the ‘statutory customers’ who use data published by HESA to perform their statutory or public functions would be classed as data users in Figure 1.<sup>26</sup> Similarly, members of the scholarly research community who use the research organisation identifiers stewarded by ROR to discover and track research outputs would be classed as data users.

## Case study: HiLo

HiLo Maritime Risk Management became an independent not-for-profit company in 2018 as a joint venture between 10 leading maritime companies. This followed a period of incubation within Shell Energy.

The main aim of HiLo is to collect data from ships, run it through a predictive risk model and offer insights back to contributors, who are also decision makers within HiLo’s data ecosystem. The data includes all internal safety data about near misses, unsafe acts or faulty equipment – “things where something has gone wrong but hasn’t led to a catastrophic incident”.

Around 55 companies joined the membership and committed to share data that is already collected, to gain collective insight about safety. The data is shared from these contributors directly to HiLo once a month – this is usually a manual process handled via the HiLo portal, but some companies opt to share via an API.<sup>27</sup>

Financially, HiLo currently funds its operations completely from subscription fees for its insight service, but hopes to diversify its revenue streams to include a more balanced mixture of subscriptions and revenue generated from new products and services.

---

<sup>26</sup> HESA (n.d.), ‘What we do: Statutory customers’, <https://www.hesa.ac.uk/about/what-we-do/statutory-customers>

<sup>27</sup> Open Data Institute (2019), ‘Case study: The value of sharing data for benchmarking and insights’, <https://theodi.org/article/case-study-the-value-of-sharing-data-for-benchmarking-and-insights/>

## Decision makers

In a data ecosystem, the ‘decision makers’ are people or organisations that are able to make better decisions informed by the products and services, analyses and insights, or stories and visualisations created by data users.

If a data ecosystem were to be drawn for a data institution like OpenCorporates, for instance, the journalists, investigators and NGOs that rely on OpenCorporates’ data to investigate corruption would be the direct data users, while the decision makers would be organisations and members of the public who use their outputs to make decisions about how to avoid corrupt organisations.

As with many of the categories in this diagram, the distinction between ‘data users’ and ‘decision makers’ can become blurred depending on the ecosystem that surrounds a data institution and the functions that a data institution performs. For instance, if a shipping company were to receive data from HiLo that required further analysis or processing in order to be useful, then that shipping company would be considered a data user. If, however, that shipping company were to receive ready-to-use, actionable insights from HiLo, then that company would be considered a decision maker, since it is in a position to make decisions based on those insights.

## Funders

In [Figure 1](#), the ‘funders’ do not directly interact with the data institution by contributing or using data stewarded by that institution, nor do they directly benefit in the same way that the decision makers do. Instead, funders tend to support data institutions with the aim of encouraging healthy data ecosystems, pursuing philanthropic goals or as part of performing their public function.

UK Biobank, the non-profit charity and long-term research project stewarding health data for 500,000 research participants, is supported entirely by revenue drawn from large funders like Wellcome and the Medical Research Council.

## Revenue models versus business models

It is important to distinguish between business models and revenue models.

An organisation’s business model describes the structure used in order to create, deliver, capture and distribute value.<sup>28</sup> Its revenue model, on the other hand, describes how it generates income. A tool commonly used to describe an organisation’s business model is Strategyzer’s Business Model Canvas, developed by Alexander Osterwalder.<sup>29</sup>

---

<sup>28</sup> Osterwalder A, Pigneur Y (2010), ‘Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers’, <https://www.wiley.com/en-gb/Business+Model+Generation%3A+A+Handbook+for+Visionaries%2C+Game+Changers%2C+and+Challengers-p-9780470876411>

<sup>29</sup> Strategyzer (n.d.), ‘Business Model Canvas’, <https://www.strategyzer.com/canvas/business-model-canvas>



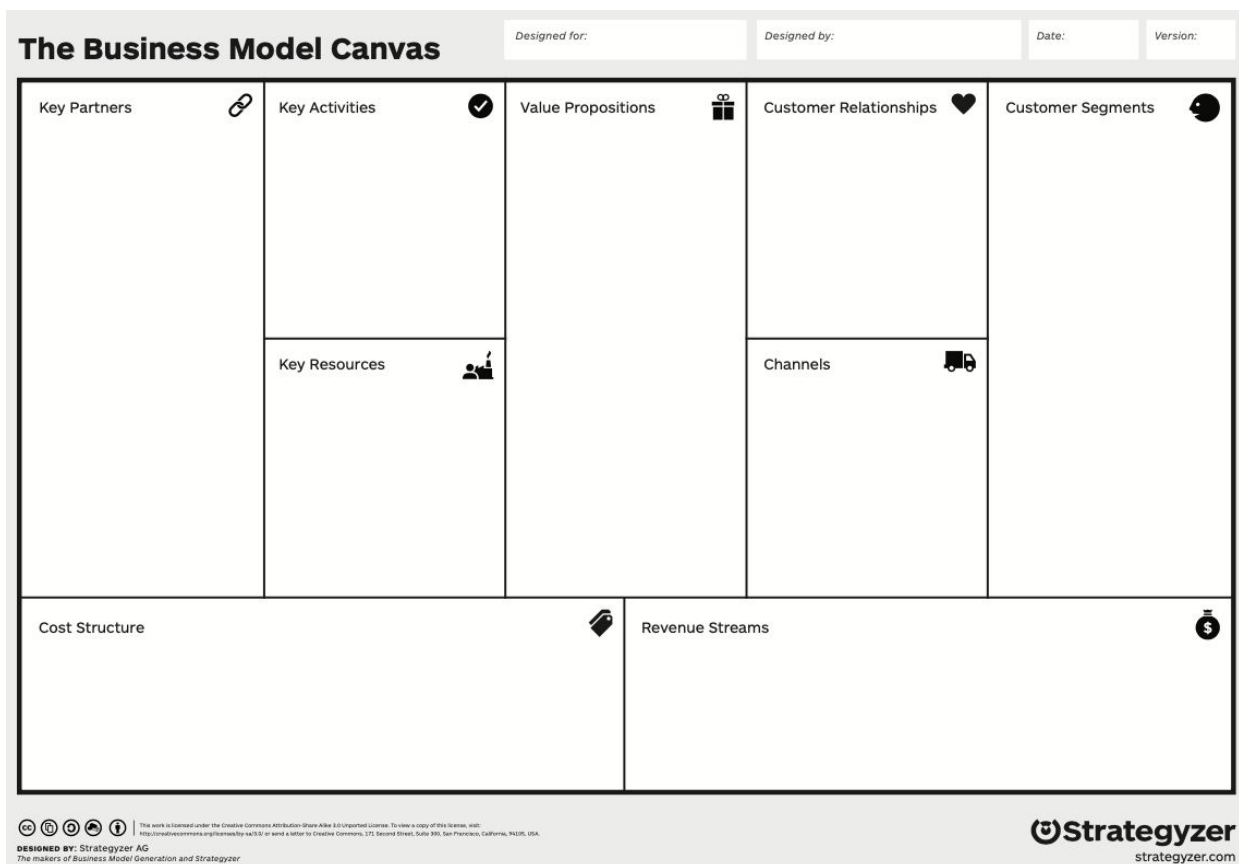


Figure 2: The Business Model Canvas

Looking at the example ecosystem in [Figure 1](#), we can see some of the varied ways an institution might generate revenue from different actors. This might be by earning money from selling insights about data, receiving subscription fees, or from grant funding. Regardless of the legal form of an organisation, a stable revenue stream is essential to its long-term sustainability.

The types of revenue that data institutions are able to access also depends on the company or organisational form that a data institution takes. In the ODI's experience, some data institutions may operate as for-profit companies with a commitment to generate profit for shareholders (for example [npm.Inc<sup>30</sup>](#)), whereas others may be set up as not-for-profit (for example HESA) with any excess funds reinvested into their mission.

In this project, our main focus has been on understanding the different revenue streams a data institution can use to become financially sustainable. In our analysis, we will continue to touch on the sustainability of the wider business model, but this warrants further research.

<sup>30</sup> npm, Inc (n.d.), <https://www.npmjs.com/>

## Applying the Business Model Canvas to data institutions

There are often said to be four factors that describe a business model<sup>31</sup>. These four factors, applied in the table below to the context of a data institution, map onto the building blocks of the Business Model Canvas.

Four factors describing a company's business model	Four factors describing a data institution's business model
1. WHO are the target customers (segments, relationships)?	1. WHO are the intended data users and those making decisions informed by the products and services, analyses and insights, or stories and visualisations created by them?
2. WHAT is the benefit that a company provides to its customers and most important partners in the value chain (value promise or value proposition)?	2. WHAT is the value that a data institution provides to its community? This can be around creating social or economic value or minimising harm.
3. HOW does the company deliver this benefit (partners, activities, resources)?	3. HOW does the data institution deliver this value through its role in the ecosystem? Key activities can include, for instance, making decisions about who should access data on behalf of an organisation or group of organisations, combining or linking data, and/or providing benchmarks and other insights back to the organisations that have contributed to them.
4. HOW does the company earn money (revenue model or revenue structure)?	4. HOW does the data institution generate revenue?

**Table 2:** Four factors that describe a business model, applied to the context of a data institution, mapping onto the building blocks of the Business Model Canvas.

A revenue model, therefore, is just one of the factors to consider when designing the business model for a data institution. In this report we are focusing primarily on revenue models.

<sup>31</sup> Lead Innovation Blog (2018), 'Business Model versus Revenue Model', <https://www.lead-innovation.com/english-blog/business-model-versus-revenue-model-what-is-the-difference>

## Funding versus financing

In our previous project to pilot data trusts,<sup>32</sup> we proposed a simple lifecycle (Figure 3) that described the stages of setting up and running a data institution.

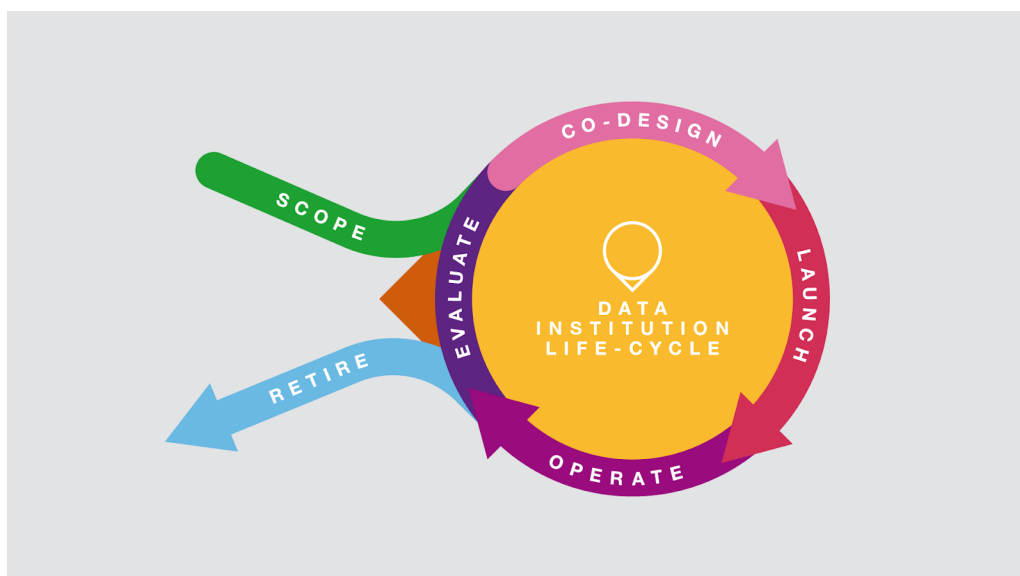
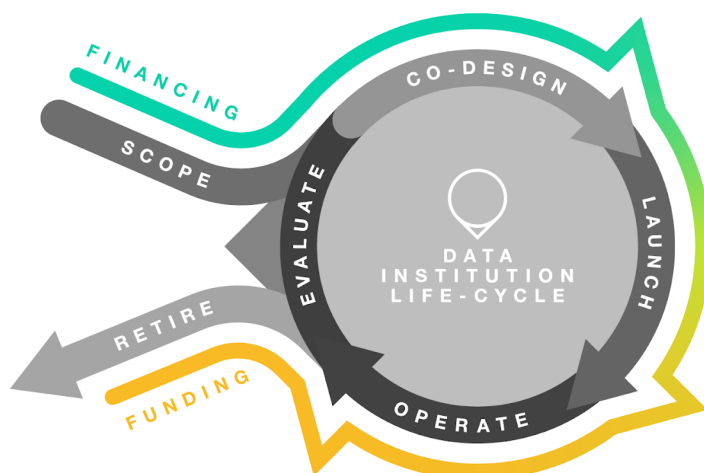


Figure 3: The data institution lifecycle

In this project, our hypothesis has been that at different stages of their lifecycle, data institutions will have different costs and revenue models. The financial costs of setting up a new data institution differ to the costs of running an established data institution. A more mature organisation will also have more sources of revenue available to it. As the data institution evolves, so might its revenue model.

When planning investments to develop physical infrastructure, a distinction is often made between financing and funding. Financing refers to covering the upfront costs of building infrastructure, while funding refers to paying for the infrastructure over its lifecycle.<sup>33</sup>

Figure 4 highlights the differences between financing and funding, in relation to the lifecycle diagram.



<sup>32</sup> Open Data Institute (2019), 'Data trusts: lessons from three pilots', <https://theodi.org/article/odi-data-trusts-report/>

<sup>33</sup> Institute for Government (2018), 'Funding infrastructure', <https://www.instituteforgovernment.org.uk/explainers/funding-infrastructure>

This distinction is useful because it highlights the need for early investment to unlock longer-term value. Developing a well-defined sustainable revenue model for infrastructure, whether physical or digital, is an important part of justifying that early investment.

Until a data institution has launched and is operational, it will have limited options for generating revenue from its activities, so it is likely to rely more on other revenue streams. Following launch, data institutions will also incur different kinds of costs: to cover their operations rather than to design their structure or policies.

During operation, data institutions need to ensure sufficient revenue not only to cover their day-to-day operational costs, but also to cover the costs of future investment in the infrastructure and costs to handle iterative change to their business model. In addition, data institutions need to be building up reserves to fund a possible retirement phase.

We explore cost structures and sources of revenue in the following sections.

## Cost structure

The Business Model Canvas describes the cost structure for a business as: “the most important costs incurred while operating under a particular business model. Creating and delivering value, maintaining customer relationships, and generating revenue all incur costs.”<sup>34</sup>

Many data institutions will have common infrastructure costs such as data storage, staff costs and fixed assets such as office space. But those costs will also vary depending on the role(s) that the data institution is taking on within the surrounding data ecosystem. The role of the data institution will differ depending on its mission; the needs of the ecosystem and community; the number of data contributors and users; the number of countries being operated in; and the sensitivity of the data being managed.<sup>35</sup>

These factors will all have an impact on operational costs. Some data institutions may focus their activities on governing access to data provided by data contributors. This means that costs of data collection and access may be carried by others. But other institutions may be more actively involved across the lifecycle of collecting, processing, managing, transforming and managing access to data. These all entail additional costs.

Anticipating the scale of costs may be quite challenging and will vary from data institution to data institution. For example, the dataset stewarded by ROR is relatively static and small, which means that infrastructure costs for providing shared services are similarly small. The small size means that it is likely data users will incorporate and use data directly in their own systems, which again reduces the type and cost of the services that ROR needs to provide.

---

<sup>34</sup> Strategyzer (n.d.), ‘How do I use the Cost Structure building block of the Business Model Canvas?’, <https://strategyzer.uservoice.com/knowledgebase/articles/1194382-how-do-i-use-the-cost-structure-building-block-of>

<sup>35</sup> Open Data Institute (2019), ‘Data trusts: lessons from three pilots’, p48, <https://theodi.org/article/odi-data-trusts-report/>

As noted in the previous section, operational costs are not the only costs that need to be considered when planning for sustainability. The ‘Full Cost Project’ initiative provides a useful categorisation of different types of organisational costs,<sup>36</sup> including:

- **Total expenses** – the day-to-day expenses of running an organisation.
- **Reserves** – savings that mitigate risks for the organisation, for example to manage uncertainty around funding streams, or to invest in researching and developing new products and services.
- **Change capital** – to invest in large, periodic change programmes that may impact the business model of the organisation, for example as part of an evaluation phase in the institutional lifecycle ([Figure 3](#)).
- **Fixed asset additions** – to cover purchase of new equipment or other assets, for example as part of improving or upgrading physical infrastructure, like sensors, to support data collection.

This section has proposed a framework for understanding the various aspects of financial sustainability by exploring them through the perspectives of a data ecosystem and the lifecycle of an institution. The next section outlines the broad patterns we identified through our research.

## Case study: ROR

The Research Organization Registry (ROR) is a community-led project that started off with 17 organisations coming together around the shared vision of creating an open, community-led organisation identifier registry for the scholarly research community. Three organisations – Crossref, DataCite and ORCID – donated staff time and/or financial resources to the ORG ID Working Group to move the initiative forward between 2016 and 2018.

The working group, set up to develop a plan for the operation and governance of the registry, considered a range of possible options,<sup>37</sup> including forming a new independent organisation – the most costly, but an option that helps ensure focus and stakeholder representation; and a hosted model sharing operational resources with existing organisations – an option that comes with reduced costs but increased risk and legal liability.

The lead organisations supporting the project decided to launch an initial start-up phase, adopting a hosted model. A pilot registry was subsequently launched in January 2019, with governance and other community layers to be ultimately built around it. As of December 2019, the registry includes identifiers and metadata for more than 97,000 research organisations. Operations are administered by California Digital Library, Crossref and DataCite, and it is advised by a larger steering group. The pilot benefited from a donation of seed data from Digital Science’s GRID database.

The project so far has been entirely funded from donations, including staff time, financial resources and seed data. To be able to strengthen its capacity

<sup>36</sup> Philanthropy CA (2018), ‘Full Cost Project’, [https://www.philanthropyca.org/full-cost-project?qt-full\\_cost\\_project=1#qt-full\\_cost\\_project](https://www.philanthropyca.org/full-cost-project?qt-full_cost_project=1#qt-full_cost_project)

<sup>37</sup> Figshare (2017), ‘ORG ID WG Governance Principles and Recommendations’, [https://figshare.com/articles/ORG\\_ID\\_WG\\_Governance\\_Principles\\_and\\_Recommendations/5402002/1](https://figshare.com/articles/ORG_ID_WG_Governance_Principles_and_Recommendations/5402002/1)

and develop new features, ROR launched a fundraising campaign in October 2019 and plans to launch an optional paid service tier from 2022 to recover costs. The ambition is to eventually generate 100% of the revenue from the optional paid-service tier. The registry's data will remain open and free for anyone to use.

While strategic decision making about operations and governance is the responsibility of ROR's steering group, the organisation engages a strong network of community advisers – publishers, researchers, librarians, funders, data analysts and metadata experts from around the world – through regular calls. ROR seeks ideas and feedback from the community about the directions it is heading, and sees this as central to the growth of the organisation.

## Sources of revenue

There is a wide range of revenue sources available to data institutions, based on drawing income from different actors and their role in the ecosystem they support.

It is helpful to draw a distinction between two main categories: earned revenue (revenue earned from key activities of the data institution like the supply of data, services or membership fees); and non-earned revenue (grants, donations and other funding sources). Distinguishing between these two main sources of revenue can help us understand how they align with the goals of a data institution, the actors providing that revenue, and the ability to access those sources at different stages in the data institution's lifecycle.

### Earned revenue

Earned revenue can be desirable for a data institution as it tends to be more stable than donations or grant funding, and there is the opportunity for growth. Earned revenue streams typically involve a direct value exchange: money is paid in return for a service (or other output) to be delivered. This means the data institution has to invest in understanding the needs of those who are paying, ensure they can service the commercial arrangement, and deliver that value.

Corporate Finance Institute,<sup>38</sup> for instance, defines four common mechanisms of generating revenue:

1. **Transaction-based revenue**, which they describe as “proceeds from sales of goods that are usually one-time customer payments”.
2. **Service revenue**, which are revenues “generated by providing service to customers and are calculated based on time”, such as revenue generated from things like consulting services.
3. **Project revenue**, which can be understood as revenue “earned through one-time projects with existing or new customers”.
4. **Recurring revenue**, which they define as “ongoing payments for continuing services or after-sale services to customers”, such as subscription fees;

---

<sup>38</sup> Corporate Finance Institute (n.d.), 'Revenue Streams', <https://corporatefinanceinstitute.com/resources/knowledge/accounting/revenue-streams/>

renting, leasing or lending assets; licensing content to third parties; brokering fees; and advertising fees.

The Business Model Canvas<sup>39</sup> lists seven ways of generating revenue:<sup>40</sup>

1. **Asset sale:** This revenue stream derives from selling ownership rights to a physical product.
2. **Usage fee:** This revenue stream is generated by the use of a particular service. The more a service is used, the more the customer pays.
3. **Subscription fees:** This revenue stream is generated by selling continuous access to a service.
4. **Lending/renting/leasing:** This revenue stream is created by temporarily granting someone the exclusive right to use a particular asset for a fixed period in return for a fee.
5. **Licensing:** This revenue stream is generated by giving customers permission to use protected intellectual property in exchange for licensing fees.
6. **Brokerage fees:** This revenue stream derives from intermediary services performed on behalf of two or more parties.
7. **Advertising:** This revenue stream results from fees for advertising a particular product, service or brand.

Drawing on these categories, the following table maps out a range of revenue streams available to data institutions that we have observed through our desk research and interviews. We anticipate that there are other revenue streams in use by existing data institutions and additional potential revenue sources which we will identify and catalogue in our future research.

Type	Actor	Revenue stream
<b>Subscription fees</b>	Data contributors, data users, decision makers	Membership fees for access to services, data and other direct benefits provided by the institution
	Data contributors, data users, decision makers	Subscription fees for use of the infrastructure. These fees might be optional, or mandated as part of the legal form of an institution
	Data contributors, data users, decision makers	Subscription fees for access to a specific product or service that uses the data
	Data contributors, data users	Subscription fees for access to a higher-level service level agreement or support arrangement for using the infrastructure
<b>Usage fees</b>	Data contributors	Fees for depositing, submitting or updating data, via a standard API or submission form. Includes registration of organisations, content and identifiers for parts, products and services into an official register.
	Data contributors	Fees for depositing, submitting or updating data, via a bespoke method or custom format. Includes registration

<sup>39</sup> Strategyzer (n.d.), 'Business Model Canvas', <https://www.strategyzer.com/canvas/business-model-canvas>

<sup>40</sup> Strategyzer (n.d.), 'How do I use the Revenue Streams building block of the Business Model Canvas?', , <https://strategyzer.uservoice.com/knowledgebase/articles/1194385-how-do-i-use-the-revenue-streams-building-block-of>

		of organisations, content and identifiers for parts, products and services into an official register.
	Data users	Fees for accessing and using data via a standard API
	Data users	Fees for accessing and using data in bulk
	Data users	Fees for accessing and using data in a customised format or delivered via a bespoke method
	Data users	Fees for more timely access
	Data contributors, data users, decision makers	Consultancy fees for provision of advice or custom analytics using data
	Data contributors, data users, decision makers	Training fees to advise on use of the data and infrastructure
<b>Licensing</b>	Data users	Data licensing fees, for example to provide additional rights for distribution or to remove need for attribution
	Data users	Revenue share in new products and services that use data provided by the institution

**Table 3:** Revenue streams available to data institutions, observed through desk research and interviews by the ODI

## Grants, donations and other non-earned revenue sources

The fact that data institutions are often established to promote particular social, charitable or educational aims can open up access to other forms of support, both financial and non-financial.

The costs of running non-profit organisations, charities or social enterprises – legal forms that many data institutions take – are most often covered by direct funding from governments, philanthropists or other funders. Financial support typically includes grants, loans, tax breaks, cash donations and sponsorship.

In contrast to earned revenue streams, this category may involve fewer direct returns of value to the funder, as grants and similar revenue are generally given due to a high-level alignment between the funders' own goals and the mission of the funded data institution. However, depending on the flexibility in how grants are awarded and what they can be spent on, there may be restrictions on how the data institution can spend the money. As an example, the UK government defines seven different types of grant funding which vary in how they are awarded and how money can be spent.<sup>41</sup>

Financial support may also come by way of loans or tax breaks. Loans may be an option for data institutions who need to cover capital costs or to boost cash flow. Governments may give a tax break to: “a particular group of people or type of organization, it reduces the amount of tax they have to pay or changes the tax system in a way that benefits them. A tax break can greatly reduce a taxpayer's liability and provides savings through tax deductions, tax credits, tax exemptions, and other incentives.”<sup>42</sup>

<sup>41</sup> GOV.UK (2017), ‘Government Grants Definitions’, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/654680/2017-09-27\\_Grant\\_Definitions.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/654680/2017-09-27_Grant_Definitions.pdf)

<sup>42</sup> Investopedia (2018), ‘Tax Break Definition’, <https://www.investopedia.com/terms/t/tax-break.asp>



Some data institutions also rely on non-financial support, such as donations of goods, and for many, the donation of services or time by volunteers is essential to achieving sustainability. These donations-in-kind can be combined with financial sources of support when appropriate, but some organisations, such as ROR, are supported primarily by non-financial contributions.

The following table lists some non-earned revenue sources for data institutions.

Type	Actor	Revenue stream
<b>Grants</b>	Funders	Grants-in-aid provided as general support for the institution in meeting its objectives
	Funders	General grants awarded in anticipation of the institution using it for a specific agreed purpose, for example, to carry out a project or invest in infrastructure
<b>Loans</b>	Data contributors, data users, decision makers, funders	Loans to help finance institution during early stages of its lifecycle
<b>Sponsorship and cash donations</b>	Data contributors, data users, decision makers	One-off or regular cash donations and sponsorship
<b>Tax breaks</b>	Funders (government)	Tax breaks to help increase reserves and support investment
<b>In-kind donations</b>	Data contributors, data users, decision makers	Donations of hardware or other physical resources
	Data contributors, data users, decision makers	Provision of staff time to provide support during setup and operations
	Data contributors, data users	Provision of technical infrastructure to support operations of the institution, for example hosting of services
	Data contributors, data users	Provision of staff time to develop and maintain software used by the institution and its ecosystem
	Data contributors, data users, decision makers	Volunteer time to support collection and curation of data, community engagement, software development, etc

**Table 4:** Non-earned revenue sources for data institutions

# Common patterns in data institution sustainability

Drawing on our desk research and interviews, we have identified some common patterns in how data institutions have been working to achieve sustainability.

## Data institutions use mixed revenue streams

Through our desk research and interviews, we have seen a range of revenue models that data institutions are relying upon at different stages of their lifecycle. These include both earned revenue, such as recurring membership and subscription fees (HiLo, HESA, OpenCorporates, MusicBrainz), or application fees for researchers (UK Biobank), and other models such as grants (OpenCorporates, HESA, HiLo) and financial and/or in-kind donations in the form of staff time or seed data (ROR, MusicBrainz). Some institutions also earn money from the provision of one-off services, for example bespoke consultancy or training (HESA).

Of the data institutions that we have researched, we noted that they cover their costs through a combination of two or three different revenue models, often with one dominant revenue source. For instance, HESA generates 90% of its revenue from mandatory subscriptions from higher education providers, and OpenCorporates generates 75% of its revenue through API subscription fees.

This mix of revenue models is very diverse and is dependent on the context of the data institution, although of the institutions we have reviewed, the majority seem to be aiming for more earned revenue over time, with an emphasis on recurring revenue streams.

## Five approaches to financing data institutions

The funding needed by data institutions changes throughout their lifecycle. A data institution initially needs financing to afford the upfront costs of setting up a new organisation. Activities that take place within the scoping, co-design and launch phases of the data institution lifecycle, including but not limited to market research, legal fees for incorporation and resources to start organisational activities, will all need to be covered financially.

None of the institutions that we have explored as part of this research were financed through venture capital.

So far we have identified five different approaches to getting a data institution started:

- 1. Grants to cover scoping, co-design and launch costs**

Some data institutions have benefited from grant funding to help get their project off the ground. This funding covers the initial scoping and pilot phase of an initiative prior to the establishment of an independent legal entity.

## 2. Being incubated by an existing organisation

HiLo, for instance, was incubated by Shell, and after it demonstrated value, was spun out as an independent organisation with the support of a grant from Lloyd's Register Foundation.

## 3. Being hosted by a non-profit organisation

There is also a hosted model, particularly prevalent in the scholarly research sector, in which an existing organisation hosts a data institution either for an initial start-up phase or on a more permanent basis, committing dedicated staff resources and funding to it.

## 4. Joint financing by a group of organisations

ROR is an example of a project hosted by a group of organisations. In its documentation, ROR describes this model as beneficial in terms of lower start-up costs (no need for separate payroll or finance) with the flexibility to expand later and hire staff directly.<sup>43</sup> Other successful institutions in the research infrastructure field that were financed by other organisations include ORCID<sup>44</sup> and Crossref.<sup>45</sup>

## 5. Financing through personal investment

For several years, OpenCorporates was funded via direct investment from its founder, Chris Taggart. This undeniably supported OpenCorporates while the organisation was trying to grow, and ultimately move towards a more sustainable business model.

## Case study: OpenCorporates

OpenCoporates, founded in 2011, is the largest open database of companies in the world, with information collected on 180 million companies, from about 140 different jurisdictions. OpenCorporates is a social enterprise and therefore a “for profit working in public benefit”.

Following a couple of years of financial support from the founder, Chris Taggart, the organisation was supported by grants from the likes of the Sloan Foundation.

OpenCorporates talks of its ‘unique’ public-private benefit business model which combines open and free access to data for all customers, with a tiered payment structure for companies to purchase access to bulk structured data and APIs. The commercial revenue is used to enhance the free offering, which increases the number of people who are likely to use the service. It believes in

---

<sup>43</sup> Figshare (2017), ‘ORG ID WG Governance Principles and Recommendations’, [https://figshare.com/articles/ORG\\_ID\\_WG\\_Governance\\_Principles\\_and\\_Recommendations/5402002/1](https://figshare.com/articles/ORG_ID_WG_Governance_Principles_and_Recommendations/5402002/1)

<sup>44</sup> ORCID (n.d.), <https://orcid.org/>

<sup>45</sup> Crossref (n.d.), <https://www.crossref.org/>

the ‘many eyes’ approach of improving the product for everyone, including commercial customers.

The team strives for around 80% of revenue to be attained via commercial routes, with the final 20% grant funding to use for research and development.

In 2018, the OpenCorporates Trust was formally established to protect the long-term sustainability of the organisation. It protects the organisation – and the data it stewards – from being sold to a third party, as has happened in recent years with other company data sites such as Bureau van Dijk<sup>46</sup> and Avention<sup>47</sup>.

## Revenue models evolve over time

Over time, several organisations have shifted away from, or talked about plans to shift away from, initial grant funding towards more dependence on earned revenue from their key activities. For example:

- HiLo has plans to generate 40% of its revenue from new analytical products and new ways of accessing and using data, with the rest coming from subscription fees.
- OpenCorporates has also moved from an approximately equal split between grants and API service subscription fees to a 25–75% split over time respectively.
- HESA is expecting to see an increase in the revenue generated through the bespoke products and services – such as tailored datasets and reports, analytical services and training<sup>48</sup> – offered in collaboration with Jisc, with whom they have a revenue sharing arrangement.
- ROR’s ambition is to eventually transition completely from grants and donations, to a commercial model (which has still to be defined).

The UK Biobank, on the other hand, has a 10-year history of being funded through grants from a number of major donors. It is the only medical research resource among the data institutions we explored in detail and it was set up with significant financial backing from major governmental and charity donors.

In addition to evolving their revenue model, some data institutions may strive to evolve their business model over the lifecycle of the institution. In particular, data institutions may seek to change how they deliver value, for example by adding analytical capacity or providing technical infrastructure for data collection. This will obviously add to the costs of a data institution and have an impact on its sustainability.

---

<sup>46</sup> Reuters (2017), ‘Moody’s Corp to buy Bureau van Dijk for about \$3.3 billion’, [reuters.com/article/us-bureauvandijk-m-a-moody-s-idUSKCN18B1CZ](https://www.reuters.com/article/us-bureauvandijk-m-a-moody-s-idUSKCN18B1CZ)

<sup>47</sup> Bloomberg (2017), ‘Dun & Bradstreet Acquires Avention, the Maker of OneSource; Expands Presence & Offerings in B2B Sales Acceleration Market’, <https://www.bloomberg.com/press-releases/2017-01-10/dun-bradstreet-acquires-avention-the-maker-of-onesource-expands-presence-offerings-in-b2b-sales-acceleration-market>

<sup>48</sup> HESA (n.d.), ‘HESA data analytics services are now delivered by Jisc’, <https://www.hesa.ac.uk/services>

Data institutions that are planning to evolve their business model will need to ensure that they have acquired appropriate capital to cover the costs of evaluation and subsequent change processes, which may require additional short-term investments (for example on fixed assets) as well as an updated future cost structure.

## Institutions plan for future investments

Data institutions also need to plan for covering the costs of operational maintenance and future investments in modernising their infrastructure, and the iterative process of evaluation and re-design. In our research, we have seen the following approaches to funding investments.

HESA highlighted the importance of not only covering its cost base, but establishing the financial headroom to be able to reinvest in the organisation and maintain operational effectiveness and clientele. HESA noted: “With a changing data environment and changing expectations from stakeholders and customers, comes a need to reinvest and update systems and processes.” Forming collaborations rather than necessarily raising new funds was one way HESA was accomplishing offering new services. A department of HESA was transferred to Jisc in 2019 to provide advanced data analytics for the higher education sector. Collaborations and sharing services with other organisations is also a way of increasing efficiency in the delivery of a public good in the case of HESA.

In our research we looked at ‘trust ports’ to see if we can learn from their practices. Trust ports are a type of authority in the UK that own and operate UK ports. Trust ports are set up by an Act of Parliament and controlled and administered by a self-governing independent statutory body. Trust port governance structures are particularly designed to reflect local markets and the needs of local communities.

We found trust ports comparable to data institutions, although their role is stewarding a physical, rather than a digital, resource on behalf of a community. There are approximately 100 trust ports in the UK. The trust model: “allows the Harbour to operate in a commercial environment with no direct public funding, but also enables it to compete in the open market with private and local authority-owned facilities that represent the vast majority of the UK’s ports”<sup>49</sup>. Trust ports are independent statutory bodies. They “[have] no shareholders who require a dividend, instead all ‘profit’ or surpluses made from Trust Port operations are reinvested into the operation, maintenance and administration of the port.”

A trust port study from 2016<sup>50</sup> listed potential ways trust ports can fund investment to improve or replace their assets and to develop or acquire new ones. These include their own resources (retained profits and reserves), borrowing from commercial sources, entering into joint ventures with third parties or accessing publicly provided support, for example through Local Enterprise Partnerships and Local Growth Funds, or from European Union sources such as the Connecting Europe Facility. The study, however, also found that “the trust port model may not necessarily be the correct one to deliver strategic objectives of ports where they involve major transformative investments”. We will discuss the corresponding challenges in the next chapter.

---

<sup>49</sup> Aberdeen Harbour (n.d.), ‘What makes a trust port’,

<http://www.aberdeen-harbour.co.uk/article/what-makes-a-trust-port/>

<sup>50</sup> Department for Transport (2016), ‘Trust Port Study Key Findings and Recommendations’,

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/525076/trust-port-study-key-findings-and-recommendations.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/525076/trust-port-study-key-findings-and-recommendations.pdf)

The ‘Power of Partnership’ report<sup>51</sup> on different approaches to public–private partnership in the provision of hydromet services – services that provide real-time weather, water, early-warning and climate information products to end users<sup>52</sup> – also highlights the need to plan for maintaining and strengthening the weather infrastructure as problems may arise over time<sup>53</sup>. Similarly, the ‘Weathering the Change: How to Improve Hydromet Services in Developing Countries’ report<sup>54</sup> provides some guidance on how to take a national meteorological and hydrological service through a process of modernisation and ensure adequate financial planning.

## Data institutions develop contingency plans

We explored whether and how data institutions are planning to ensure the sustainability of data and related infrastructure beyond their own existence. We were interested in whether, in line with the idea of putting in place insurance mechanisms for when something goes wrong,<sup>55</sup> they had considered succession planning and had put in place any mechanism to enable the community to replicate the system if their organisation was to wind down.

Most of the institutions we talked to are thinking about contingency planning. For HESA, as an agency with a statutory function, this is focused on maintaining its role as the provider of that function in the long term. To do that, HESA does scenario planning and reassesses the situation periodically. For other data institutions, the focus is on ensuring the sustainability of the data in case they were no longer operational. Some institutions have built on open-source practices and integrated relevant considerations into the design of their institutions. ROR’s ‘Governance recommendations’<sup>56</sup> from 2017 – before the official launch of the registry – highlights that “the new Registry organisation should have protections in place for assets to go to another non-profit with a similar mission if the organisation is wound up.” In line with the recommendations, all software developed to run and support the registry is open source, which means others could feasibly run the same code if ROR ceased to exist.

The ‘Data Trust for Open Access Ebook Usage’, a project in its pilot stage to improve the measurement and analysis of open-access ebooks, plans to create “a technological infrastructure released as open-source code with full documentation”.<sup>57</sup>

OpenCorporates, besides making everything openly available, also ensures that the data it stewards cannot be sold, by incorporating the OpenCorporates Trust. The Trust helps to ensure that the organisation is independent and the data will always be available to public benefit and commercial users.<sup>58</sup>

---

<sup>51</sup> Global Facility for Disaster Reduction and Recovery (2019), ‘The Power of Partnership: Public and Private Engagement in Hydromet Services’, [https://www.gfdrr.org/sites/default/files/publication/The%20Power%20of%20Partnership\\_WEB.pdf](https://www.gfdrr.org/sites/default/files/publication/The%20Power%20of%20Partnership_WEB.pdf)

<sup>52</sup> The World Bank (2017), ‘Hydromet’, <https://www.worldbank.org/en/results/2017/12/01/hydromet>

<sup>53</sup> Ibid, p80

<sup>54</sup> Global Facility for Disaster Reduction and Recovery (2019), ‘Weathering the Change: How to Improve Hydromet Services in Developing Countries’, [https://www.gfdrr.org/sites/default/files/publication/weathering%20the%20change\\_web\\_0.pdf](https://www.gfdrr.org/sites/default/files/publication/weathering%20the%20change_web_0.pdf)

<sup>55</sup> Bilder G, Lin J, Neylon C (2015), ‘Principles for Open Scholarly Infrastructure’, <https://wiki.lib.sun.ac.za/images/f/f6/2015-principle-for-open-scholarly-communication-infrastructures.pdf>

<sup>56</sup> Figshare (2017), ‘ORG ID WG Governance Principles and Recommendations’, [https://figshare.com/articles/ORG\\_ID\\_WG\\_Governance\\_Principles\\_and\\_Recommendations/5402002/1](https://figshare.com/articles/ORG_ID_WG_Governance_Principles_and_Recommendations/5402002/1)

<sup>57</sup> Hawkins K (2019), ‘Grant Proposal: Developing a Data Trust for Open Access Ebook Usage’, <https://digital.library.unt.edu/ark:/67531/metadc1596980/?q=legal>

<sup>58</sup> OpenCorporates (n.d.), ‘Governance’, <https://opencorporates.com/info/governance/>

## Good governance and community support are critical to sustainability

Good governance and a strong supporting community are frequently discussed in the context of sustainable organisations. Although we have focused much of this initial research on financial sustainability, we recognise that “the problems of sustainability are not merely ones of finance but of political economy, which means that focusing purely on financial sustainability in the absence of considering governance principles and community is the wrong approach.”<sup>59</sup>

The ‘Principles for Open Scholarly Infrastructures’<sup>60</sup> paper details the interlinked nature of governance and community: “trust is built on three pillars: good governance (and therefore good intentions), capacity and resources (sustainability), and believable insurance mechanisms for when something goes wrong.”<sup>61</sup> When “an infrastructure is successful [it] becomes critical to the community” it needs to be ensured that “it is not co-opted by particular interest groups. Similarly, we need to ensure that any organisation does not confuse serving itself with serving its stakeholders.”

The article ‘Sustaining Scholarly Infrastructures through Collective Action: The Lessons that Olson can teach us’<sup>62</sup> sets out that: “Crises of financial sustainability (or challenges of expansion) for these organisations are often coupled with or lead to a crisis in governance and/or community trust.”

In the interviews we conducted, several organisations highlighted one and/or the other aspect as key to their sustainability.

ROR, for instance, mentioned the importance of community uptake and adoption with “community [being] at the heart of what ROR is trying to do... [They] need the data to be used and adopted and the data trusted by the people who are using it.” ROR sees its work, including decisions around its revenue model, as guided by community needs and interests. MusicBrainz also emphasised that “the bottom line is about helping customers take and use the data” and “having an offering to customers that you are serious, credible and also technically credible”.

The good governance guidance for statutory harbour authorities<sup>63</sup> provides detailed guidance for trust ports in particular on aspects of good governance, including stakeholder engagement, given their nature as independent statutory bodies. The key principles established in the guidance include being guided by the interests of the wide community of stakeholders at all times, reinvesting surpluses to support the long-term success of the port, and consulting stakeholders on significant decisions.

---

<sup>59</sup> Neylon C (2017), ‘Sustaining Scholarly Infrastructures through Collective Action: The lessons that Olson can teach us’, <https://kula.uvic.ca/articles/10.5334/kula.7/>

<sup>60</sup> Bilder G, Lin J, Neylon C (2015), ‘Principles for Open Scholarly Infrastructure’, <https://wiki.lib.sun.ac.za/images/f/f6/2015-principle-for-open-scholarly-communication-infrastructures.pdf>

<sup>61</sup> Neylon C (2017), ‘Sustaining Scholarly Infrastructures through Collective Action: The lessons that Olson can teach us’, <https://kula.uvic.ca/articles/10.5334/kula.7/>

<sup>62</sup> Ibid.

<sup>63</sup> GOV.UK (2018), ‘Ports Good Governance Guidance: Moving Britain Ahead’, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/684839/ports-good-governance-guidance.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684839/ports-good-governance-guidance.pdf)

In general, good organisational design is essential to the creation of both sustainable and trustworthy institutions. For more discussion on this, see the section on ‘Organisational design’ in our report on ‘Designing trustworthy data institutions’.<sup>64</sup>

# Issues and tensions with achieving sustainability

In the previous section we highlighted some of the common approaches that institutions are taking towards building sustainable revenue models. In this section we highlight some of the common issues and tensions that we have observed.

## Difficulty in attracting early financing

As our research on data trusts has highlighted, all data institutions will need to go through adequate scoping and co-design stages. These stages will help to identify the purpose for the data institution, align its community around a common vision and ensure that it is designed to achieve its goals, reflect the need to build trust through strong governance, and plan for longer-term sustainability.

These phases may take some time to complete. They are likely to be iterative as new ideas and approaches are tested with the community. Completing this process requires both time and investment.

With a limited ability to raise revenue before being fully operational, in order to finance these early stages, data institutions are likely to rely heavily on grants, voluntary time and other in-kind investments. This emphasises the need to clearly articulate a shared vision and the ability to engage with funders who are comfortable with a strong level of uncertainty in outcomes.

Loans, debt financing and venture funding are alternatives to grants, but these have their own limitations. Loans and debt financing will require some certainty around long-term income so that funders can be certain they will be repaid. This implies that some early groundwork must be completed which will still require financing. Venture funding is often geared around equity which, as we noted in the previous section, may create additional tensions around sustainability and the long-term goals of the institution. Impact investment, where success is measured in terms of social or environmental impact alongside financial benefits, may be a better fit.

OpenCorporates explained during our interview how many forms of finance are off the cards for it because it is a social enterprise, and therefore, there are no profits or assets to reclaim or benefit from. The end goal of venture capital funding is an exit – or “someone else taking control, whether it’s stakeholders in the stock market, or other owners squeezing it” – and the independent structure of OpenCorporates prevents that.

In our research we also came across an issue trust ports had in attracting financing for major investments: “trust ports generally face a more limited range of options on

---

<sup>64</sup> Open Data Institute (2020), ‘Trustworthy data institutions’, <https://theodi.org/project/rd-trustworthy-data-institutions>



how to deliver [major transformative] investment compared to ports in private ownership... [Trust ports] do not have access to new equity and are not part of wider groups of ports potentially with access to wider and deeper sources of funding.”<sup>65</sup>

Existing incubation models, which can provide a range of financial and non-financial support to early stage organisations, also tend to be geared more towards businesses rather than social enterprises. However, there are a growing number of accelerators and incubators that cater for non-profit organisations.<sup>66</sup>

## Aligning revenue models with institutional goals

There are a number of potential tensions between the purpose of an institution and the sources of revenue that are available.

The following table summarises a range of tensions in relation to some common institutional goals we have seen across data institutions. A specific institution may be trying to fulfil several of these goals as part of its broader purpose. Resolving these tensions may help to increase sustainability and trust in the institution.

Institutional goal	Tensions related to specific revenue models and possible alternatives
<p><b>Governing access to data on behalf of a group of data contributors</b></p> <p>For example, to mitigate harms as a trusted intermediary or broker</p>	<p>Charging data users to access data creates tension by incentivising the data institution to share data with more users to secure greater revenue.</p> <p><b>Alternatives:</b> charging for services to analyse stewarded data to provide insights, rather than charging for data access; charging data contributors to steward data on their behalf.</p>
<p><b>Increasing access to data to enable it to be used by a broader ecosystem</b></p> <p>For example, to provide a centralised pool of authoritative, standardised data</p>	<p>Charging data users to access or license data creates tension because to avoid data users providing onward access to that data, data institutions are incentivised to limit the ability of data users to enrich and redistribute data for other purposes.</p> <p>Providing services to analyse data creates tension because a data institution may be competing with its own data users.</p> <p><b>Alternatives:</b> providing optional services, such as training or consulting, to support use of data that is available via a range of options, for example, using standard or custom data feeds, under open licences.</p>
<p><b>Long-term stewardship of a data asset</b></p> <p>For example, to archive or ensure ongoing access to data over the long term</p>	<p>Relying on time-limited funding, such as grants, or uneven funding, such as donations, creates tension by reducing the ability for the data institution to predict future revenues and plan investments in its core infrastructure.</p> <p><b>Alternatives:</b> charging both data contributors and users recurring membership and subscription fees to cover provision</p>

<sup>65</sup> Department for Transport (2016), ‘Trust Port Study Key Findings and Recommendations’, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/525076/trust-port-study-key-findings-and-recommendations.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/525076/trust-port-study-key-findings-and-recommendations.pdf)

<sup>66</sup> Donorbox (2018), ‘9 Best Startup Accelerators For Nonprofits And Social Entrepreneurs’, <https://donorbox.org/nonprofit-blog/best-startup-accelerators-for-nonprofits/>

	of shared infrastructure; providing services to support the analysis and use of data.
<p><b>Maintaining a data asset on behalf of a community</b></p> <p>For example, supporting the collaborative maintenance of a shared resource</p>	<p>Charging data contributors for making contributions to the dataset creates tension as it discourages regular contributions or participation by some contributors.</p> <p><b>Alternatives:</b> providing optional services to support use of data that is available via a range of options, under open licences; charging both data contributors and data users membership fees to cover provision of shared infrastructure.</p>
<p><b>Providing a data-enabled service to a community</b></p> <p>For example, benchmarking or secure analysis of an aggregate dataset</p>	<p>Reliance on membership fees, donations or time-limited funding creates tension as these sources may not scale in line with the demand for service provision.</p> <p>Generating revenue from providing access to data may create tension if the core needs of the community aren't met by the services developed by data users.</p> <p>Revenue sources based on charging for updating or depositing data may also create tension if it discourages a wider community from contributing to the common datasets, leading to reduced utility.</p> <p><b>Alternatives:</b> directly charging for use of the primary service; charging membership fees for decision makers and contributors.</p>
<p><b>Supporting a broad community</b></p> <p>For example, enabling a community of small and large businesses and non-profit organisations to participate in the data ecosystem</p>	<p>Use of standard fees for membership can create tension by discouraging startups or not-for-profits in contributing to, or from using, data.</p> <p>Charging for all data uses may create tension by discouraging use of data by smaller organisations and not-for-profits that are unable to cover costs.</p> <p><b>Alternatives:</b> use scaling membership fees, freemium models and tiered pricing to enable a broader set of uses; provide models that support use of data for research or product development purposes.</p>
<p><b>Generation of revenue for a third party</b></p> <p>For example, shareholders, investors or funders that are not directly participating in the data ecosystem</p>	<p>Using a data institution as a means of generating and extracting revenue from a community creates tension as it means the data institution is not able to return that value to its data ecosystem, and invest in and grow its core infrastructure.</p> <p><b>Alternatives:</b> use institutional forms and revenue models that ensure that value is retained in the data institution, with surplus either reinvested or shared back with the ecosystem in other forms.</p>
<p><b>Independent stewardship of data</b></p>	<p>The need to maintain independence around the purpose of a data institution and the goals of its community can create tension when using revenue from funders or investors that have their own goals or priorities. This is especially true when funding comes with limitations on how money can be spent, or with expectations on specific deliverables.</p> <p><b>Alternatives:</b> find funders or investors whose goals and values align with the mission of the organisation; look for core/unrestricted funding rather than restricted or project grants.</p>

**Table 5:** Summarising a range of tensions in relation to some common institutional goals we have seen across data institutions

These tensions and the alternatives can be illustrated with reference to some of the data institutions and similar organisations that we have explored in our desk research and interviews. For example, OpenCorporates stewards a wealth of company data that it could use to generate revenue by selling access to it. But, it believes that this data should be open and available to everyone for societal benefit: “if people don’t have access to this dataset then there will be an asymmetry of access... which is highly problematic for society.” Therefore, OpenCorporates does not “restrict access to data or its website based on how much money you have, nor do you get access to more or different data”. Their model is therefore largely based around balancing tensions between openness and the need to apply some restrictions on the use of the data to be sustainable. ROR similarly suggested that this is the reason behind its reticence in turning their offering into a commercial product.

OpenCorporates believes that having many ‘open eyes’ on the data – people to look at, access, use, share and query the data – will drive in commercial customers. All of OpenCorproates’ data is available openly for all to use, but additional services, such as access to bulk structured data, or use of the API, are charged for. The commercial customers provide revenue for OpenCorporates to improve the data, and the services on top of it, for a better product for everybody. In a blog from 2017, they stated that they “don’t see commercial income as a necessary evil, but as an essential and intrinsic part of OpenCorporates”.<sup>67</sup>

OpenCorporates does not want to rely on grant funding as it is “fickle, and rarely long-term”.<sup>68</sup> This is partly due to the time and costs of repeatedly applying for grants: “getting grants is not easy and we don’t assume it’s going to happen.” In our interview they indicated that they don’t want to “base [their] sustainability and survival” on this instability. However, the UK Biobank is one data institution that, due to the nature of their work, has been able to achieve sustainability while existing solely on grant funding. As a long-term research project, grant funding may be appropriate as the costs are unlikely to change and growth isn’t an ambition.

Other data institutions also demonstrate how important maintaining a balanced revenue model can be to achieving sustainability in ways that don’t conflict with the core mission of the institution.

A good example is that of the Israel Meteorological Service (IMS). A case study<sup>69</sup> reviewing the migration of IMS from a model of charging for access to data to a purely open data model highlights a range of tensions. For example, the organisation incurred additional costs from supporting its commercial services that reduced its investment in delivering on its broader public good mission. According to that case study, its commercial model led to unfair competition in the marketplace and drove users to lower quality data.

HESA described to us how its business model has adjusted over time. It wanted to move away from commercialisation of data resources towards commercialisation based on expertise, such as data analysis and data visualisation. Its board of

---

<sup>67</sup> OpenCorporates (2017), ‘A corporate structure for the public benefit and the long term’, <https://blog.opencorporates.com/2017/07/06/a-corporate-structure-for-the-public-benefit-and-the-long-term/>

<sup>68</sup> OpenCorporates (2017), ‘A corporate structure for the public benefit and the long term’, <https://blog.opencorporates.com/2017/07/06/a-corporate-structure-for-the-public-benefit-and-the-long-term/>

<sup>69</sup> World Meteorological Organization (2016), ‘Case Study: Free data access – the experience of the Israel Meteorological Service’, <https://public.wmo.int/en/resources/meteoworld/case-study-free-data-access-%E2%80%93-experience-of-israel-meteorological-service>

directors have always been cautious in ensuring that HESA's commercial activities are not in conflict (or perceived by stakeholders to be in conflict) with HESA's core functions. This has meant that potential revenue streams, such as being paid to facilitate direct marketing by commercial third parties, have been deemed unacceptable by HESA.

OpenStreetMap has a number of corporate sponsors<sup>70</sup> that provide funding and are increasingly involved in the project by contributing data, open source code and other resources.<sup>71</sup> These contributions help make the project more sustainable, but have raised concerns from the existing community about the ease with which those organisations could shape the future of the project.

## The ever-evolving relationship between a data institution and its community

The relationship between data institutions and the communities that surround and support them is one of mutual interaction and influence. A change or evolution of either of these is bound to have an impact on, or bring about changes in, the other. Over the life of a data institution, this changing relationship can present a number of challenges.

For instance, as a data institution grows, the community around it is likely to grow as well. Yet as a community grows, the needs, motives and expectations of that community are almost certain to evolve. This will necessarily force the data institution to reassess not only how it serves its community, but how it attempts to achieve sustainability. HESA touched on this point in response to a question about how to remain sustainable: "Update systems and processes: the data environment is changing, expectations from the public are changing, so it is needed to adapt to changing expectations of stakeholders and customers".

Not surprisingly, different data institutions have sought to adapt in different ways. Some, like OpenCorporates, have sought to introduce commercial products that meet the needs of specific parties in their ecosystem. These products need to remain useful, of course, so as circumstances and the needs of these parties change, the commercial products must change as well.

In seeking to evolve, scale and adapt to changing circumstances, however, it is important that data institutions assess whether they will be changing in ways that will upset or alienate members of their current community, clientele or ecosystem.

For instance, as a data institution grows, it might seek to add a commercial offering to remain sustainable, but in doing so, might disappoint or draw the ire of members of their ecosystem who interpret that change as going against the initial 'open' ethos of the institution. npm Inc. serves as a useful case in point. The package manager was created in 2009 as an open-source project aimed at helping JavaScript developers share packages of code, but made the decision to transition to become a private company, npm Inc, in 2014. The company has since found itself at the centre of a number of controversies wherein questions have been raised about whether its

---

<sup>70</sup> OpenStreetMap Foundation, 'Corporate Members', [https://wiki.osmfoundation.org/wiki/Corporate\\_Members](https://wiki.osmfoundation.org/wiki/Corporate_Members)

<sup>71</sup> Open Data Institute (2018), 'How Facebook, Apple and Microsoft are contributing to an openly licensed map of the world', <https://theodi.org/article/how-are-facebook-apple-and-microsoft-contributing-to-openstreetmap/>

commitment is to commercial companies or the open source community. Some in the community contend that a for-profit company should not be allowed to steward such an important piece of infrastructure and have begun working to create an open source service to replace npm.<sup>72</sup>

Similarly, as a data institution grows and evolves, it may attempt to evolve its business model (possibly by offering new services or by welcoming new data contributors, users or decision makers to the ecosystem), but by doing so, may disconcert the original members of the community who interpret these changes as disrupting the original balance of the system.

Take the example of a data institution that has been set up to steward data on behalf of data contributors and exclusively share analyses and insights back to those data contributors. At some point in its lifecycle, that data institution may receive requests from non-contributors who want access to the data it stewards, particularly if the data would have value for parties beyond the direct data contributors – for example, for academic researchers. These researchers might be willing to pay for access to the data – thereby offering the institution a new source of revenue to support their sustainability – but might not have relevant or useful data to contribute. In this case, a data institution would have to weigh the benefits of a new form of revenue against any potential tensions caused by the introduction of new data users. For instance, the original data contributors may feel that providing access to non-contributors goes against the terms or spirit of the original data sharing arrangement, or may contend that providing access to non-contributors creates an asymmetrical value distribution, since non-contributors are sharing in the benefits without contributing data.

Crossref has encountered a similar tension due to an increased move to open access publishing models in its sector. This has changed perceptions of the value proposition of Crossref within the ecosystem and the ways in which it might generate revenue in future.<sup>73</sup> Crossref has recently commissioned research to help identify and explore those issues with its community.<sup>74</sup> The results clearly identify tensions across different parts of its ecosystem and an evolving role from a service specifically for commercial publishers to one of key infrastructure serving a broader community.<sup>75</sup>

As data institutions, and the communities they serve, evolve and expand, it is important that data institutions are conscious of the consequences of over-expansion. A data institution that expands to offer additional services or take on more functions within an ecosystem, may end up hurting the overall health of the ecosystem.

Some public sector data institutions, for instance, offer value-added products or consulting services alongside the datasets they steward. These are valuable and worthwhile products and services, but a 2006 study by the Office of Fair Trading found that when public sector information holders offer these types of products and services, they run the risk of competing with the people and organisations within that ecosystem who also offer value-added products or consulting services.<sup>76</sup>

---

<sup>72</sup> The Register (2019), 'Hey, NPM. How do you like your Bogensberger? He's, well, done: CEO Bryan ejects from biz', [https://www.theregister.co.uk/2019/09/20/npm\\_ceo\\_bryan\\_bogensbergers\\_off/](https://www.theregister.co.uk/2019/09/20/npm_ceo_bryan_bogensbergers_off/)

<sup>73</sup> Brand A (2019), 'Guest Post — Crossref at a Crossroads: All Roads Lead to Crossref', <https://scholarlykitchen.sspnet.org/2019/10/22/crossref-at-a-crossroads-all-roads-lead-to-crossref/>

<sup>74</sup> Pentz E (2020), 'A turning point is a time for reflection', <https://www.crossref.org/blog/a-turning-point-is-a-time-for-reflection/>

<sup>75</sup> Shift Learning (2019), 'Crossref Value and Benefits', [https://docs.google.com/presentation/d/1RsqtHssBkaFNphdWoq20\\_ewruYP04n8j\\_dYB9wvphM/edit#slide=id.g65af51c04a\\_1\\_807](https://docs.google.com/presentation/d/1RsqtHssBkaFNphdWoq20_ewruYP04n8j_dYB9wvphM/edit#slide=id.g65af51c04a_1_807)

<sup>76</sup> DotEcon (2015), 'Independent evaluation of the OFT's 2006 market study into the Commercial Use of Public Information',

Over-expansion by a data institution could also lead to an ecosystem that becomes overly centralised with too many functions running through a single data institution. That institution might not be able to satisfy the needs of a broader community as well as one that adopts a more limited role but enables a more diverse and distributed ecosystem. In the end, this might limit the data institution's ability to satisfy its purpose. Mitigating this centralising effect can involve: providing a variety of ways to access data, to allow alternative and competing approaches; maintaining a separation between the institution's primary purpose and commercial operations; and early consultation with data users to gather feedback and to give time for the wider ecosystem to evolve and adapt as the institution itself changes.

At the same time, it is also true that many data institutions support value creation beyond the data users or decision makers found in their immediate ecosystems. These competing interests represent a tension that data institutions will need to be aware of as they work to balance their own sustainability and growth with the sustainability and growth of their community.

The tensions outlined in this section help reinforce the point, mentioned above, that when exploring options for evolving or expanding, data institutions would be well advised to engage with their community early on, to ascertain whether plans are sustainable, whether it would negatively impact the rest of the community, or whether it would be seen as going against the existing agreement with the community.

For further discussion on how expectations and agreements between parties in an ecosystem can change – and how those changes can influence not only the sustainability of a data institution but also its perceived trustworthiness – see our report on 'Designing trustworthy data institutions'.<sup>77</sup>

---

<https://www.gov.uk/government/publications/commercial-use-of-public-information-evaluation-of-oft-market-study>

<sup>77</sup> Open Data Institute (2020), 'Trustworthy data institutions', <https://theodi.org/project/rd-trustworthy-data-institutions/>

# Conclusion and recommendations

This report summarises an initial exploratory study to help understand the factors that contribute to the creation of sustainable data institutions. To conclude, we identify some areas for further research and provide some initial recommendations for those currently working on scoping, designing and running data institutions.

## Areas for further research

We plan to apply the framework introduced in this report to support further analysis of the business and revenue models adopted by data institutions. This will allow us to expand our initial investigation and explore additional examples across a range of sectors.

In particular, there are several areas which merit further work. These include:

- Developing a more comprehensive review of the business and revenue models adopted by different types of data institutions across sectors, and in particular exploring some failed or retired institutions.
- Further work to understand the tension between different revenue streams and the goals of data institutions, with a view towards providing recommendations of specific revenue models for different types of data institutions operating in different contexts and communities.
- Exploring how the sector, domain and context of a data institution affects the available financing and funding opportunities.
- Understanding the role of funders in supporting the growth and sustainability of data institutions, particularly in the transition from scoping and co-design, to launch and operation,
- Further research on the role of governance and community in the sustainability of data institutions, to explore the interplay between this research on sustainability and the concurrent work on trustworthiness.<sup>78</sup>
- Conducting economic assessments on the value of data institutions to help demonstrate their value and impact, to support organisations applying for financing and funding.
- Exploring the operational, evaluation, re-design and retirement costs associated with different types of data institutions, to help support financial planning.
- Developing and testing guidance and tools, for example a canvas or financial planning tool, to help support data institutions develop business and revenue models and estimate their costs and investment requirements.

We welcome collaboration with other organisations working in this area.

---

<sup>78</sup> Open Data Institute (2020), 'Trustworthy data institutions', <https://theodi.org/project/rd-trustworthy-data-institutions>

# Suggestions for those scoping, designing and running data institutions

While our exploratory research was not intended to produce a comprehensive survey of the many data institutions that exist across sectors, through our desk research and interviews we have already identified some common issues and challenges in the design of sustainable business and revenue models.

Based on those insights, we offer some initial suggestions for those currently scoping, designing and running data institutions, to help them navigate the complex decisions needed to reach sustainability.

1. **Ensure your revenue model aligns with your organisational goals**
  - **Develop a clear understanding of the data ecosystem that your institution will be creating and supporting**, and the ways in which value will be created and shared.
  - We have highlighted a number of tensions that can arise between specific sources of revenue and the broader goals of an institution. It is important to **consider how adopting specific revenue streams will impact the operational focus, cost structure and ecosystem of your institution**. Fundamentally, the revenue model should be aligned with the goals of your institution and not run contrary to them.
2. **Understand how your revenue sources will change during your institution's lifecycle**
  - **Recognise the difference between the financing needed to scope, design and launch a data institution, and your ongoing sources of revenue**. The organisations and individuals providing you with revenue and their needs and expectations will be different at different stages in your institution's lifecycle.
  - **Develop a clear plan for your future revenue model to help unlock early financing** and support from other organisations.
  - **Plan to evolve your revenue model over time** to adapt to changes in circumstances and the needs of your ecosystem, for example by regularly evaluating this aspect of your operation. This will help to ensure that you are able to remain relevant and sustainable.
3. **Consider both financial and non-financial aspects of sustainability**
  - **Your data institution should have a governance structure that aligns with your goal for sustainability**. This might include adopting specific legal forms.
  - **Data institutions should invest in maintaining and growing their community** and ensuring that their product or service remains useful to them. If your data institution relies on the community for data, ensure that the process is as seamless as possible.
4. **Identify and mitigate future risks**
  - **Ensure that your revenue model is resilient** and protected from shocks, such as the loss of a specific grant.
  - **Ensure you have a contingency plan if you are not able to achieve sustainability**. If your data institution is no longer able to operate, then you need to have planned for what happens to data, code and other assets. Making these open, where possible, will allow others to build on your work.



## 5. Learn from others

- Across sectors there are many examples of long-standing data institutions that can provide a model for success. **Engage with and learn from those organisations** to help you navigate some of the complex choices you are faced with.
- **Work openly to share your successes and failures.** It is helpful for data institutions to share the decisions they have made, the challenges they have faced and to be open about their business and revenue models. This report relied significantly upon blog posts and reports that existing organisations published about their journeys.

We plan to apply these recommendations as we continue to work with a range of organisations working to establish new data institutions. As our research and practical work continues, we plan to revisit, revise and expand these recommendations.

# Methodology

To inform this exploratory research project, we conducted broad desk research into concepts of business and revenue models and sustainability relevant to the context of data institutions. We then defined a set of criteria to select different types of data institutions we would look at in more detail. We conducted desk research about these institutions and, where possible, interviewed representatives of the data institutions.

The brief timeframe of this research resulted in several limitations, including:

1. Limited number of interviews.
2. Potential sector bias. Our desk research has been informed significantly by our prior knowledge of specific institutions.
3. Focus on open infrastructure. Linked to the above, the revenue and business models we have encountered might be specific to data institutions operating an open infrastructure.

## Research questions

**Main research question:** What sustainable business models can be adopted by data institutions, and what are their respective benefits, costs and risks?

### Research questions:

1. What revenue sources have contributed to the creation of data institutions?
2. What revenue models have been adopted by successful data institutions?
3. How does the revenue model evolve throughout the lifecycle of data institutions, eg during scoping and design, launch and early stages of operating, and later stages of operating?
4. How might different business models align to the goals of different types of data institutions, for example, by supporting or conflicting with their core goals and mission?

## Criteria to select institutions

In our research, we wanted to understand the experience of a diverse set of organisations. We decided to focus on organisations that met one or several of the following criteria:

1. Type of data institution:
  - a. Organisations that steward data on behalf of a community
  - b. Organisations that steward digital resources or a platform on behalf of a community
  - c. Organisations that steward a physical resource on behalf of a community or set of stakeholders
2. Life stage of the data institution:
  - a. Institutions in **scoping or co-design stage**, so we can learn how they're thinking about covering the costs of their launch and operation
  - b. Institutions that are **operational** but relatively young, so we can learn about early experiences

- c. **Well-established** or ‘successful’ institutions, so we can learn what makes them successfully sustainable
  - d. Organisations that tried to set up a data institution but could not overcome the challenges, so that we can learn about specifics of those challenges
3. Main source of revenue:
    - a. At least one data institution that is funded by a large funding body

## Desk research

For each of the organisations studied, we gathered information related to:

- General information about the organisation – core goals or mission
- Establishment
- Type of data institution – the functions it fulfils or services it provides
- Revenue model – current, past and anticipated future revenue sources
- Alignment of the organisation’s business model with its core goals

## Interviews

We complemented the information gathered through desk research with interviews with representatives of data institutions. We interviewed representatives from five organisations covering some of the defined criteria:

- **HESA** – as a well established (operational) organisation within the educational sector
- **MusicBrainz** – as a well established (operational) and community-maintained project
- **OpenCorporates** – as a well established (operational) organisation within the business information field
- **ROR** – as a relatively young organisation offering open identifiers for every research organisation in the world
- **HiLo** – as a relatively young organisation in the maritime industry

The questions asked during the one-hour interviews covered themes related to:

- Revenue streams
- What sustainability looks like
- Alignment of business models with the data institution’s core goals
- Mechanisms for ensuring trustworthiness