Data institutions: reducing costs and improving sustainability

Open Data Institute

Contents

About

This report has been researched and produced by the Open Data Institute, and published in September 2020. The lead authors are Diána Szász, Ben Snaith and Joe Massey, with research support from Sonia Duarte. Leigh Dodds, Jeni Tennison and Jack Hardinges provided additional contributions. To share feedback by email or would like to get in touch, contact the Sustainable Data Access project lead Diána Szász.

To share feedback in the comments, highlight the relevant piece of text and click the 'Add a comment' icon on the right-hand side of the page.

How can it be improved? We welcome suggestions from the community in the comments.

Executive summary

In spring 2020, the Open Data Institute (ODI) published a report[1] summarising the results of an initial investigation into the sustainability of data institutions.

Data institutions are organisations whose purpose involves stewarding data – that is, collecting, maintaining and sharing data – on behalf of others, often towards public, educational or charitable aims. The sustainability of these organisations is vital to ensure that they can continue to deliver on their mission, and drive social, economic and environmental impact. But we found that achieving financial sustainability is often challenging for data institutions as some of the possible revenue models can hinder their goal of increasing access to data.

This report is the result of our continued research into the business models of data institutions and the surrounding funding landscape. We have further explored the revenue models used by data institutions by looking at a broader set of data institutions through our desk research. We then focused on examining the cost structures of a small number of data institutions and the approaches they have taken to lower costs so that they do not have to generate as much revenue. We looked at data institutions that are relatively young as well as others that have been operational for a long time. Finally, we explored what drives public and philanthropic funders' decisions to fund data institutions.

We found further validation that a mix of revenue sources, while not always feasible or indeed necessary, seems to be common among 'successful' data institutions. We also found that a mixed revenue model can help distribute the costs of running the data institution more equitably among the stakeholders that gain value from the data institution's work. Some revenue streams, such as membership and subscription fees, seem to be used only when data institutions play particular roles while others, such as grants, appear to be common across different roles. Understanding the value of data institutions for key actors in the ecosystem that surrounds them – for data contributors, data users, decision makers and funders – may help to identify the revenue models available to data institutions.

Continuing to draw on the Business Model Canvas developed by Alexander Osterwalder, we examined the key activities and key resources that drive data institutions' costs. Staff costs appear to make up the majority of costs for data institutions, driven by activities such as technical infrastructure development and maintenance; data collection, curation and maintenance; strategy development; governance; and community engagement.

The strategies to reduce costs come either from the choice of services used – for example, choice of data hosting service – or from the design of the data institution and its services. Most strategies to reduce costs come with risks that need to be balanced. Some strategies – say, adopting a crowd-sourcing model for data – may reduce costs in some areas (for example data collection) but raise it in other

---

[1] Open Data Institute (2020), 'Designing sustainable data institutions', https://theodi.org/article/designing-sustainable-data-institutions-paper/

areas (for example in community engagement).

A data institution can make decisions about when it will incur certain costs. It may be able to decide when it would like to scale its services and incur related costs, or when it would like to develop its own platform if using an existing one is suboptimal. Considering these choice points can help to better plan and manage specific costs.

Strategies to lower costs do not eliminate the need to identify long-term, sustainable revenue and funding models. If data institutions do not have access to suitable revenue and funding streams, they cannot fulfil their role of creating and maintaining essential infrastructure for increasing access to data for decision-makers and for society more broadly.

We have previously seen that grant funding is a common source of income for data institutions, especially at the initial stages of their lifecycle. It is likely that some data institutions, such as the long-term health research project Biobank, will continue to rely on public or philanthropic funding throughout their operation. We found that some funders are conscious of the need for ongoing funding and, recognising that the data institution delivers value aligned to their philanthropic goals, are able to provide long-term funding.

Our research has identified broader forms of support that can help data institutions' growth and sustainability. These can come from actors other than funders. Support can come from firms providing specific services such as legal, accounting or technology services. It can also come from organisations that aim to support data infrastructure either at the system level or focused on specific sectors of the economy. The scholarly research and communication space, for instance, has given rise to several initiatives aimed at supporting data institutions with guidance and resources as well as supporting funders' decisions through research and recommendations on which institutions should be considered for funding support in the sector[2]. Other sectors could explore these efforts and see if there are models to learn from.

We have identified recommendations for data institutions, funders and other actors who are interested in supporting the creation and sustainability of data institutions.

Recommendations

Recommendations for those scoping, designing and running data institutions

- Be transparent about the full costs involved in running data institutions
- Understand the consequences of cost management strategies
- Understand and communicate the value you offer to key actors in your data ecosystem
- Develop and iterate on an evaluation plan to demonstrate impact

---

[2] Educopia Institute (2019), 'Why Are So Many Scholarly Communication Infrastructure Providers Running a Red Queen's Race?', https://educopia.org/red-queens-race/

Recommendations for those funding or supporting data institutions

- Identify opportunities to align your approach with other funders
- Support data institutions to grow beyond their core activities
- Clearly articulate your approach to funding data institutions

Introduction

Access to data is crucial for enabling innovation. We need data to be stewarded – collected, maintained and shared – responsibly to ensure that it's available to those who need it in ways that retain trust. If we don't act responsibly, then we are unlikely to realise the full benefits that data could have, and could expose people and communities to harmful impacts, such as exclusion or profiling.

At the ODI we are exploring the role of data institutions – organisations whose purpose involves stewarding data on behalf of others, often with public, educational or charitable aims.[3] The economic and social value of existing data institutions is evident in examples such as the below.

- The **Open Apparel Registry**, an open-source map and database of global apparel facilities, helps inform efforts to improve working conditions in factories and enables brands to manage their own supply chains more effectively.
- The **Creative Diversity Network** stewards and publishes data about the diversity efforts of UK broadcasters.
- The **UK Biobank** aims to improve the prevention, diagnosis and treatment of serious and life-threatening illnesses. It stewards the biological samples and data of 500,000 volunteer participants, and makes the data available for research and development.
- The **Carbon Disclosure Project (CDP)** promotes a sustainable economy by helping companies and cities disclose the environmental impact of major corporations.

We're also interested in data access initiatives[4], which we describe as initiatives or programmes that:

- have a clear challenge, in the form of a specific social, environmental or economic problem that is the focus for collaboration
- involve multiple stakeholders that are actively working together to solve the problem
- include a strong focus on collecting, using and sharing data as part of their work.

OpenActive[5], a community-led initiative to help people in England get active, hosted at the ODI, is an example of a data access initiative. Another is the Emergent Alliance[6], whose aim is to build economic resilience and support recovery from COVID-19.

Over time we think that data access initiatives might lead to new data institutions being created, or existing institutions will have a strengthened role as stewards of

---

[3] The Open Data Institute (2020), 'What do we mean by data institutions?', https://theodi.org/article/what-do-we-mean-by-data-institutions/
[4] The Open Data Institute (2020), 'What do we mean by data access initiatives?', https://theodi.org/article/what-do-we-mean-by-data-access-initiatives/
[5] Open Active (n.d.), https://www.openactive.io/
[6] Emergent Alliance (n.d.), https://emergentalliance.org/

data infrastructure. This will be needed to embed and secure the change necessary to solve particular challenges.

Setting up and running data access initiatives and data institutions can be costly. If an organisation wants to ensure it can provide continued stewardship of data, it needs to be financially sustainable.

The costs of data access initiatives that do not take an institutional form may be met in various ways. An existing organisation may be paid to run it, as has been the case with OpenActive. A group of stakeholders may contribute in-kind and/or financial resources as with the Emergent Alliance[7], a nonprofit community made up of corporates, individuals, non-governmental organisations (NGOs) and governments. Its aim is to better inform economic decision making, in particular to aid societal recovery post Covid-19. Often a mix of these funding models might be used.

Setting up a dedicated organisation – a data institution – might arise from the need to be able to receive direct funding, as has been the case with Open Banking Limited[8], or the need for the initiative to govern itself as a standalone entity separate from organisations that may have previously funded or run it.

Our focus in this report is on the institutional forms of data stewardship – organisations which require guaranteed resources over the longterm to fulfil their purpose.

This is challenging. Traditionally, generating revenue has tended to rely on charging directly for access to data. However, revenue models that include charging for access to data entail restricting access to it, for example through licensing conditions on its use or reuse, or through price. This means only organisations with sufficient resources can access and use the data. This in turn reduces the economic and social value that data institutions aim to deliver by increasing access to data. We have been looking at ways data institutions can become financially sustainable without using paid for licences, or otherwise restricting access.

We explored some of the challenges around finding sustainable revenue or funding sources in our first report, *Designing sustainable data institutions*[9], published in April 2020. We proposed a framework for thinking about the sustainability of data institutions that involves three different elements.

- The **role** that the data institution plays in its data ecosystem, which directly impacts its options for 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 and can acquire, and how those revenue sources may evolve over time as the institution and its ecosystem mature.
- The broader **business model** of the data institution, that is, the structure used to create, deliver, capture and distribute value. This in turn informs the choice of revenue model that best aligns with its purpose.

---

[7] Emergent Alliance (n.d.), https://emergentalliance.org/?page_id=1517
[8] Open Banking (n.d.), https://www.openbanking.org.uk/
[9] Open Data Institute (2020), 'Designing sustainable data institutions', https://docs.google.com/document/d/1k0tcVStnXb7jpS7P9hhFQ2d3Spu45J5GkVtxpUdXwCg/edit

The previous report explored the following areas, which are not elaborated on in this report.

- The role of community support and good governance as facets of sustainability.
- Common legal forms of data institutions and how they affect the types of revenue sources available.
- The tensions that arise with some revenue models between the purpose of a data institution and its need to generate revenue.

This report builds on that research and complements it with research on:

- what costs data institutions need to cover at different stages of their lifecycle
- what strategies can be used to manage them
- an exploration of what makes public and philanthropic funders interested in funding data institutions
- how funding practices align with data institutions' needs.

For more details on how we conducted this research, see the Methodology section.

This research and development project is part of the ODI's broader work on data access initiatives and data institutions. This includes our data institutions programme[10] that aims to make significant advances in the understanding of data institutions, and the roles they can play, as well as supporting organisations to create new data institutions or to recognise their role as a data institution. The ODI's research and development programme also includes a project that looks at mechanisms such as certification to formalise trusted relationships between data stewards and data users[11]; and a project that supports sectors to address specific challenges through data sharing.[12]

This research will inform the next phase of our project, which involves developing practical tools for supporting data institutions at different stages of their lifecycle.

---

[10] Open Data Institute (n.d.), 'Project: Data Institutions',
https://theodi.org/project/rd-data-institutions/
[11] Open Data Institute (n.d.), 'Project: Building trusts through audit and certification',
https://theodi.org/project/building-trust-through-audit-and-certification/
[12] Open Data Institute (n.d.), 'Project: Data Infrastructure for Common Challenges',
https://theodi.org/project/data-infrastructure-for-common-challenges/

New insights on revenue generation

In our previous report, *Designing sustainable data institutions*[13], we introduced a framework which considers the role of the data institution in its ecosystem, its lifecycle stage and its broader business model, that is, how it creates, delivers and captures value.

In this research we looked at data institutions across the six roles we have found they play[14] and identified the revenue models adopted by data institutions that fulfil these roles.

Any given data institution may not fall precisely into one of these categories, and instead may perform a number of the roles.

Role	Examples	Revenue sources observed
Data institutions that create and maintain open databases that other people and organisations can contribute to and that anyone can access, use and share	Wikimedia Foundation, OpenStreetMap, Musicbrainz, Open Apparel Registry	Donations, sponsorships, fees or donations for commercial use of data, donations for speaking at conferences, consultancy fees, membership fees, grants
Data institutions that aggregate existing open or public data and make it available for anyone to access, use and share	OpenCorporates, imin, 360Giving, OpenOwnership	Fees for commercial use of data, premium data services, grants
Data institutions that assign and maintain identifiers for a sector or field	CrossRef, ORCID, ROR, other registration agencies	Membership and subscription fees, service fees, content (data) registration fees, grants and donations (in early stages)
Data institutions that hold data on behalf of an organisation or person, or group of them, and process,	Data institutions that exist with some degree of public mandate: HESA, Oil and Gas Authority, Companies House and other business registers, Office for	Government funding, mandatory subscriptions, grants, service fees

---

[13] Open Data Institute (2020), 'Designing sustainable data institutions'
https://theodi.org/article/designing-sustainable-data-institutions-paper/
[14] The Open Data Institute (n.d.), 'Data Institutions',
https://theodi.org/project/rd-data-institutions/

publish or share it with others (such as for research)	[National Statistics](#), [Eurostat](#) and other statistical agencies, [Data Communications Company](#)	
	Data institutions that do not have a public mandate: [UK Biobank](#), [Health Data Research Hubs](#), [Creative Diversity Network](#), [Land Portal](#), [UK Data Service](#), [MEDIN](#)	Grants, application fees, access charges, membership fees
Data institutions that combine or link data from different sources, and provide insights and other services exclusively to those that have contributed data	[HiLo](#), [Data Trust for Open Access Ebook Usage](#), [FarmBench](#), [LG Inform](#)	Subscription and membership fees (incubation in early stages). If organisations in this category are private companies, raising investment from the market may be a feasible option.
Data institutions that support people and communities to steward data themselves for particular purposes or causes	[OpenHumans](#), [Health Bank](#)	Private foundation grants, commercial partnerships, academic partnerships, usage fees

Table 1: Examples of data institutions and their revenue sources across a range of roles which we have found data institutions play.

Sources of revenue

In *Designing sustainable data institutions*[15], we identified a range of revenue models that data institutions rely on at different stages of their lifecycle. These include earned revenue, such as recurring membership and subscription fees, and other models such as grants, and financial and in-kind donations.

Our current research confirmed that organisations can bring in revenue by selling added-value services where the data user pays for access to a more reliable service than the free public service or API.

For example, the Metadata Plus service of Crossref[16], a membership organisation that assigns and maintains identifiers for research outputs, offers extra guarantees and features over the public service. To make such a paid-for service attractive to data users, Crossref told us that the most important thing it can do, even though it requires significant effort, is to improve the value of the data by enhancing its quality and coverage and/or by supplementing or combining it with additional data.

In our research, we found that membership rather than subscription fees might offer additional benefits for some data institutions. Crossref is committed to the idea of membership. Membership typically encourages a more active involvement and is used as a way of ensuring that people contributing data are the ones governing access to it. Members may receive a range of benefits as part of their

---

[15] Open Data Institute (2020), 'Designing sustainable data institutions'
[16] Crossref (n.d.), 'Fees', https://www.crossref.org/fees/#metadata-plus-subscriber-fees

membership, and also may be expected to contribute in both monetary and non-monetary ways, for example with their time, energy and expertise.[17] Subscriptions tend to be a simpler, transactional exchange of services for a fee.

Crossref told us that this focus on membership emphasises that the data stewardship is a collective endeavour. According to Crossref, it is the best way to 'ensure that they do not become an organisation that develops its own agenda'.

Another long-term strategy that has been used by some data institutions is endowments. Endowments are long-term donations of money that can be used by nonprofits for a specified purpose. They have been typically used in education, but organisations such as Wikimedia are also using them.[18] Building or fundraising for an endowment can help with an organisation's long-term sustainability, and smaller organisations such as Permanent, a nonprofit organisation whose mission is the preservation of the digital legacy of all people, has managed to convert donations into an endowment fund.[19]

Mixed revenue models

In our previous research we established that data institutions tend to cover their costs through two or three different revenue streams, often with one dominant revenue source. Other research, such as the recent 2019 Census of Scholarly Communication Infrastructure Providers[20] that mapped data institutions in the scholarly research and communication space, also notes that 'a healthy diversity in revenue sources likely accounts for some of the success of [these organisations]'.

In this research we found that, for some data institutions, a mix of revenue sources, beyond managing risk, can also help distribute the costs of providing data infrastructure more equitably among different stakeholders. Crossref is looking at diffusing the costs among its broader constituency to reflect who is using the data and therefore has a stake in ensuring its quality.

We have also seen that for some organisations, diversified revenue streams may not be feasible, or indeed necessary.

When diversified revenue streams may not work

- Existing funding may not allow an organisation to spend time or expense on fundraising or developing other revenue streams.
- There are benefits to a single source of funding, or one dominant source of funding. For example, it means that a direct correlation can be drawn between funding and impact. It is also simpler to build and maintain relationships with, and report to. a single stakeholder.

---

[17] INMA (2019), 'Conference Blog: Membership versus subscription models: Which one is better?',
https://www.inma.org/blogs/conference/post.cfm/membership-versus-subscription-models-which-one-is-better
[18] Wikimedia (n.d.), 'Wikimedia Endowment',
https://meta.wikimedia.org/wiki/Wikimedia_Endowment
[19] Permanent.org (n.d.), 'How We Use Donations',
https://www.permanent.org/what-we-do/our-endowment/how-we-use-donations/
[20] Educopia Institute, (2019), 'Mapping the Scholarly Communication Landscape 2019 Census',
https://educopia.org/wp-content/uploads/2019/06/Census2019_EducopiaPublications.pdf

- Some data institutions may be able to rely on revenue coming from a single group within their ecosystem as long as that group continues to see the value the institution provides. This might be, for example, a group of committed funders or members who are mandated by law to pay membership fees.
- Some data institutions, for example public bodies, may not be able to accept funding from other sources.

Revenue models across different roles

While we have not observed definitive patterns, we have seen that some revenue models are only used when data institutions play particular roles, while others are common across different roles. For example, data institutions that maintain open databases such as OpenStreetMap and Musicbrainz often tend to receive at least part of their revenue from donations, sponsorships and fees for commercial use of data. Data institutions that combine or link data from different sources, and provide insights and other services exclusively to those that have contributed data, such as HiLo or Farmbench, seem to often use subscription and membership fees. Public and philanthropic grants, on the other hand, seem to be a common source of funding across most roles.

We think that exploring the value of data institutions for key actors in their ecosystem can help us understand the revenue models that a data institution might be able to use. In our previous report, *Designing sustainable data institutions*[21]*,* we introduced the ecosystem diagram below, outlining the different actors that contribute to and/or gain value from the presence of a data institution: data contributors, data users, decision makers (those able to make better decisions informed by the data analysed by their staff) and funders. By understanding what problems a data institution solves for these actors, we can start to explore who would be willing to pay for its services.

---

[21] Open Data Institute (2020), 'Designing sustainable data institutions', https://theodi.org/article/designing-sustainable-data-institutions-paper/
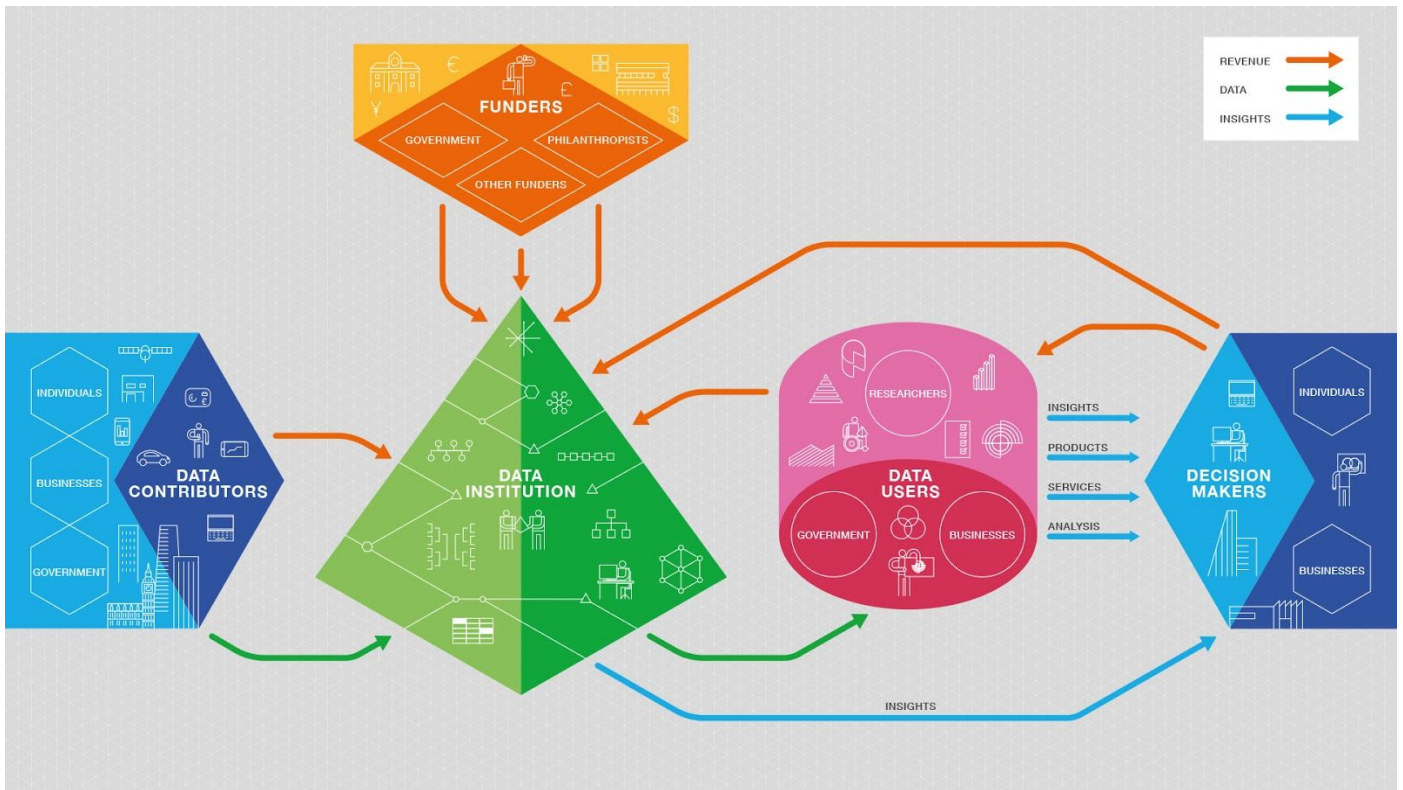
Figure 1: Common actors and flows of value and data related to data institutions

This thinking is also in line with the way the social sector tends to think about business models. Nonprofits often use adapted versions of the lean and business model canvases[22] [23]. These canvases are one-page templates for developing new or documenting existing business models. The versions adapted for nonprofits highlight the two-sided nature of the markets in which they operate, addressing the needs of both their customers/beneficiaries and their funders. In the case of data institutions, we see multi-sided markets with data institutions serving the needs of (and potentially gaining revenue from) data contributors, data users, decision makers and funders.

---

[22] Maurya, A (2018), 'Why and How to Model a Non-profit on the Lean Canvas', https://blog.leanstack.com/why-and-how-to-model-a-non-profit-on-the-lean-canvas-514e4acf1051
[23] Social Business Model Canvas (n.d.), https://www.socialbusinessmodelcanvas.com/

Costs and cost management approaches

To understand commonly occurring costs that data institutions need to cover, we used the Business Model Canvas[24] as a framework. According to the canvas, the cost of operating an organisation's business model is driven by its key activities and key resources. We asked eight data institutions across five of the six roles described above about their key activities and resources and how these shaped their costs. For more details on the research questions, see the Methodology section.

We also asked them about the approaches that data institutions have taken to manage or minimize their costs, for example in response to financial pressures.

According to the canvas, the key resources of an organisation – the most important assets needed to carry out the key activities – can be classified as follows:

- **Physical:** such as computer equipment, office supplies, buildings and facilities, servers (although over time many of these have become rented resources)
- **Intellectual property:** such as registered patents, brands, or copyrights
- **Human:** staff
- **Financial:** cashflow, working capital management and funding

To understand the key activities of data institutions, we referred to the ODI's Data Trust report[25] that described activities that can take place in different stages of the lifecycle of a data trust, a type of data institution: scoping, co-design, launch, operate, evaluate, retire. In our interviews with data institutions, we focused on the main activities driving data institutions' costs during each of these stages.

---

[24] Strategyzer (n.d.), 'The Business Model Canvas', https://www.strategyzer.com/canvas/business-model-canvas
[25] Open Data Institute (2020), 'Designing trustworthy data institutions', https://theodi.org/article/designing-trustworthy-data-institutions-report/
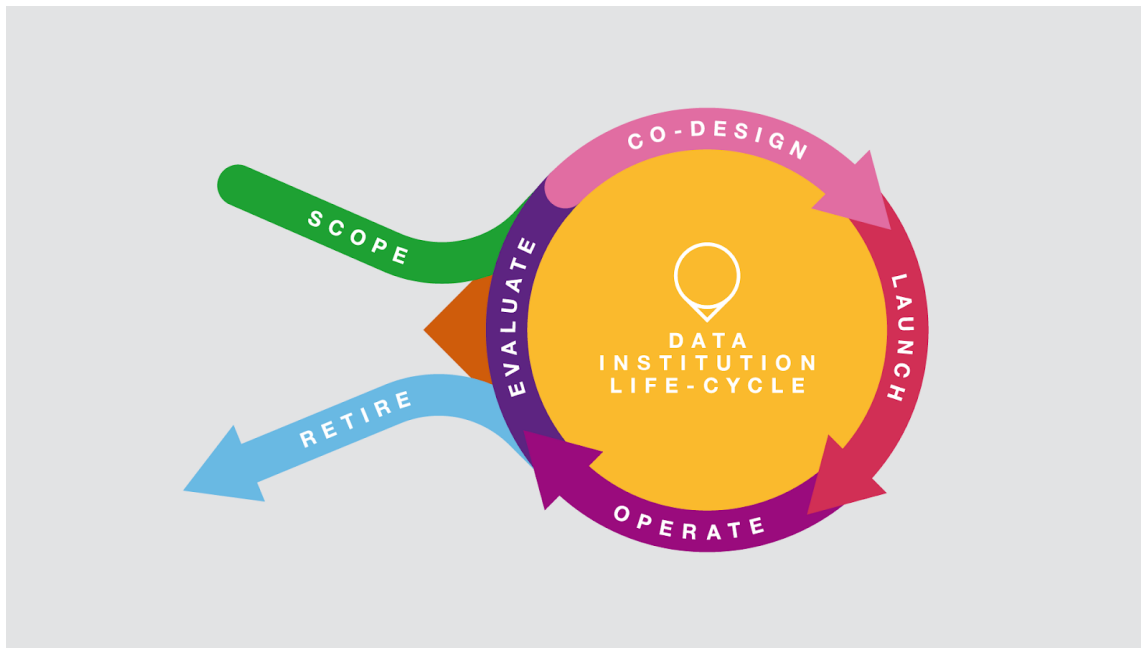
Figure 2: The data institution lifecycle

Key activities in different stages of a data institution's lifecycle

Some activities grow or shrink, are born or disappear, at different stages of an organisation's development.

For instance, initial scoping activities may include the development of a proof of concept and extensive talks with relevant stakeholders. In the co-design stage, the focus often shifts to technical platform development, user testing, communications, business development and marketing.

As data institutions then become operational, development costs may decrease, although technical infrastructure maintenance and the development of new features may remain a significant cost. Continuous outreach and support to the community becomes a key priority as data institutions launch.

Some activities, while present throughout several stages of a data institution's lifecycle, tend to become more significant and cost more as data institutions move from scoping to co-design and launch and then into their operational stage.

In many startups, the costs of support operations such as legal and accounting need to be covered as organisations become more established. As the organisation expands, so too will the volume of legal and accountancy activity needed. For example, one organisation noted that now that they're 'beginning to roll out paid-for services, [they] need to establish formal accounting processes. [They] will also need to fulfil certain financial reporting obligations as a nonprofit organisation and have sought support with this.'

Another cost category that may become significant is marketing and business development more broadly, that is the development of an organisation's revenue and funding streams. This is especially true if the data institution focuses on generating its own commercial revenue or diversifying its government or

philanthropic funding sources (which each have their own costs associated with fundraising activity).

Data institutions, in some cases, can choose at what point they want to incur certain costs. This may be the case, for instance, for investment in long-term or capital assets such as data platforms which require significant investment of cash up front. Other costs might be present and not possible to defer. Data institutions, like other organisations facing capital investment decisions, can determine their capital needs, assess their ability to invest, and decide which capital expenditures are the best use of their resources.[26]

In some situations it may be preferable to develop a simple initial product or service more cheaply in order to launch faster, even if this may lead to greater costs down the line. For example, a data aggregator may choose to begin their project using a platform that already exists and customise it for their purpose. This could lead to an inferior product, but may allow the organisation to go live and/or to start to generate revenue sooner. Such a decision has to consider that building a purpose-built platform at a later stage may be a more complicated process than if it was done originally. In other situations, however, this slower investment may not be an available strategy for the organisation. For instance, if the data institution holds personal data so that it can be shared with others, it will need to reach higher levels of technology and security solutions quicker to build and maintain trust among data contributors. Similarly, there may be particular needs from data users, decision-makers or funders that a data institution has to take into account when making investment decisions.

Most expensive resources and activities

Resources

Staff costs were the largest expense across the data institutions we have interviewed, which is in line with the general trend of 'human resources being particularly prominent and valued in knowledge-intensive and creative industries', perhaps due to the retention of experience and knowledge.

Managing physical assets such as **data servers** was also described as costly.

Centralised **office spaces** can be costly, and some organisations have decided to do without the cost and shift to a remote working model.

One organisation, however, discussed how they were determined to keep the overhead costs of physical assets to a minimum in order to prioritise spending on key activities such as data collection.

---

[26] Graybeal, P., Franklin, M., & Cooper, D. (2019), 'Principles of Accounting, Volume 2: Managerial Accounting. Describe Capital Investment Decisions and How They Are Applied', https://opentextbc.ca/principlesofaccountingv2openstax/chapter/describe-capital-investment-decisions-and-how-they-are-applied/

The activities that the data institutions mentioned as requiring extensive dedicated staff time and/or the time of external providers – and that therefore incur significant costs – include the following:

- **Technical infrastructure development and maintenance** such as the building and maintenance of a platform, including the implementation of new features. Having good technical developers and engineers is of paramount importance to data institutions and developer salaries or rates tend to be high compared to many other skills that an organisation might need access to.
- **Data collection**, **data curation** and **keeping the data up to date**. If a data institution operates in a sector with poor data-publishing maturity, it may need to do more work to manage the data. This could be the manual collection of data and the development of data 'scrapers' etc – for instance in the case of aggregators of publicly available or open data. Or it could mean that data contributors require significant support when uploading data. Data collection may also be costly if a data institution collects new primary data and needs to deploy new sensors or systems.
- Developing a **strategy and governance structure** to manage the activities of an institution and foster trust by the community. Developing a strategy to guide what the organisation does and doesn't do takes time and requires collaboration with different stakeholders. Designing appropriate governance mechanisms – the 'systems and processes concerned with ensuring the overall direction, effectiveness, supervision and accountability of an organisation'[27] is also a time-intensive and important activity.
- **Community and user engagement.** Running events and meetups, and offering onboarding of customers and ongoing customer support are also important and costly. This is even more challenging when a data institution is operating globally, like Crossref and ORCID, without having offices in other countries. How to staff the organisation to be able to engage the community effectively can be a big challenge.
- **Marketing and sales costs**. Where possible data institutions carry out direct marketing activities to grow their user base, to build the demand for their services. Marketing activities require planning and specific budgets, as do pursuing business development opportunities.

Although these activities are fundamental to an organisation's ability to operate successfully, it can be hard to raise public or philanthropic funding for them, as we will discuss later.

The following section looks at how data institutions are, or could be, minimising costs to allow for the limited resources to be dedicated to activities that are key to their ability to achieve their mission.

Ways to manage costs

---

[27] NVCO (n.d.), 'An introduction to governance', https://knowhow.ncvo.org.uk/governance/getting-started-in-governance/getting-started-in-governance-1

The table below summarises the main strategies we found for managing and/or minimising the costs of running data institutions. Many of these are methods commonly used by startups[28] – such as outsourcing instead of hiring or operating in a virtual rather than a physical office space; and nonprofits[29] – such as seeking in-kind donations, benefitting from nonprofit discounts, using volunteers and partnering with other organisations.

Resource or activity category	Ways to manage or minimise costs	Consequences of cost-minimisation approach
Marketing and sales costs	Request that funders and similar supporting organisations market your organisations through their networks.	This typically happens already and there are few downsides. It should be standard practice for organisations to market any organisation that they are funding or supporting in some way.
	Speak at events and panels in order to share your organisation and experiences.	Events do not always pay speakers or offer expenses, therefore it is still a cost to cover.
	If offering paid-for services, having a free tier can: ● attract users who may then pay for premium services ● attract users who do not use the service frequently enough to justify paying for it, but can benefit from the service and grow the user base.	There needs to be a clear distinction between the free and premium tier to ensure that the organisation is generating revenue from users. Some organisations may feel tempted to start commercialising more of the data gradually, which can squeeze the free tier.
Strategy and governance	Operate with minimal strategy or governance to oversee activities. Alternatives could include adopting governance structures that are run by volunteers, or leaning on pro-bono services and unremunerated advisory boards to help develop strategies.	If they operate with minimal governance and strategy, organisations run the risk of not being able to access funding from philanthropic or grant givers, who routinely consider these issues when considering funding. Alternative governance structures will require time, planning and legal advice to establish.
Hosting costs	Finding cheaper physical hosting options.	Outsourcing hosting means sacrificing a degree of control over the servers, meaning little input into maintenance or down times. This is not usually a significant challenge.
	Moving to cloud hosting.	Cloud hosting has become the norm for most

[28] BusinessLoad (2018), 'Best Ways to Minimize Startup Costs', https://www.businessload.com/startups/best-ways-to-minimize-startup-costs/
[29] AZCentral (n.d.), 'Fundraising Ideas for a Nonprofit Organization', https://yourbusiness.azcentral.com/fundraising-ideas-nonprofit-organization-4775.html

		organisations. Cloud servers may have bespoke offerings that can tie organisations into an ecosystem and then struggle to move to a different provider.
	Taking advantage of in-kind services or free credits for nonprofits/startups from a cloud provider.	There is a risk that the in-kind or free credits for hosting may be withdrawn or charged for at a future date, leading to a later increase in costs.
	Reducing the volume of data stored.	This might involve reducing the scope or scale of the product or services offered, which can affect users.
Office rent costs	Moving to smaller, cheaper and/or serviced offices.	There may be a small loss of organisational identity.
	Using office space provided by others, as in-kind support.	Working within another organisation's space may obstruct the building up of a unique organisational identity.
	Shifting to a fully remote workforce.	There will be new challenges around coordination and collaboration involved in managing a remote workforce.
Staff costs	Exploring in-kind services such as pro-bono legal support.	Pro-bono support can be withdrawn or only offered infrequently. Organisations should ensure that budget is allocated for this eventuality.
	Using contractors or external service providers, saving cost by not paying a full annual salary and overheads.	This is typical practice for legal and accountancy activities in small and medium-sized organisations. There is a risk that the organisation relies too heavily on the experience of external contractors, who are usually more costly over time. It also means that their experience and knowledge will be lost when their contract ends. Therefore finding a balance between core staff and freelance expertise is important.
	Inviting volunteers to contribute	Running a volunteer system can be unreliable, and still requires staff members to organise, facilitate and train the volunteers. It can be difficult to obtain volunteer help for all activities.

Technical development and data collection costs	Seeking voluntary/crowd-sourced contributions can save costs in the case of collaborative maintenance or membership-based data institutions.	Volunteer staffing is unpredictable and cannot necessarily be relied upon as a fixed level. Relying on community-led contributions may reduce costs for technology and data, but will require more time spent in activities such as community management, training, review and quality assurance.
	Using and developing open-source software can be less costly than using proprietary software and can help to avoid vendor lock-in. Using or building common platforms with similar organisations can reduce costs not only for individual data institutions but also for the sector as a whole.	Open-source models are public and have no clear structure, and therefore rely on contributors and maintainers. There is a risk that the platform is not able to meet the needs of all individual organisations and that at some stage a data institution may need to decide to build its own platform – at which stage it might be more complicated to do that. The common platform may or may not continue to serve the needs of various organisations and may cease to exist if it doesn't.

Table 2: Examples of how data institutions can manage costs

We discuss some of these strategies in more detail below.

Using voluntary or crowd-sourced contributions can save costs, as mentioned above, but come with challenges. The benefits have been highlighted through various examples, such as the use of crowdsourcing by membership-driven news organisations to support the freedom of information request process, 'taking some of the load off journalists, while also bringing members into the reporting process so that they value the final output more'[30]. In this case, the support provided by members did not require specialised skills. In the case of OpenHumans, a project that enables individuals to steward personal data and to connect it with research and citizen science, the types of work needed require highly skilled labour which organisations can get through motivated volunteers for free. While benefiting from such experience is a huge boon for the organisation, it does mean that technical development is done at the whim of busy engineers. If no individuals are prepared to dedicate their time to do a particular piece of development at the time it is needed, then the organisation will have to pay high fees to compensate people to do it professionally.

Crossref, in our conversations, has also mentioned how **getting the community and members involved in building and tailoring tools** can be an alternative to contracting that work out. 'We have to look at our instinct of first ways to build traditional tools as any kind of commercial service would. When we have a community that is so willing to engage and do stuff, maybe we have to do things differently.' This model worked in the beginning of the Wikimedia Foundation, which was run almost exclusively by volunteers in the first few years. Relying on

[30] mySociety (2020), 'Introducing WhatDoTheyKnow Projects', https://www.mysociety.org/2020/07/24/whatdotheyknow-projects/

volunteers, however, as highlighted in our previous report, is not always a sustainable solution. "Relying on volunteers working in their spare time is not a reliable basis for essential infrastructure."[31] An exception to this may be OpenStreetMap, which has managed to operate on a largely voluntary basis since its inception. Indeed, the voluntary nature has enabled them to develop a global infrastructure that simply would not have been possible in the same timeframe, or without a large investment of funds, if they relied on paid staff.

A general challenge in collaborative or open source development projects tends to be around the **ability to plan ahead**. As Nadia Eghbal wrote for the Ford Foundation, 'Current efforts to support digital infrastructure tend to be reactive and ad hoc. For existing projects, maintainers will benefit greatly from being able to plan for the next three to five years, not just six months to a year.'[32]

One dilemma faced by organisations looking to set up a data institution that has fundamental cost implications is **whether to create a new organisation or to launch a new initiative within an existing organisation** or group of organisations. This dilemma has been particularly prominent in the scholarly research space where new organisations have been created for assigning and maintaining the identifiers of different things. There have been calls in the sector to 'further explore the value of mergers, migrations, and other mechanisms that may provide the necessary administrative, fiscal, and social infrastructure to help support the technical development and maintenance scholarly communications resources require. Scaled, leveraged efficiencies (for example, multiple programs hosted by a single entity with shared leadership and staffing) may help to bring needed expertise while also maintaining a lower overhead.'[33]

Not launching a standalone organisation, as we saw in our previous report, can help keep costs low. The Research Organization Registry (ROR), a community-led project to develop an open, sustainable, usable, and unique identifier for every research organisation in the world, has remained a collaboration between partner organisations with most of its resources – human (team members), physical (servers) and financial (donations) – being provided as in-kind support to the project by these organisations. It may also help the organisation to be developed and grown within a safer setting. It can then be 'spun out' when ready.

The counter argument has been that keeping a new project within an existing organisation can endanger the project's success and may mean that the project cannot grow to the same extent as if it had to stand on its own two feet. For ROR, that risk has been diffused by the project being run by multiple existing organisations.

Several data institutions we spoke with – particularly the relatively young ones – mentioned that **running the organisation with minimal overhead** is a key strategy to sustain their operations. ROR highlighted that 'the Covid-19 pandemic has reinforced its existing assumptions and goals that running ROR with minimal

[31] The Scholarly Kitchen (2019), 'Building for the Long Term: Why Business Strategies are Needed for Community-Owned Infrastructure', https://scholarlykitchen.sspnet.org/2019/08/01/building-for-the-long-term-why-business-strategies-are-needed-for-community-owned-infrastructure/
[32] Ford Foundation (2016), 'Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure' https://www.fordfoundation.org/media/2976/roads-and-bridges-the-unseen-labor-behind-our-digital-infrastructure.pdf
[33] Educopia Institute, (2019), 'Mapping the Scholarly Communication Landscape 2019 Census', https://educopia.org/wp-content/uploads/2019/06/Census2019_EducopiaPublications.pdf

overhead and with flexible and diverse sources of funding is key to its sustainability and will allow it to remain nimble and adaptable in a constantly shifting landscape'. ROR also benefited from managing a comparatively small dataset, making it easier to not have the need to launch a new organisation to manage it. Open Apparel Registry said that they would find ways to reduce overheads by any means necessary to support certain key activities. They would much rather cut office space, for instance, so that they could afford to pay the best technical developers in their space.

The strategies we discussed to lower costs can help data institutions better manage some costs. A too-heavy focus on cost minimisation, however, could risk leading to data institutions adopting a cost-driven business model[34], where keeping overheads low becomes the primary focus and less emphasis is placed on delivering value. This is likely a reality of operating in a competitive, nonprofit space but must be considered by funders interested in supporting these organisations. Cost reduction must be balanced against any knock-on effect on the impact that can be delivered.

---

[34] 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

Support from funders and other actors

In our previous report, *Designing sustainable data institutions*[35], we talked about the difficulties data institutions encounter accessing certain types of funding. Data institutions are often nonprofits so may struggle to access debt financing and venture capital for example. As nonprofits tend to rely more on philanthropic and government funding, especially in the early stages of their existence when their ability to earn revenue is also limited, we focused our research on public and philanthropic funders.

We explored what makes funders interested in funding data institutions, and whether the way grant funding is awarded is aligned with the needs of data institutions.

In our research we have seen examples of data institutions successfully getting off the ground through grants they have secured to cover scoping, co-design and launch costs. Existing research, like the 2019 Census of Scholarly Communication Infrastructure Providers[36] has, however, suggested that it can be challenging to attract longer-term funding for data institutions: 'If you have a good idea, it's not that hard to get a grant or financial backing to build out that idea, but once it exists, how do you keep it running? There's not a lot of funding out there for long term maintenance (and by maintenance, I mean continuous updating and upgrading) of existing services or tools.'[37] This experience was echoed by one of the data institutions we spoke with that described how "grant makers want to fund something new."

This becomes an issue in two ways for data institutions. First, funders are more likely to fund a new project or organisation than continue to pay to maintain an existing one. Second, existing institutions who need money to maintain their infrastructure and services may find it harder to acquire funding for their ongoing operations. They are not offering a new product, they simply need money to maintain their existing one – which is less appealing for funders.

We explored whether continued public or philanthropic funding into the operational stage of data institutions is a viable option to help support the growth of data institutions and allow them to develop long-term sustainable funding and revenue models. For more details on the research questions, see the Methodology section.

---

[35] Open Data Institute (2020), 'Designing sustainable data institutions' https://theodi.org/article/designing-sustainable-data-institutions-paper/
[36] Educopia Institute (2019), 'Why Are So Many Scholarly Communication Infrastructure Providers Running a Red Queen's Race?', https://educopia.org/red-queens-race/
[37] The Scholarly Kitchen (2019), 'Building for the Long Term: Why Business Strategies are Needed for Community-Owned Infrastructure', https://scholarlykitchen.sspnet.org/2019/08/01/building-for-the-long-term-why-business-strategies-are-needed-for-community-owned-infrastructure/

Government and philanthropic funding

What are funders interested in?

Philanthropy has begun to see the benefits of funding projects with a specific focus on increasing access to data in order to achieve their desired outcomes. These outcomes are typically framed around specific social, research, educational or charitable goals in areas such as health, education and arts.

This is demonstrated through examples such as the UK Biobank being funded for the past 16 years by Wellcome and others to move health research forward, and the OpenActive initiative being funded by Sport England to promote healthy lifestyles and increased physical activity. Data institutions can help government and philanthropic funders deliver on these strategic goals. Being different to the kind of charitable projects which funders traditionally support, there is, however, a pronounced need for a strong narrative around how data institutions contribute to the community and broader social aims.

Other funding organisations, such as the McGovern Foundation, have been established to focus on delivering change through data and technology. These funders, although fewer in number, seek to more directly fund organisations whose mission is focused on maximising the benefits of data and technology for society rather than delivering predefined outcomes across more traditional charitable focus areas.

Type of funding provided

There has long been discussion about the persistent underfunding of overhead costs in the nonprofit sector and the consequent vicious cycle, in terms of nonprofits not being comfortable enough to be transparent about their true costs and funders continuing to have unrealistic expectations around the costs needed to run a nonprofit[38]. As we have seen, many of the activities that are central to data institutions' ability to operate successfully – and that incur significant costs – tend to fall into the category of overheads, which tend to not be fully covered in project funding.

The funders we interviewed said that the philanthropic sector is changing how it funds. Long term, core, often unrestricted funding is becoming common practice among some funders, with the Esmée Fairbairn Foundation almost exclusively granting unrestricted core funding. Research from the Ford Foundation shows that 727 different grant makers have pledged to loosen or eliminate restrictions on their funding, and put trust in their nonprofit partners to find the best solutions for the people they serve. This has led to better results during the coronavirus pandemic[39]. Despite this positive outlook, investigating the wider trend data (from funders

---

[38] Gregory A., & Howard, D., (2009), 'The Nonprofit Starvation Cycle',
https://ssir.org/articles/entry/the_nonprofit_starvation_cycle
[39] The Chronicle of Philanthropy (2020), 'How This Crisis May Upend Grant Making for Good',
https://www.philanthropy.com/article/How-the-Coronavirus-Crisis-May/248352

who have shared data with 360Giving) from 2013 to 2018 suggests that the overall amount of core funding has remained consistent[40].

In our interviews, funders who gave core funding expressed how it 'enables the people who do the work the greatest flexibility in order to maximize the chances of success' (Wellcome). This is supported by the findings of recent research from the Esmée Fairbairn Foundation, which suggest that grants for organisations' core costs, and in particular unrestricted grants, pay dividends as they are more likely to help unlock other funding sources for organisations and therefore diversify their revenue streams[41]. The research also shows how core funding enables evolution, as it allows organisations to try new things, take risks and learn, and to respond to opportunities and challenges in rapidly changing contexts.

The same research outlines the benefits of providing long term grants, five years or more, in freeing organisations to concentrate on delivery and impact.

Conversely, other funders continue to have reservations about unrestricted funding, and are nervous about the money being used for purposes which they hadn't envisaged. One way which the funders we spoke to were able to alleviate concerns around unrestricted funding is by building effective due diligence frameworks, for example through audit requirements, seats on the data institution's board, and staggered release of funding based on meeting pre-agreed milestones.

Funding for different lifecycle stages

Several funders we spoke to expressed a keener interest in funding the early stage start-up costs of organisations rather than funding them perpetually. There is an expectation that data institutions will diversify their funding streams over time and to reach some level of sustainability. For example, McGovern 'enthusiastically seed early stages and pilots, with the hope that successful models then find additional funding to take them further'. The funders we spoke with recognised that for many data institutions who seek to deliver a public benefit, it may be inappropriate to look for sustainability through earned revenue.They said that data institutions may instead reach sustainability through a diverse range of funding sources: from other philanthropic donors, public funding and, in cases where appropriate, earned revenue models.

UK Biobank, funded by Wellcome and the Medical Research Council

The UK Biobank is a registered charity and major health resource. It aims to improve the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses. 500,000 people aged between 40 and 69 years old in 2006–2010 from across the UK have undergone testing, provided blood, urine and saliva samples for future analysis, given detailed information about

---

[40] The Institute for Voluntary Action Research (2018), '3 things to notice from 'Slice and Dice' - a tool to explore GrantNav data visually',
https://www.ivar.org.uk/3-things-to-notice-from-slice-and-dice-a-tool-to-explore-grantnav-data-visually/
[41] The Esmee Fairbairn Foundation (2019), 'Insights on core funding',
https://www.esmeefairbairn.org.uk/userfiles/Documents/Publications/InsightsOnCoreFunding.pdf

themselves and agreed to have their health followed to create one of the largest health databases in the world. The UK Biobank relies on grant-based core funding and project funding for delivering specific enhancements. Charges to access the data are based on a cost-recovery model, with reduced fees for student projects and those from low and middle-income countries.

Wellcome is a politically and financially independent foundation, which supports researchers, takes on big health challenges, campaigns for better science, and helps people get involved with science and health research. They are flexible and generally long term with their funding in order to help them achieve their goal of moving science forward. They believe that where they are appropriate, flexible grants allow organisations with expertise to push forward without restrictions. They have supported the UK Biobank for about 15 years, reviewing their support every five years.

The UK Biobank is an interesting example of a data institution, or a scientific resource as it's more commonly described, being able to attract steady funding directly alongside third party investment in enriching the resource as it continues to evolve at the cutting edge. Wellcome described to us how the UK Biobank has continued to be enriched as a resource through additional projects undertaken by third parties, such as the Whole Genome Sequencing project. These kinds of projects continually add value to the UK Biobank: the Whole Genome Sequencing project reflected a £200 million in further scientific research.

Requirements for measuring impact

Measuring impact is an important condition of any public or philanthropic grant. Funders want to ensure that their spending has contributed to the change they want to see in the world. This can be difficult for data institutions and other data access initiatives, particularly due to the ongoing challenge of measuring the value of data and its impact on the economy.[42]

Some of the funders we spoke to, such as the Esmée Fairbairn Foundation, McGovern and the National Lottery Fund, understand these difficulties and emphasised that they do not require organisations to conduct rigorous impact assessments. These funders expect organisations seeking funding to outline a set of short term quantifiable measures, alongside a clear hypothesis of what the organisation believes they can achieve, by when, and what good would look like to them.

---

[42]The Open Data Institute (2020), 'The Value of Data',
https://theodi.org/article/the-value-of-data/

HiLo Maritime Risk Management, initially funded by the Lloyd's Register Foundation

HiLo Maritime Risk Management is an independent nonprofit organisation which aims to save lives in the shipping industry by changing the way companies address risk. HiLo collects data from shipping companies, runs it through a predictive risk model, and offers insights back to the companies. HiLo is currently funded through subscription fees from these companies, but has received significant early stage funding from, among others, the Lloyd's Register Foundation.

The Lloyd's Register Foundation is a UK charity established in 2012. Its mission is to secure high technical standards of design, manufacture, construction, maintenance, operation and performance for the purpose of enhancing the safety of life and property at sea, on land, and in the air, alongside the advancement of public education including within the transportation industries and any other engineering and technological disciplines. It sees opportunities in helping low-technology organisations and sectors to get the benefits of technology and data. It sees its impact manifesting in two ways: through the improved capacity of the funded organisation itself, and through the broader impact that the data creates.

The Foundation mainly funds early stage costs, and expects organisations to become sustainable in the long term. It sees its role as a catalyst for organisations to embrace data and technology to save lives. It funds a mix of start-up costs and project work for organisations to develop new competencies and capacities.

Through funding projects like HiLo, the Foundation is working towards achieving its goal of making the world a safer place. In a world where safety levels were flatlining over the last 10 years, investing in a data project like HiLo has, for the foundation, seen a greater impact than investing in more conventional applied safety spending'. According to 2019 figures, thanks to HiLo there has been a 72% reduction in the risk of lifeboat accidents and a 65% reduction in the risk of engine room fires on participating ships.

Other support

Funding is just one type of support that data institutions can benefit from, and there are several others that can help data institutions grow and become sustainable. Some of this support can come from funders who may be in a position to offer specialised expertise directly in some areas or may have contacts at organisations who in turn can offer support. At other times this support may come from private sector companies – who may see this as part of their corporate social responsibility efforts or an opportunity for staff development – or from governmental actors.

Data institutions will need to think of the entirety of the support that they need to operate successfully and pursue different avenues to obtain this support. They will also need to recognise that there are organisational risks if they rely on free or

reduced-rate support and assume they will not need significant investment. Pro-bono or discounted services may not be available indefinitely, and data institutions may eventually have to start covering the costs of the associated services.

The table below summarises the types of support that data institutions may need and where this support may come from.

Type of support	Description of support	Who can offer that support?
Technology support	Data institutions can benefit from advice on the best technologies to use, such as cloud servers, web hosting, and coding, and support in setting them up. Auditing and assessing suitable external suppliers can also be beneficial. Data institutions may also need support in back-end operations, or skills training so that their own staff can develop expertise.	Technology support can be offered by technology companies and consultancy firms. For some organisations with charitable aims, some tech companies may be able to provide free or reduced-rate access to their offerings, for example Slack for Charities, Google for Nonprofits, or discounted cloud services.
Legal, administrative and accountancy support	It is not uncommon for relatively new data institutions to operate with minimal staff members, and for small teams dedicated legal, administrative and accountancy staff may not be necessary.	Pro-bono support can come from legal firms. Fiscal hosting[43] or fiscal sponsorship[44] is a model that enables administrative costs to be borne by a host organisation who acts as a fiscal sponsor. Fiscal hosts - see some examples here[45] and here[46] - can have different missions and cover different geographies.
Organisational development and other specialised support	The founders of data institutions may require mentorship or training in key areas of organisational development, including but not limited to: revenue generation, vision and strategy development, service design, marketing, staffing, evaluation, and	One of the institutions we interviewed spoke about the positive experience of their main funder connecting them with similar organisations who could offer them advice or collaborate with them. That funder also helped to source participants for user research. Support in key areas of organisational development may be

---

[43] Open Collective (n.d.), 'FAQ - What is a fiscal host?', https://docs.opencollective.com/help/fiscal-hosts/fiscal-hosts
[44] Wikipedia (n.d.), 'Fiscal sponsorship', https://en.wikipedia.org/wiki/Fiscal_sponsorship
[45] Open Collective (n.d.), 'Hosts', https://opencollective.com/hosts
[46] Global Impact (n.d.), 'Fiscal sponsorship', https://charity.org/fiscal-sponsorship

	governance – areas where it is difficult to find volunteers with expertise.	offered 'pro bono' by consultancy firms.

Funders and other actors may use their networks to provide free marketing for data institutions by sharing their successes. |
| **Sector level coordination and support for data institutions** | There are initiatives that support both data institutions and funders with the aim of coordinating efforts to increase access to data in specific sectors. Support can include the provision of advice and shared services to data institutions, advocacy on behalf of them, the provision of funding structure for the sector or advice to funders on where their investment may contribute to most impact. | Invest in Open Infrastructure[47] aims to enable durable, scalable, and long lasting open scientific and scholarly infrastructure. Invest in Open coordinates efforts across the scholarly research and communications space, providing data institutions with guidance and resources as well as supporting funders' decisions through targeted research and recommendations for investment to support high-impact projects and organisations.[48]

Another example in the open access and open science space is the Global Sustainability Coalition for Open Science Services (SCOSS)[49] formed by a network of organisations to 'help secure infrastructure well into the future'[50]. SCOSS helps 'to identify non-commercial services essential to Open Science, and to make qualified recommendations on which of these services should be considered for funding support. SCOSS [then] provides the framework and funding structure, vetting potential candidates based on a defined set of criteria.'

Proposals for similar initiatives in the scholarly research space include the Open Platform Initiative and David Lewis's '2.5%' provocation.[51] The |

---

[47] Invest in Open Infrastructure (n.d.), https://investinopen.org/
[48] Ibid
[49] SCOSS (n.d.), 'What is SCOSS?', https://scoss.org/what-is-scoss/
[50] SCOSS (n.d.), 'Help sustain open infrastructure', https://scoss.org/help-sustain-open-infra/
[51] Educopia Institute (2019), 'Why Are So Many Scholarly Communication Infrastructure Providers Running a Red Queen's Race?', https://educopia.org/red-queens-race/

		Open Platform Initiative[52] aims to bring financial, strategic, and institutional coherence to community-built, interoperable, service-driven infrastructure and software projects that support the preservation and access of scholarship and knowledge. David Lewis argues in his article that 'academic libraries should commit 2.5% of their total budgets to organizations and projects that contribute. to the common digital infrastructure need to support the open scholarly commons'[53].

Table 3: Types of support offered to data institutions and who can support them

---

[52] Carter, L., Cooper, Y., & Thornton, A. (2019), 'The Open Platform Initiative: A proposal for coherence and interdependence in Community-owned Infrastructure', https://www.arl.org/wp-content/uploads/2019/05/mm19sp-OpenPlatformInitiative-Carter-Cooper-Thornton.pdf

[53] Lewis, D. W. (2017), 'The 2.5% Commitment', https://scholarworks.iupui.edu/handle/1805/14063

Conclusions and recommendations

Below we include our recommendations for

- **data institutions** and those interested in **setting up** data institutions,
- **funders**, and **other actors** who are interested in **supporting** the creation and sustainability of data institutions.

These recommendations complement the recommendations in our previous report in two ways:

- We include recommendations for data institutions around cost management and demonstrating the impact of their work
- We include recommendations for funders and other actors in addition to those for data institutions

Recommendations for those scoping, designing and running data institutions

Understand and communicate the value you offer to key actors in your data ecosystem

Articulating a data institution's value for key actors in its ecosystem can go a long way in making the case for appropriate revenue models. For instance, having a clear narrative on how increased access to data enables a particular social, educational or charitable outcome in support of a funder's strategy can help demonstrate the data institution's value in a way that's not always evident when describing the organisation's role in terms of stewarding access to data. Existing institutions can run activities such as evaluations of previous work or developing robust Monitoring, Evaluation and Learning (MEL) frameworks.

Develop and iterate an evaluation plan to demonstrate impact

Measuring impact is a challenge for long-term projects, particularly when there are no key performance indicators available to match that impact. It is important to establish, early, a hypothesis of how your work will contribute to long-term impact. Use any available broader evidence about the impact of increasing access to data, including the impact similar organisations have had, to demonstrate your contribution to change. Consideration of what would not be possible without the work of your data institution can also be used to indicate impact. In addition, outline the short and medium term expected results and corresponding indicators that you will use to track progress. Review and iterate these indicators over time.

A variety of resources are available to help you develop and conduct evaluations. The Better Evaluation[54] knowledge platform for instance has information on methods and processes plus approaches and thematic pages, events and resources on evaluation. Guidance on evaluation planning and impact measurement might also be available from government or philanthropic funders, often tailored to relevant sectors, such as Sport England's Evaluation Framework[55] , as well as from international organisations such as the World Bank's guidance on designing a results framework[56]. A useful list of resources is available through gov.uk[57].

To do all the above, make sure that during the design of your data institution you put aside some budget and resources for monitoring and evaluation purposes.

Be transparent about the full costs of running data institutions

Our research has demonstrated that there are a number of costs associated with running a data institution that are not typically funded. For example, in the Nonprofit Finance Fund 2015 survey into the nonprofit sector, only 7 percent of organisations reported that foundations always cover the full costs of the projects they fund.[58] Data institutions have spoken about the need for core funding and overhead support. Documenting these costs and selling the value in the activities they help to cover can boost the chances of obtaining funding for the full costs of running data institutions.

Understand the consequences of cost management strategies

In the absence of comprehensive funding, data institutions will have to make decisions about the resources and activities they spend their limited resources on. Strategies to manage costs come with their own risks and data institutions should understand and consider the long-term consequences of different cost management approaches.

Recommendations for those funding and supporting data institutions

Identify opportunities to align your approach with other funders

Data institutions are new types of organisations that can offer a new opportunity for funders to deliver on their public and social impact goals. But these institutions would benefit from funders and other actors telling their success stories and

[54] Better Evaluation (n.d.), https://www.betterevaluation.org/
[55] Sport England (n.d.), 'Evaluation Framework', https://evaluationframework.sportengland.org/
[56] Independent Evaluation Group (2012), 'Designing a results framework for achieving results: A how-to guide',
https://www.oecd.org/dac/peer-reviews/WB%202012%20designing%20results%20framework.pdf
[57] Public Health England (2018), 'Introduction to logic models',
https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/introduction-to-logic-models
[58] Nonprofit Finance Fund (2015), '2015 State of the Nonprofit Sector Survey - A Field in Flux: Reaching for Sustainability',
https://nff.org/sites/default/files/nff/docs/2015-survey_natl_summary.pdf

reinforcing the narrative about how their work is contributing to wider social and economic impact. Funders can help create a better understanding of the role and value of these institutions and build more support for them by talking with other funders and developing alignment across funding goals. Sharing stories of why and how they fund data institutions, and what successes and challenges they have experienced along the way, can be the first step towards that alignment.

Funders should also engage in conversations about the role they see for themselves and others in supporting sustainable data institutions. Data institutions are seen in many different ways by different funders which, considering the public benefit that data institutions aim to bring, raises questions about who is ultimately responsible for funding them. One interesting suggestion, which came from Wellcome, argues that resources which create global impact, like the UK Biobank, could be funded at a global level, opening up funding from a broader range of organisations. Coordination and alignment among those funding or supporting data institutions – through shared approaches and complementary funding – could help both international and domestic data institutions to become more sustainable.

Support data institutions to grow beyond their core activities

Unlocking additional funding or exploring new revenue sources takes time and resources. Funders can support data institutions by specifically funding activities to develop sustainable business models or providing funding that can be used for that purpose. They can more directly support some of these activities if, for instance, they are able to provide expert business development support, or help connect the data institution with organisations that can provide such support and more broadly through helping the data institution build a stronger network.

Funders should also consider supporting data institutions in conducting evaluations and measuring their impact over time. This may be something which data institutions typically find to be difficult, so ensuring they have the necessary skills and funding to evaluate their projects successfully would be a further lever to create change.

Clearly articulate your approach to funding data institutions

Several funders suggested that it was not the role of philanthropy or government to support data institutions perpetually. So, in order for funders to feel confident about the sustainability of a data institution, and for data institutions to have reassurance about the financial support they need to get started, funders must set clear requirements for data institutions to follow.

An example of this, which some of the funders we spoke to use, may be to require that a data institution put forward a business plan for the short, medium and long terms alongside an agreement around staggered milestones for the release of funds based on hitting pre-agreed targets. This enables the funders to see a path to exiting their support for the organisation once it has reached sustainability, and reassures the data institution that they'll be supported throughout the early stages of their journey. Plotting the path to sustainability together builds trust, and gives each party a stronger sense of co-ownership over the project.

Next steps

At the ODI, we've established the data institutions programme[59] to help organisations, people and communities build new data institutions, such as the types described in this report, and help existing organisations to recognise and strengthen their roles. Our data institutions programme seeks to

- provide advice to people, communities and organisations building data institutions, and enable existing organisations to recognise their role as data institutions and to steward data on behalf of others more effectively

- support data institutions in their early stages, through stimulus funding[60], dedicated peer networks and incubation

- advocate for data institutions and responsible data stewardship, through exploring the role of legislation and other policies in creating an enabling environment for data institutions

- collaborate with like-minded organisations to create new data institutions in different sectors and fields.

- make significant advances in the theory and understanding of data institutions, and the roles they can play, as well as developing tools and guidance to support data institutions through the problems they face.

To follow on from this report, we will continue to explore and test our findings through a product discovery phase with existing and nascent data institutions and data access initiatives. We will be working to develop a tool, service or guidance which will support existing data institutions and those looking to set up new ones, to overcome the barriers to achieving sustainability. In order to develop a tool or service to benefit data institutions, we may specifically look at

- The details of specific activities where they encountered difficulties, and what they would need to overcome these barriers.

- The support they have previously received or would have liked to receive.

- The key moments in the journey to sustainability where support is most effective.

- How the costs of running a data institution are driven by the role the data institution plays in its ecosystem.

---

[59] Open Data Institute (n.d.), 'Project: Data Institutions', https://theodi.org/project/rd-data-institutions/
[60] Open Data Institute (n.d.), 'Project: Data Access Stimulus Fund', https://theodi.org/project/rd-data-access-stimulus-fund/

Methodology

We identified a list of organisations that met the criteria of data institutions and we mapped them against different categories defined in previous research. Once those organisations were identified, we carried out desk research into their approach to sustainability – their business and revenue models and their main activities, resources and associated costs. We conducted desk research about these institutions and, where possible, interviewed their representatives to expand the information. We also identified a list of government and philanthropic funders who could feasibly fund the creation of data institutions in the UK and, where possible, spoke with their representatives.

The short timeframe of our research led to several limitations.

1. Limited number of interviews with data institutions and public and philanthropic funders.
2. Focus on either relatively new or 'successful' data institutions. Our research has not covered any failed institutions and therefore has not resulted in any lessons about challenges that could not be overcome.
3. Potential bias towards funders who fund or have some interest in funding data institutions or other data stewardship models.
4. Potential sector bias. We prioritised the health and energy/climate sectors both for data institutions and funders in line with the UK government's industrial strategy themes[61].

Research questions

Main research question: How can we support data institutions to become sustainable?

Secondary research questions

1. What activities are these organisations required to engage in while scoping, designing, operating, evaluating, re-designing and closing a data institution?
2. What are the associated costs of these activities?
3. What challenges do these organisations encounter while carrying out these activities?
4. What methods do these organisations engage in to reduce the costs of these activities?
5. What methods do these organisations have for generating funding to support these activities?
6. What factors should organisations choosing different business models or funding strategies take into account?

---

[61] Department for Business, Energy & Industrial Strategy (2019), 'The Grand Challenge missions', https://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/missions

Desk research

For each of the data institutions studied, we gathered information related to

- General information about the organisation – core goals or mission, legal form, sector and 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
- Main activities and costs.

Interviews

Interviews with data institutions

Criteria to select data institutions

In our research we wanted to understand the experience of a diverse set of organisations. We aimed to cover a range of data institutions that:

- Have different maturity levels
- Fulfil different data ecosystem roles
- Operate across a variety of sectors with a priority for health and energy/climate
- Use different revenue models (based on our existing knowledge and additional desk research)
- Offer applicability of learning for others.

We've interviewed or sought information by email from representatives of eight data institutions:

- **Musicbrainz**
- **Open Apparel Registry**
- **CrossRef**
- **ROR**
- **OpenCorporates**
- **UK Biobank**
- **Creative Diversity Network**
- **Open Humans**

We asked data institutions about their:

1. Revenue models
2. Key activities and resources and associated costs
3. Challenges in carrying out key activities
4. Strategies to reduce costs of key activities and resources
5. Funding challenges and desired support
6. About any recent changes in context – for example the Covid-19 pandemic.

Funder interviews

We also wanted to understand the perspective of funders around supporting the growth and sustainability of data institutions.

Criteria to select funders

We aimed to cover funders that:

- do and don't fund data institutions or other data access initiatives (based on our existing knowledge and information available from 360Giving)
- fund UK focused projects
- fund projects in the health and/or energy/climate sectors.

To compile our list, we contacted the first 10 funders in the list of the top ten grantmaking charities and trusts in the UK.[62] We also reviewed the first ten foundations on a list of the wealthiest charitable foundations[63] around the world with a view to prioritising those that would potentially find UK based projects.

We've interviewed representatives of eight different foundations:

- **The National Lottery Community Fund**
- **Lloyd's Register Foundation**
- **Sport England**
- **Esmee Fairbairn Foundation**
- **Patrick McGovern Foundation**
- **Paul Hamlyn Foundation**
- **Indigo Trust**
- **Wellcome**

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

1. funding strategy
2. types of project funded and types of funding provided (for instance core or project funding and length of funding)
3. openness and/or ability to fund different stages of a data institution's lifecycle
4. expectations around data institutions becoming sustainable.

Sampling method

Research participants, both data institutions and funders, were identified using the ODI's existing network supplemented by desk research and snowball sampling, building on the network of existing ODI partners.

---

[62] fundsforNGOS (2014), 'Top 10 Grantmaking Charities and Trusts in the UK', https://www.fundsforngos.org/featured-articles/top-10-grantmaking-charities-trusts-uk/
[63] Wikipedia (n.d.), 'List of wealthiest charitable foundations', https://en.wikipedia.org/wiki/List_of_wealthiest_charitable_foundations