



Financing Models for Sustainable Data Reuse Infrastructure

RESEARCH PAPER

ROB W. W. HOOFT

MARCO ROOS

**Author affiliations can be found in the back matter of this article*

]u[ubiquity press

ABSTRACT

We explore financing models that support data resources and associated services in a long-term data infrastructure. Data infrastructure components often originate in research projects dealing with digital data and may be discontinued when projects end. Long-term sustainability is challenging and requires a detailed understanding of the elements that need to be in place for successful continuation and operation of a digital infrastructure. In this paper we recognise the different kinds of activities that need to be undertaken in order to run research data infrastructure and individually match the costs of these activities with potentially suitable financing. We show that a single source of financing that is suitable for all activities is really hard to find and suggest an approach to pay for different activities from different sources of money. We find that some of the essential activities of research data infrastructure are systematically difficult to finance. We conclude this paper with recommendations for infrastructure providers/creators, infrastructure users, and national and international research funding organisations. Although the experience leading to this work was gained in life science data infrastructure, it is thought that the concepts presented in this paper can also prove their value for data service infrastructure in other research fields.

CORRESPONDING AUTHOR:

Rob W. W. Hoof

Health-RI, Utrecht,
The Netherlands

rob.hoof@health-ri.nl

KEYWORDS:

Research infrastructure; data infrastructure; data service; financing model

TO CITE THIS ARTICLE:

Hooft, R.W.W. and Roos, M.
2025 Financing Models for Sustainable Data Reuse Infrastructure. *Data Science Journal*, 24: 29, pp. 1–18. DOI: <https://doi.org/10.5334/dsj-2025-029>

Scientific progress is incremental: every scientific research project relies on the results of past research, and future research plans build on a foundation of current research and development results. In the case of research based on digital data, that foundation consists of both *data* and *software* operating on it. Mature software can be maintained and operated as a *service* for others; among these are catalogues, curated databases, data processing services, data annotation services, and much more. Research projects often initiate these services to address specific problems. When a research project successfully develops internally used data and software services, and these are envisioned to have broader applicability, it could be valuable to turn them into broader available data service *infrastructure* (see [Box 1](#)). Solely publishing software development in traditional papers is not as effective, as redeploying for each reuse can become very time-consuming. Services become part of the infrastructure if users routinely rely on them, mostly without thinking; this can only happen if they are very reliable.

Research increasingly relies on *research infrastructure*. Why? The process of scientific exploration building on previous results becomes more difficult when it relies on technology that is at the limits of what can be constructed. It would be very inefficient to have every researcher who needs such technology acquire and operate it separately. Therefore, exploration using such technology makes use of shared *physical infrastructure* that contains a shared instance of the technology, run as a *service* for researchers. Examples of such infrastructure are large particle accelerators and supercomputers.

More recently, research has increasingly relied on *research data infrastructure*. We can recognise that science in many disciplines becomes very data intensive. Data-intensive science has been developing incrementally from traditional science, but now that the volume and variety of data increase, data integration is becoming a dominant task, and repeating data acquisition and processing for different research projects becomes inefficient. Therefore, in a parallel to the physical infrastructures described above, curated data resources are collected in *virtual infrastructures*: the data are hosted by a specialised institute and available for new research, including common *services* operating on the data and the expertise that is necessary to run these. Since this kind of infrastructure is not linked to a physical instrument, it is not bound to a physical location and can be operated as a distributed network of expertise, making it a *virtual infrastructure*.

Box 1 Research infrastructure in the digital domain.

Planning for availability of results beyond the lifetime of a research project is the subject of a project sustainability plan. In this context, sustainable is defined ([Oxford Dictionary, 2024](#)) as ‘the ability to be maintained at a certain rate or level’; a project sustainability plan describes how the activities and/or outcomes of the project can be maintained in the long term, after the project has finished. For data services, this requires attention to two different aspects: *maintainability*, influenced by technical architecture, and *financial sustainability*, at the lowest level describing how much it costs and who is paying. In this work, we focus on the financial sustainability of data services.

Over our years of working in the life sciences, both as contributors to research projects and as contributors to research data infrastructure, we have learned that there are many mistakes and misconceptions about data service sustainability that are visible through the handling of sustainability plans describing data services. These can be summarised as ‘too little, too late’:

1. Too little: assuming that there will be sufficient resources.
 - a. Plans rely on the continued delivery of free-tier services by commercial data-service providers (such as GitHub ([no date](#)), Docker ([no date](#)), Google ([no date](#)), and Slack ([no date](#))). In reality, the conditions for these free-tier services undergo regular changes.
 - b. Plans rely on the existence of dedicated funding for infrastructure services. In reality, there is very little such funding (e.g., [European Commission: Directorate-General for Research and Innovation, 2017](#)).
 - c. Plans rely on the assumption that existing infrastructures can take responsibility for sustaining services. In reality, existing infrastructures often do not have funding

for such activities e.g., in the ELIXIR data infrastructure for the life sciences ([ELIXIR Website, no date](#)), a service provider must demonstrate sustainability *before* applying for an ELIXIR label.

- d. Plans rely on the assumption that a service will require little to no maintenance, and they expect that this service can be delivered as a minor addition during future development work, utilising computer hardware purchased with a previous grant and managed by the same people. In reality, servers have limited lifetimes, and maintenance of a service is often demanding and can require a different mindset than the pursuit of innovation ([Kern et al., 2020](#)).
 - e. Plans underestimate the effort of providing a scalable infrastructure service. The process of taking a service developed for a single use case and generalising its application, thus turning it into an infrastructure service, often requires significant reengineering. In addition, whenever a freely provided service is catching on, there is a serious risk that its use will soon outgrow available financing.
 - f. Plans rely on the assumption that if something good is made, its future will be guaranteed by someone in due time. In reality, even a combination of quality and need is not a guarantee for sustainability, e.g., the experience with Tranche as a proteomics raw data repository ([‘A Home for Raw Proteomics Data’, 2012](#)).
2. Too late: sustainability as an afterthought.
- a. Sustainability is not taken into account from the beginning of the new service’s development. Design choices do not take sustainability into account, resulting in services that are fundamentally harder (and thus more expensive) to maintain.
 - b. Sustainability plans are often among the last deliverables of a project and conceived exclusively by project members. This means that there is no planning of knowledge transfer, but only a description of what will need to be done with the resulting services after the project has ended.

In the current work, we draw on our own experience with life science data infrastructures to create guidance for sustainability efforts in research projects that aim to deliver data and data services. We start by sketching the data services landscape as the environment that we observe to be operating in and that we expect will remain constant. Within the confines of this environment, we then explore what types of costs come into play when providing research data services and also explore a landscape of financing sources. We then develop our analysis model, which can help organisations providing research data infrastructure identify which sources of financing best fit their ambitions and activities.

Our goal is to create reusable guidance for sustainability plans in research projects covering data services, within and outside of our own field of data in the life sciences.

THE STATE OF AFFAIRS: WHAT WE CANNOT CHANGE

To develop a new model for sustaining data services, it is important to describe the current state of infrastructure financing models, including aspects that cannot be changed in the short term.

State of affairs 1: There is almost no long-term funding for digital research infrastructure.

Financing of scientific research often comes from dedicated organisations or programmes, either nationally or internationally (e.g., EU’s framework programme) or philanthropic (e.g., Rockefeller Foundation), which grant money to projects through a competitive process. Many such funders of scientific research say that research infrastructure is not their remit; they assert other funders pay for that. Unfortunately, that assertion is mostly incorrect. Even what funders label as *infrastructure funding* is often *project funding* or so-called *seed funding* that assumes some other source of money is paying for continued operations. Existing infrastructure providers have done a lot of analysis but have not come to satisfactory structural funding. An analysis of the sustainability of research infrastructures conducted by the European Commission ([European Commission: Directorate-General for Research and Innovation, 2017](#)) also noticed the scarcity

of financial instruments, especially for the operational phase of research infrastructure, and explicitly calls for the development of dedicated budgets and for synchronisation at the European scale of long-term financial commitments towards infrastructures on the national infrastructure roadmaps.

Good infrastructure remains invisible to its users (Bilder, 2014; Bilder, Lin & Neylon, 2015). It would almost be possible to define infrastructure as ‘the stuff that only becomes noticeable when it is broken’. Consequently, many take infrastructure for granted. This holds for research data infrastructure like it holds for the electricity grid: no bioinformatician is ever worried about e.g., the European Bioinformatics Institute (EBI) (EBI Website, no date) services until they malfunction.

In stark contrast to this, Drysdale *et al.* (2020) describe how even the most core data resources used in biology, used by bioinformaticians and biologists worldwide, have very short funding horizons and rely on frequently securing new short-term project funding in order to maintain their long-term services. Bastow and Leonelli (2010) also identify the lack of public funding sources and describe alternative financing methods for data resources, each having their own drawbacks. Gabella *et al.* (2018) similarly go through a list of twelve different existing financing models and describe how they could apply to the Uniprot resource. They do not come to a satisfying model and propose a radically different way of collecting money for digital infrastructure services. Virapongse *et al.* (2024) similarly describe difficulties for the sustainability of what they call cyberinfrastructure for the earth sciences, lacking models beyond the current in-kind support from academics, researchers and their institutes.

State of affairs 2: Researchers won't readily pay for research data services.

Many people recognise that for societal infrastructure like the electricity grid and the road system, economies of scale and requirements for resilience favour a centrally maintained infrastructure. Nevertheless, there are frequent complaints about the shared costs, e.g., road taxes. Users of research data infrastructure underestimate the value provided to them by the service and, at the same time, underestimate the amount of effort it takes to ‘do it yourself’, in effect paying with their own time that is then not available to perform the core research work. Researchers rarely fully understand what it means to run research infrastructures (Hooft *et al.*, 2023). In addition, if a price is put on academically provided infrastructure, it is often unrealistically low; the reasons for low prices may be either that the infrastructure is subsidised or that the provider has (political) reasons to make sure its capacity is fully booked.

As a consequence of the wrong ideas about the cost of infrastructure services, provisional users of new infrastructure are hard to convince to pay an honest fee for use of a centrally provided service. The situation is even more difficult in an organised data infrastructure like ELIXIR (ELIXIR Website, no date), the European infrastructure for life science data, where one of the *principles* is that in most cases users get free access to their services, and neither they nor their organisations need to be a member of the infrastructure to benefit from its services. The members (in ELIXIR's case countries or selected institutes in member countries) pay for the maintenance of the infrastructure organisation in addition to bearing the cost of running their services. There is a component to this that can be referred to as a *prisoner's dilemma* or *the tragedy of the commons*: if services are free for non-members, there must be other incentives to be a paying member and service provider in such infrastructure, for example, to influence the priorities of the infrastructure or to be involved in international collaborations.

There is another aspect to expecting end-users to pay for data services: it could lead to a future in which researchers need to subscribe to dozens of research data infrastructure services to do their work (Bilder, Lin and Neylon, 2015). That situation would be like scientific publishing before open access. Not only is the maintenance of so many subscriptions a large overhead and requiring a lot of planning, but in addition, there is a serious risk that an individual service becomes a monopoly, thereby losing the need to operate efficiently and/or potentially gaining a goal to become (very) profitable.

State of affairs 3: Money is unnecessarily wasted on reinventing wheels.

Before buying a car, one considers the roads that it will need to drive on. And since a car is bought to serve for several years, the state of the roads in a few years is as important to that decision

as the state of the roads right now. If roads will not be maintained, buying an all-terrain vehicle makes the most sense. Considerations of a research project for research infrastructure are very similar (Hoof et al., 2023); if a digital infrastructure cannot convincingly show to its prospective customers that they will still be operating in five years, how can they convince projects to rely on their services? As a consequence of the absence of long-term funding for infrastructure, noted above, infrastructures will often have difficulties making these necessary long-term commitments. This lack of long-term commitment in turn means that potential users will decide not to make use of the services and may choose technically inferior solutions or build their own.

Another money-waster is the disconnect between the expectations of a service as developed through a project grant and the requirements of a reliable operational data service. Research projects will rarely take a new idea beyond Technology Readiness Level (TRL) 3 or 4, whereas it will only become usable as a service for others at TRL 6 or 7 (Héder, 2017). A reliable operational service requires several people who share the expertise to operate it; a single programmer who developed it in a project grant will not suffice. Furthermore, since neither funding for research (preferring ‘something new’, often called ‘excellent science’, over continuing work that has been published before) nor funding from industry (needing more proof of a suitable business model than can be shown at TRL 4) is available to take services through the TRL 4 to TRL 6 gap, concrete realisations of good ideas end up being described in the scientific literature, and their implementation is abandoned. The techniques may subsequently be independently reimplemented or even redeveloped by others using new project funding.

STEPS TOWARD BETTER SUSTAINABILITY PLANS

As described, there are several issues affecting research data infrastructure funding that we will not be able to tackle. We do see possibilities to lead to better sustainability even within these constraints. This is based on a simple thought: rather than looking for a single source of financing that would cover the entire *long-term sustainability of a service*, identify *what* activities must be undertaken to operationalise it, and consider *separate* sources of financing for different classes of activities.

The approach we advocate is related but not identical to ‘Activity-Based Costing’ (ABC) (Kaplan and Bruns, 1987). In ABC, the cost for providing a service is calculated from the cost of activities that lie at its basis, including indirect costs, and this cost is presented to the buyer or financier of the service. In our approach we also decompose the provision of the service into different required activities, but we propose to find stakeholders who are willing to invest in each activity rather than using the calculation to substantiate the total costs to users. Both ABC and our approach require an analysis of the cost of the composing activities, something that appears to be notoriously difficult for academic organisations since they often do insufficient accounting to be able to calculate where each part of their money is spent. Starting with such analysis early, in the context of the development of a sustainability plan, can improve that situation.

In this paper, we develop a model that describes the different activities that an infrastructure must undertake in order to provide its services to research. We argue that different kinds of activities call for different types of financing, and we make a systematic analysis of the possibilities. We will show that some essential activities of infrastructure providers have poor financing options.

The analysis and general categorisation of both service-related activities and potential classes of financing, as given below, will show that no single source of financing is suitable for all required activities, but that different combinations can be made that can cover the entire spectrum.

CATEGORIES OF COSTS IN A DATA INFRASTRUCTURE

Anyone who has tried to provide a data service can attest that this is not simply a question of putting functionality that was developed for a specific use case into a web service for others to use. Turning an *occasional service* into something that other researchers can rely on for their own projects requires the buildup of a *proper data infrastructure*. A range of different activities are required to operate such an infrastructure.

For our analysis, we need a cost analysis for infrastructure activities. There are existing cost classification systems for research projects, but these are not separating different activities in an infrastructure service operation. Such existing classifications include the following:

- Large research projects in Europe are usually budgeting Direct Costs and Indirect Costs (Antonopoulos, 2024). Direct Costs need to be auditable (through timesheets or bills), and eligible Indirect Costs are a fixed percentage of the Direct Costs.
- In the UK (UKRI, 2022), the distinction is usually made between Directly Incurred Costs (auditable time of people working on the project, travel costs, and equipment costs) and Directly Allocated Costs (costs based on estimates, such as time of supervisory people involved, estate costs, and indirect costs).
- Costs for data-related projects are sometimes split between People, Infrastructure for People, and Computers.
- The StR-ESFRI project's cost estimation guidelines for infrastructures (StR-ESFRI, 2019) make a distinction between costs incurred in different phases of the lifecycle of a research infrastructure: construction, operation, upgrade and decommissioning; it also separates 'investment' and 'operating' (fixed and variable) costs.

Mayernik (2023), specifically on research infrastructure, makes a distinction between infrastructural work and institutional work that are both needed to make infrastructure function in practice. They do not apply this classification to costing.

The given categorisations for project costs are not specifically applicable to our analysis of data infrastructure costs, although many of these classifications can be partially applied to our case. We therefore carefully considered the existing models and developed, based on our own experience and discussions with peers, a specialised hierarchy of cost categories. At the top level of that hierarchy, they are split into three different categories of activities that need to be supported:

1. **Development of Services.** Any activity that can change what is exactly offered to users.
2. **Infrastructure Operations.** Any activity that is needed to offer the services to customers.
3. **Maintaining the Organisation.** Any activity to maintain an organisation that runs service development and infrastructure operations.

Each of these three can be split into different subcategories. We will introduce these in three subsections.

DEVELOPMENT OF SERVICES

It can be observed both in academic and industrial information technology projects that when a data service is *no longer developed* that is not a sign that it is mature, but that it has been *retired*. Any service that is in active use will be under continuous development. We subdivide developments that are undertaken in a research data service infrastructure into seven different subcategories: five (labelled A through E) that directly change the service itself and two (F and G) that make the service more valuable to users.

- A. Exploratory work.** Small projects with a focus on testing the feasibility of new functionality or new applications of existing functionality. This kind of project can be called pilot studies or implementation studies (in the ELIXIR ESFRI infrastructure these fall under the name of commissioned services (ELIXIR, no date)). Exploratory work can lead to the proposal of larger development projects, e.g., sub C.
- B. Adaptations.** Users who wish to apply the services in a new context, such as their own projects, can engage the infrastructure to help develop new aspects of the service. These should be developed in collaboration between the user and the infrastructure so that the new aspects will serve a more generalised problem and be usable by other customers.
- C. New functionality.** Development of new options or entirely new services in the portfolio. This can be done through collaborative projects, potentially involving the first users of the new functionality.

- D. Professionalisation.** Making new functionality, either added through projects in the infrastructure (see B) or through external acquisition, robust enough to be served to a large audience. Focus lies on increasing the Technology Readiness Level (TRL) (Héder, 2017) from a standard academic product to a level that can be guaranteed to customers of the infrastructure.
- E. Maintenance.** Security patches, bug fixes and upgrades to frameworks or operating systems. Minor changes can be carried internally by the infrastructure; major changes could be run as separately financed projects. Maintenance is considered part of the development process because, in practice, it usually results in the software's evolution.
- F. New Data Content.** The value of many data services comes from the data that is available (e.g., through a repository or database) or used in the service (e.g., a trained model). The incorporation of new data evolves the service even without new functionality. Not all new data additions fall under this category; the additions here are often batch-wise additions, while other data additions could be seen as more part of the actual service offering (sub H below).
- G. Documentation and training materials.** This includes the development and improvement of documentation and training materials describing the use of the services. It could also include changes to the service that are made solely to improve the usability. These activities are usually assigned low priority.

INFRASTRUCTURE OPERATIONS

We all hope that data services can be run using a 'set and forget' mentality: you set them up once, and they will do their job without any human intervention. Expected effort needed for operations can be small if the need for personal support to users is ignored (see Box 2). However, while the scale of operational costs can vary significantly between different services, these costs are never zero. Some of the operational costs are independent of the number of users, and these are referred to as fixed costs. Others scale with the number of users, and these are referred to as variable costs. The additional costs to serve one more customer can be referred to as the *marginal cost* (O'Sullivan and Sheffrin, 2007, p. 111) for that customer. The ratio of variable to fixed costs is a very important parameter: if marginal costs are significant, a service is limited in the number of customers it can serve. Examples of such services are training events and computer storage capacity. This stands in contrast to the maintenance of highly curated databases, where the (variable) cost of the actual usage is only a minimal fraction of the total cost of operations, and the bulk of the costs are fixed (curation) costs.

For physical infrastructure, it is obvious that it requires support personnel. Nobody expects that a researcher who wants to make use of a particle accelerator will be happy with the keys to the building and a box of manuals for the magnets and vacuum pumps (reductionist view). The effective use of the infrastructure relies on the availability of experts who know how to use the ensemble (holistic view).

The expectations are (unfortunately) often different for a data services infrastructure. Researchers who want to make use of data services are often provided with access to web resources ('keys to the building') and lists of data resources and services ('manuals for the magnets and vacuum pumps') and are expected to know how to apply these in their own research. In some cases, (e-)learning is offered. However, the most effective use of a data infrastructure *also* relies on the availability of experts who know how to use the ensemble.

Box 2 Data infrastructure needs operators too.

We classify operational activities into six subcategories, three internal (H, I and J) and three user-facing (K, L and M). For each internal subcategory, we identify how the costs scale with the number of customers.

- H. Basic infrastructure.** The underlying infrastructure that the service relies on, including compute and storage systems and network connections but also subscriptions to other data services and human infrastructure for continued data curation. These costs are all independent of the number of customers of the service.

- I. **Operating a basic service level.** Running the lowest level of operations of the service. This can be the 'Free' tier of a 'Freemium' service (Marín de la Iglesia and Labra Gayo, 2009). The costs in this bucket generally scale with the number of users of the service.
- J. **Resources for premium users.** Extra resources needed to run large numbers of queries or other premium services for premium users; something that not all users of the service will have access to. These costs scale with the needs of the premium users.
- K. **Running user training events.** Training events run at customer's or at your own premises or virtually, for users of different levels of experience with the services.
- L. **Help desk.** Manning e-mail and phone help desks to help customers to the level that can be expected from the service. There may be basic and premium versions of the help desk too.
- M. **Expertise.** Helping customers set up their projects using the components of the infrastructure. This can be seen as an extension of the helpdesk. In general, this will concern non-routine use of the services, where the internal expert works together with an expert of the customer to design a solution for the customer, and that could lead to lessons learned for the infrastructure.

Data additions can fall under Development (F) as well as under Operations (H). The distinction may not always be obvious, but in first approximation (batch-wise), data additions to a *random* data collection fall under Development activities, whereas the continuous maintenance of a *complete* or *timely* data collection for a service falls under Operational activities.

For both the internal and the user-facing activities, the scale can be strongly dependent on the level of service provided. It is therefore important to make explicit what level(s) of service provision the organisation aims for.

Some possible service levels for tools and databases, in order of increasing operational costs to the infrastructure, are:

1. User can download software and/or data from a repository that it is no longer developed and serviced (**dead**)
2. User can download software and/or data from a repository; there are an active mailing list and developer group (**alive**)
3. Software is run as shared web service without SLA (**Free** or Freemium)
4. Software is run as dedicated web service for users with an SLA (**SaaS**)
5. Service is partly developed and deployed for users through participation in project (**participatory**)

Service levels for expertise, in order of increasing costs, are:

- A mailing list or forum (**public**)
- Ticketing or personal contacts (**incidental**)
- Participation in projects (**participatory**)

MAINTAINING THE ORGANISATION

The organisation is an essential foundation for running a service. Virapongse *et al.* (2024) observe that data infrastructure initiatives in earth sciences tend to move towards becoming an organisation in order to pursue sustainability. Even if no new organisation is started, sustainability relies on an existing organisation accommodating its operations.

Organisational costs are roughly proportional to the size of the operation, which is only partly determined by the number of users. Costs grouped under this header are all taken for granted in the context of research projects, and a budget for these is often not made explicit in sustainability plans. We recognise eight subcategories of organisational activity (N through U):

- N. **Housing.** Rent, office space, lab space, basic meeting rooms, heating, electricity.
- O. **Organisation services.** Office costs and office services like human resources, finance, travel office, secretarial support, etc.

- P. Spreading the news** (within the direct professional network of the organisation). Maintaining a network, attending and presenting at conferences, attending and presenting at meetings; this includes travel costs associated with these activities. Networking has as its primary goal the promotion of the availability of the infrastructure and also gathers input about its use and limitations.
- Q. Outreach** (outside of the professional network of the organisation). Representing viewpoints in society, materials for public education and schooling; this includes travel costs associated with these activities.
- R. Funding acquisition.** Grant writing, lobbying, analysing impact, showing impact to decision makers.
- S. Getting new users.** Acquisition of new users or members/partners to the organisation. This includes helping them become users by writing letters of recommendation.
- T. Maintaining user contacts.** Stay connected to existing users and partners in order to make sure they get what they expect. This can take the form of one-on-one meetings and/or organising community meetups.
- U. Memberships.** Both work and fees required to be part of other (e.g., international) organisations/infrastructures.

SOURCES OF FINANCING FOR A DATA INFRASTRUCTURE

A *viable* infrastructure has, by definition, a financing model that recovers all the costs associated with the activities listed above. This can be achieved through various means. Here we explore possible sources to finance the activities associated with a data infrastructure.

Just like for costs, we need a categorisation of different sources of finance.

Our categories are:

- 1. Public project funding.** Any public funding source that can deliver a predetermined budget for a fixed period of time.
- 2. Public infrastructure funding.** Any public funding meant to sustain a longer-lasting infrastructure.
- 3. Institutional financing.** Long-term financial commitment from the hosting institute available specifically if a service is considered one of the primary outputs of the institution, one of its *raison d'être*.
- 4. Private Financing.** Private parties who see the services as beneficial or aligned with their organisational goals.
- 5. Organisational Partnership Fees.** Money paid by organisations who, in return, can influence decisions and/or get privileged or exclusive access to services.
- 6. Industry Partnership.** Industry partners who see the services as beneficial for their customers may gain a competitive advantage.
- 7. Commercial Services.** Income generated by selling (other) services at a profit.
- 8. User Fees.** Any money paid by the user or by an organisation that the user represents that specifically allows the user to gain access to the service.
- 9. Submission Fees.** Money paid by the data provider to a data service provider, which uses that data to provide services to *other* users.

OECD (2017) has analysed financing sources for 48 repositories, coming to five classes of financing for these services. Their category *Structural Funding* corresponds to our category two, Public infrastructure funding; *Host institution funding and support* is like our Institutional financing; *Data Deposit Fees* correspond to our Submission Fees, *Access Charges* correspond to our User Fees; and *Contract Services or project funding* corresponds to our Public project funding. Kitchin et al. (2015) identify 14 different funding streams and classify these into 6 different classes; they map to our nine sources except for their 'Razor and Blades' and 'Free with Advertising' models.

We will go into a little more detail about each of our funding categories in subsections below. Some could be further subdivided into subcategories, but since most of these subdivisions are not behaving differently in our financing model, we will not explicitly name each of them. This list is not exhaustive: you may have access to financing sources that have aspects of multiple of these categories or are different altogether. In sections below, it will become clear how such different financing sources could be added to the model.

PUBLIC PROJECT FUNDING

National and international (e.g., EU) research funding organisations are structured in topical calls that can be responded to with proposals. Some funding instruments specifically target regional development, limiting their use to initiatives within a specific geographic region. Each proposal must have a budget detailing how the money will be spent. Often, one of the things that should be delivered by a project is a sustainability plan, explaining how the results will be sustained (using other financing) after the project term finishes.

PUBLIC INFRASTRUCTURE FUNDING

Funding intended to establish and sustain research infrastructures can come from national or international funding institutes or directly from a government. Beware of funding that is labelled as part of an ‘infrastructure programme’ but that is only meant to help with the startup of an infrastructure (so-called seed funding). Under this categorisation, this is likely better captured under project funding. While it may have a slightly longer running time, it is expected by the providers of such financing that the actual running of the infrastructure will continue based on *other* financing when the seed funding ends.

The NIH has experimented with a Common Fund model in BD2K ([BD2K, no date](#)), where it provided funding to (cloud) infrastructure services based on the needs identified in research proposals. To facilitate this, eligible research proposals were given ‘cloud credits’ in addition to their own funding, which they could spend on participating services. This is an example of a system where funders pre-allocate dedicated infrastructure funding. The experiment by NIH was successful but has not been continued, and the initiative for a similar model of Cloud Coins ([Pringle, 2018](#)) for the European Open Science Cloud (EOSC) ([EOSC, no date](#)) did (so far?) not materialise.

INSTITUTIONAL FINANCING

There can be different reasons for organisations to finance services from their core budgets. Two of the more generally applicable reasons are:

1. A research institute can see the benefit of centralising a service that is used by different researchers and pay from central financing for a centralised service that can be used by everyone in the institute.
2. A government-funded institute with a specific societal goal can have great internal freedom to allocate their available budget to different activities that support that societal goal. Running data infrastructure services can be part of that. This is relevant especially when the data service is essential internally in the institute (see first point) and can be provided to a much larger external audience at a relatively low additional cost.

PRIVATE FINANCING

Just like institutes, private funds often pursue specific societal goals. Publicly run services can be a very good match to such goals. Examples can be private foundations like the Gates Foundation ([Gates Foundation, no date](#)) and the Chan Zuckerberg Initiative ([Chan Zuckerberg Initiative, no date](#)), charities, patient federations, and funds brought together by other stakeholder groups.

Similarly, individuals may see the goals of the infrastructure as worth pursuing, and they may be willing to donate. Such donations can be tax-deductible under certain circumstances (e.g., in the USA, if the infrastructure has a 501(c)(3) tax-exempt status, and in the Netherlands, if it is recognised as ANBI, *Algemeen Nut Beogende Instelling*).

ORGANISATIONAL PARTNERSHIP FEES

Financing through the membership fees of an organisational partnership, whereby the partnership is not restricting the usage of the service to members of the partnership (if the service is restricted to members, it is a form of user fee; see below).

We can recognise partnerships where:

1. Members are stakeholders/organisations (e.g., Dutch Techcentre for Life Sciences) (Eijssen *et al.*, 2016).
2. Members are countries/national organisations mandated by government, such as the European Research Infrastructure Consortium (ERIC) (ERIC, no date), BBMRI (BBMRI-ERIC, no date) and other ESFRI (ESFRI, no date) infrastructures.
3. Large-scale country membership where governments are directly involved (e.g., CERN (CERN, no date) and ITER (ITER, no date)).

For the goals of our financing model, these partnerships are similar.

INDUSTRY PARTNERSHIP

Industrial partners may have a stake in a public service, for example, if their product builds on top of the service or if the public data service increases the value for users of their own commercial product, maybe leading to a competitive advantage. Such a dependency relationship can result in a collaboration in which the industry partner finances aspects of the public service provision. This may be accompanied by online advertising (an example of such a relationship is given in Gabella *et al.*, 2018).

COMMERCIAL SERVICES

Financing through commercial services uses income that is generated by providing services that are based on an infrastructure service, at a profit. Such commercial services may be offered by your own organisation (in order to generate money that can be spent for any purpose by the organisation, for example, a systematic surplus from organising training events or conferences) or by other (commercial) parties who thereby have an interest in the sustainability of the infrastructure and are willing to pay a fee that helps this sustainability.

USER FEES

User fees are payments made by the user, their employer, a national organisation, or any other group to access the service.

Table 1 collects a number of different models for user fees that are in common use.

HOW A SERVICE IS OFFERED	HOW MUCH COST IS RECOVERED
Free	None
Freemium (Marin de la Iglesia and Labra Gayo, 2009)	Marginal costs for premium services
Only marginal costs charged, if significant (this model is propagated by ELIXIR)	Marginal costs only
Realistic operational costs (including some indirect costs)	Operational costs only
Full costs, including organisational and development	100%
Full cost + profit margin (commercial)	> 100%

Table 1 User fee-based cost recovery models.

A very important distinction is between ‘academic’- and ‘commercial’-style user fees, where we use the term ‘commercial’-style user fees if it is known to the users that they pay more than just the bare running costs for the service.

Providing services for free is often considered the ideal for academic services. However, completely *free* is not always the best model to deliver services to as large an audience as possible. A *free* service offers large advantages in that it avoids a lot of administrative overhead and that it is equally available to users who would not be able to afford paying any fee. However,

if a free service grows too large, the resources needed for operation may exceed available capacity. Requesting a fee also benefits the provider by making it easier to demonstrate to governmental funding organisations that the service provides significant value to its users. Note that policies driving for open data may limit the possibilities to charge user fees for data infrastructure services, especially repository services (OECD, 2017).

SUBMISSION FEES

Submission fees are payments by providers of content, especially applicable to a repository service. These fees only make sense when the repository provides a required service to the submitter of data, for example, when the submitter is obliged to make use of the repository through institutional or funder policy. Submission fees are comparable to article processing charges in the scientific open-access publishing world, where a paper is considered an essential service to the authors (as well as to the readers). Submission fees also have the same problem in that they give an advantage to researchers from rich countries and institutes.

Independent of the cost structure of the service, a submission fee can be deemed inappropriate by a repository that wants to maximise value to the users of their content, because it favours a focus on quantity of content over quality.

NOT EVERY SOURCE OF FINANCING WILL PAY EVERY COST

Given the two hierarchies for cost types and financing sources, we can now examine which type of financing is suitable, less suitable or unsuitable for each cost type. This can be expressed as a matrix, where suitability for each combination is expressed as a grid of traffic light colours: green means suitable, yellow marginally suitable and red unsuitable (Table 2).

		Sources of Finance									
		1. Public Project finance [1]	2. Public infrastructure	3. Institutional financing	4. Private financing	5. Organisational Partnership	6. Industry Partnerships	7. Commercial services	8a. Academic user fees	8b. Commercial user fees	9. Submission fees
Categories of Activities / Spending	Development	A. Exploratory	Green	Red	Yellow	Green	Yellow	Green	Yellow	Red	Yellow
		B. Adaptations	Green	Yellow	Yellow	Green	Red	Green	Yellow	Green	Yellow
		C. New functionality	Green	Yellow	Yellow	Green	Yellow	Green	Yellow	Red	Yellow
		D. Professionalization	Green	Yellow	Yellow	Green	Yellow	Green	Yellow	Red	Yellow
		E. Maintenance	Green	Yellow	Yellow	Green	Yellow	Green	Yellow	Green	Red
		F. New data content	Green	Yellow	Yellow	Green	Yellow	Green	Yellow	Green	Yellow
		G. Documentation	Green	Yellow	Yellow	Green	Yellow	Green	Yellow	Green	Yellow
	Operations	H. Basic operation	Red	Green	Green	Green	Green	Green	Green	Green	Yellow
		I. Free service level	Red	Green	Green	Green	Green	Green	Green	Green	Yellow
		J. Resources for premium users	Red	Yellow	Red	Red	Yellow	Yellow	Red	Green	Red
		K. User training	Red	Green	Green	Red	Green	Green	Red	Yellow	Green
		L. Help desk	Red	Green	Green	Yellow	Green	Green	Green	Green	Yellow
		M. Expert usage assistance	Yellow	Green	Green	Green	Green	Green	Green	Green	Red
	Organization	N. Housing	Red	Red	Green	Green	Green	Green	Green	Yellow	Green
		O. Organisation services	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Green
		P. Spreading the news	Green	Yellow	Green	Green	Green	Green	Green	Green	Green
		Q. Outreach	Green	Yellow	Green	Green	Green	Green	Green	Green	Green
		R. Funding acquisition	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow
S. Getting new users		Red	Yellow	Green	Green	Green	Green	Green	Green	Green	
T. Maintaining user contacts		Red	Yellow	Green	Green	Green	Green	Green	Green	Green	
U. Memberships		Red	Green	Green	Green	Green	Green	Green	Green	Green	

Table 2 Infrastructure funding matrix, showing suitability of different financing sources to cover costs of different activities undertaken by a data infrastructure service provider. Colour code: green = suitable, yellow = marginal, red = unsuitable. (Source document see: [matrix](#)).

Important to note is that this is a single prototype of a funding matrix built up with relatively generic considerations. In a different context, for a specific infrastructure, the hierarchies of activities and financing sources as well as the suitability colouring can be different.

Some of the considerations used in colouring a matrix can be:

1. **Explicit eligibility:** If there are documents describing activities that are eligible for financing from a certain type of source, do these include the activity at hand?
2. **Implicit eligibility:** Does the target of the activity belong to a (strategic) target of the financing source? Or would the finance source disapprove of the cost if they discovered their money is being used? For example, project funding in scientific research targets the financial support of breakthrough novel science, which is one group having access to a technique; they generally do not target the improved availability of techniques and services to larger groups of researchers.
3. Is the timeline of the financing compatible with the timeline for the activity? Temporary grants are less suitable to pay for continuing costs ([Bilder, Lin and Neylon, 2015](#)).
4. Could the financing match result in a perverse incentive, e.g., a monopoly or other kind of vendor lock-in situation? We would like to avoid situations with infrastructure services that are effectively multiple monopolies.¹ One consequence of this is that money for developing and maintaining the Commons is not suitable to request from individual users. Individual users do not expect to pay for this.
5. Is it legal? It may be illegal to charge users significant fees for access to data. Data is not protected by copyrights, and database rights may offer limited possibilities to charge for access. For example, European Database Law ([Directive –96/9- EN – EUR-Lex, 1996](#)) only allows selling access to a database by the creator of the database if the creator has made a significant investment in creating the database (which is arguably not the case if the investment was paid for by public funds), and the European Data Governance Act ([Regulation –2022/868- EN – EUR-Lex, 2022](#)) gives data sharing conditions for public sector bodies.
6. Is there a match between the stakeholder that finances a cost and the stakeholders that benefit from the related activity? This rule is violated if the entire user group pays for services that are used only by a few, or if a single user is (without knowing it) financing services that are provided to all. This rule is not violated if a research funder finances a service that is made available for free to those researchers they would otherwise have to finance to “run their own” (but this may not be the way they strategically operate). A good example of a match is adaptation (B), which could be co-financed through participation of the infrastructure in the user’s financed project.

More specific considerations for individual cells or small groups of cells were:

- *Public infrastructure funding does not readily pay for organisations.*

Almost all of the organisational costs in the column of public infrastructure funding are yellow, with housing a solid red, because infrastructure grant providers are not willing to pay for the organisational infrastructure needed to provide the service ([NWO, 2021](#), section 3.2.2 ‘Costs not eligible for funding’).

- *Offering premium resources on internal funding is dangerous.*

Several services have ceased to operate in the past because they offered premium services (costing serious money to provide) for free to their users, resulting in unsustainable growth of premium user numbers and cost. This fact results in the red colour for premium services offered from institutional financing. Pay per use may be the only real sustainable way of offering premium services. This consideration is also made in [OECD \(2017\)](#), which lists it as a key issue matching financing sources to key cost drivers.

- *Commercial service income may not be willing to pay for some new developments.*

Some part of the new development for your free infrastructure services could compete with the commercial services.

¹ Compare this to the state of scientific publishing, where there are multiple publishers, but since each has different content, it is still a requirement to have subscriptions to each one in order to do research.

- *Income from submission fees could be in antiphase with development needs.*

Expecting to pay for the development of new features or explorative development from submission fees could be risky because the need for more development may become clear from reduced numbers of new submissions; therefore, the need for money may be higher when the supply is restricted.

- *Paying operational costs through submission fees relies on continued growth.*

Even basic operational costs for providing a service, except for the submission service itself, require that the income from new submissions pays for the service on all existing data. This requires continued growth to be sustainable. Submission fees are probably most suitable if the submission process is dominant in the operational costs.

Note that if a certain cost is very small in comparison to the rest, it can be paid by ‘overheads’. However, the suitability of such practice still depends on the time horizons of the financing source and activity being comparable; even when they are small, one cannot expect to pay long-term operational costs as overhead from a short-term development grant.

HIGHLIGHTS FROM THE MATRIX

A few observations can be made at a glance from [Table 2](#).

Every column in the matrix, representing a single type of financing, contains yellow squares, and all but two columns contain at least one red square. If an infrastructure is seeking a single source of financing for the full spectrum of activities, that leaves only industry partnership or commercial user fees as options. Both have their own serious implications for the organisation’s structure and services. For example, we have heard of services that ran out of serial development grants and, as a final foothold, switched to commercial user fees. As a result, their user population significantly changed, and they had to become completely different organisations.

An alternative financing strategy for an infrastructure-providing organisation involves various combinations of multiple funding sources that collectively cover all cost types, and these combinations can be adapted to many different organisational models. For each active cost type, a research infrastructure should ideally have a *green* funding source at any moment. If none of those can be found, at least one of the *yellow* funding sources for that row should be explicitly considered and probably explicitly asked to provide funding for that cost.

DISCUSSION

We propose that stakeholders in need of or involved in financing the delivery of dependable, long-lasting data infrastructure could create a matrix mapping sources of financing against their infrastructure tasks. We created a first version of such a matrix based on our experience and suggest that most of the contents are broadly applicable. We encourage prospective data service providers to try it out, expand it with variations of their own (for example, by substituting alternative financing models like those from Kitchin *et al.* (2015)) and share additional insights.

The technique for creating a funding matrix and the implementation presented in this paper were developed to gain insight into how financing tools can be applied to organisations managing data infrastructure, drawing on the experience of the authors. By our delineation of infrastructure activities and our experience of acquiring financing for infrastructure, we were able to document why it is difficult to finance ‘state-of-the-art’ data infrastructure services.

Implementing a pluriform financing strategy for data infrastructure adds complexity to the data service organisation, but it also aims to improve the long-term viability of the infrastructure and facilitate communication with stakeholders through additional clarity in the cost structure.

Lack of good financing models incurs unnecessary costs to society. New services are developed based on technology that was initially delivered by publicly funded projects. Without good long-term financing, such services that should form components of reliable data infrastructure could decay. Decaying infrastructure negatively impacts the cost-effectiveness of public funding in the long term. The societal costs are exacerbated by the negative impact on the trust of user communities in the robustness of new infrastructure: infrastructure without consistent long-

term financing will not be able to deliver on long-term promises needed for their customers, and if customers decide not to use existing services, this will lead to repeated reimplementation or reinvention.

Our analysis shows that most common sources of financing can only pay for part of the activities that need to be undertaken to run data infrastructure services successfully. Our suggested solution is to use different financing to pay for different activities. We have not seen this approach elsewhere before. OECD (2017) already remarked that approximately half of the data repositories they surveyed are relying on more than a single source of financing but does not decompose the usage of the different financing for different types of costs and instead bases their analysis on several desirable characteristics of a *business model*. Erway (2012) also remarks that the seven repository services they profiled often rely on multiple sources of financing but do not specify whether the funding from different sources is used in different ways; in fact, they suggest that the most straightforward approach is to have 100% government funding or funding from an endowment, and they conclude, as we do, that such financing is hard to pursue. Virapongse *et al.* (2024) contains an inventory of possible financing sources but does not decompose the activities of the infrastructure.

Some sources of financing that are part of our analysis, such as public infrastructure funding or institutional financing, can be extremely hard to find in practice. For example, the ELIXIR Sustainability plan (Smith and Martin, 2023, p. 5) writes on this topic:

There are very few national or international funding organisations that have dedicated, bespoke funding schemes to support the long-term sustainability of research infrastructures, including databases, software, tools, standards, ontologies, training materials and cloud computing.

This is unfortunate, because as the matrix shows, these would be the most flexible sources of financing for a data service infrastructure. As an example of this elusiveness in practice, most of what is advertised as ‘public infrastructure funding’ is actually one-off, temporary funding, sometimes called ‘seed funding’; we would characterise this as Project Funding.

Taken together, this leads to three recommendations for key stakeholders.

For **projects developing new data services**, we recommend using the matrix as a tool to evaluate where they are in the development process: is the development still exploratory, is the tool being professionalised, or is it operational for concurrent use by many users? For complex tools, different components can be in different stages and thus fit multiple product types. This analysis can help identify which components match best with which financing options, for instance, when writing a grant or approaching infrastructure, and it can also help inform the sustainability plan.

To **infrastructure providers**, we suggest taking an approach using a financing matrix to plan consistent and complete financing for all their activities. A consistent plan avoids lumping all undertaken activities under ‘infrastructure provision’ and paying percentages of these lumped costs from different sources. Instead, a plan should aim to acquire a selection of different funding streams and pay for different activities *explicitly* from different sources of money. In the plan, flexible money should be used for tasks that are difficult to finance. In other words, the financing that offers the most freedom in its use should pay for activities outside of the remit of the other financing sources. For example, if an infrastructure has a budget based on flexible institutional financing or flexible private financing and another budget from public project funding that is specifically meant for scientific research, then the flexible funds should be used to pay for organisational tasks (e.g., P, Q, R) and for exploratory work that needs to be allocated at short notice (A), and not to do the continued developments (C) that could also be paid from specific research funding grants.

Organisations that finance infrastructure we encourage to study our analysis and take a more holistic view of data infrastructure to ensure that their investment will be cost-effective beyond their immediate scope in the matrix. Research funders should make sure that the portfolio of programmes they run matches their goals and that it does not rely on non-existing *other* sources of financing. Research funders could also point to the financing-matrix model in their instructions regarding the content of sustainability plans.

We see several other possible applications of our technique, applicable to different phases in the life cycle of a data services infrastructure:

- We started out with the idea that this methodology can inform the creators of data services about the creation of their sustainability plan.
- Next to being a tool to analyse and plan future infrastructure, the methodology can be used to identify weaknesses in the financing structure of an existing resource and suggest roads to improvement. For instance, when infrastructure operation relies on development projects with short-term funding horizons.
- Additionally, the methodology can be used to guide infrastructure maintainers and providers to develop alternative financing models before current funding runs out.

In all of these cases, we envision that it will be used to engage stakeholders in co-developing appropriate financing models.

ACKNOWLEDGEMENTS

The authors would like to thank Stefan van Aalst (0000-0002-2161-1009), Regina Becker (0000-0002-6711-8375), Erik van den Bergh (0000-0001-9865-574X), Gijs van den Beucken, Eva Bermejo (0000-0001-7282-2714), Ruud Esveld, Fotis Karayannis (0000-0001-7080-4600), Marion Koopmans (0000-0002-5204-2312), Corinne S. Martin (0000-0002-5428-2766), Rick van Nuland (0009-0000-1659-274X), Tommi Nyronen (0000-0002-5569-5183), Troels Tvedegaard Rasmussen (0000-0002-9896-2341), Merlijn van Rijswijk (0000-0002-1067-7766), Maria del Carmen Sanchez Gonzalez (0000-0002-8523-9218), Susanna Assunta Sansone (0000-0001-5306-5690), Daoud Sie (0000-0001-6762-2582), Andrew Smith, Bruna dos Santos Vieira (0000-0001-7893-0505) and Wim Vranken (0000-0001-7470-4324) for valuable discussions and contributions. In addition, the anonymous peer reviewers are acknowledged for their valuable suggestions, corrections and additionally suggested literature.

FUNDING INFORMATION

For the writing of this paper, Rob Hooft was supported by the Dutch Techcentre for Life Sciences (DTL), Health-RI, the European Joint Programme on Rare Diseases (EJP-RD) (Horizon Europe #825575) and the Genomic Data Infrastructure (GDI, Digital Europe #101081813) project.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Rob Hooft: Conceptualization, Methodology and Writing.

Marco Roos: Conceptualization, Methodology, Writing and Supervision.

AUTHOR AFFILIATIONS

Rob W. W. Hooft  orcid.org/0000-0001-6825-9439

Health-RI, Utrecht, The Netherlands

Marco Roos  orcid.org/0000-0002-8691-772X

Leiden University Medical Center, Leiden, The Netherlands

REFERENCES

- 'A home for raw proteomics data', (2012) *Nature Methods*, 9, p. 419. Available at: <https://doi.org/10.1038/nmeth.2011>
- Antonopoulos, P.** (2024) *Calculating budget for Horizon Europe Projects*. Available at: <https://www.emdesk.com/horizon-2020-horizon-europe-basics-guide/calculating-budget-for-a-horizon-europe-project> (Accessed: 1 July 2024).

- Bastow, R.** and **Leonelli, S.** (2010) 'Sustainable digital infrastructure', *EMBO Reports*, 11(10), pp. 730–734. Available at: <https://doi.org/10.1038/embor.2010.145>
- BBMRI ERIC.** (no date) *Home – BBMRI-ERIC*. Available at: <https://www.bbmri-eric.eu/> (Accessed: 1 July 2024).
- BD2K.** (no date) *Cloud Credits*. Available at: <https://commonfund.nih.gov/bd2k/cloudcredits> (Accessed: 1 July 2024).
- Bilder, G.** (2014) *Mind the GAP*. Available at: https://youtu.be/g_yvRXYl-3c?feature=shared&t=603 (Accessed: 28 June 2024).
- Bilder, G., Lin, J.** and **Neylon, C.** (2015) *Principles for open scholarly infrastructures*. Available at: <https://doi.org/10.6084/M9.FIGSHARE.1314859>
- CERN.** (no date) *CERN: Home*. Available at: <https://home.cern/> (Accessed: 1 July 2024).
- Chan Zuckerberg Initiative.** (no date) *Chan Zuckerberg Initiative – Building a future for everyone*. Available at: <https://chanzuckerberg.com/> (Accessed: 5 July 2024).
- Directive –96/9- EN – EUR-Lex.** (1996) *EUR-Lex.europa.eu*. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31996L0009> (Accessed: 1 July 2024).
- Docker.** (no date) *Docker: Accelerated container application development*. Available at: <https://docker.com/> (Accessed: 5 July 2024).
- Drysdale, R., Cook, C.E., Petryszak, R., Baillie-Gerritsen, V., Barlow, M., Gasteiger, E., Gruhl, F., Haas, J., Lanfear, J., Lopez, R., Redaschi, N., Stockinger, H., Teixeira, D., Venkatesan, A., ELIXIR core data resource forum, Blomberg, N., Durinx, C. and McEntyre, J.** (2020) 'The ELIXIR Core Data Resources: Fundamental infrastructure for the life sciences', *Bioinformatics*, 36(8), pp. 2636–2642. Available at: <https://doi.org/10.1093/bioinformatics/btz959>
- EBI Website.** (no date) *EMBL-EBI homepage | EMBL-EBI*. Available at: <https://www.ebi.ac.uk/> (Accessed: 28 June 2024).
- Eijssen, L., Evelo, C.T., Kok, R., Mons, B.** and **Hooft, R.W.W.** (2016) 'The Dutch Techcentre for Life Sciences: Enabling data-intensive life science research in the Netherlands [version 2; peer review: 2 approved, 1 approved with reservations]', *F1000Research*, 4(33). Available at: <https://doi.org/10.12688/f1000research.6009.2>
- ELIXIR.** (no date) *Commissioned Services | ELIXIR*. ELIXIR Europe. Available at: <https://elixir-europe.org/about-us/commissioned-services> (Accessed: 1 July 2024).
- ELIXIR Website.** (no date) *About us*. Available at: <https://elixir-europe.org/about-us> (Accessed: 28 June 2024).
- EOSC.** (no date) *EOSC Association: Advancing open science in Europe*. Available at: <https://eosce.eu/> (Accessed: 1 July 2024).
- Erway, R.** (2012) *Lasting impact: Sustainability of disciplinary repositories*. Dublin, Ohio: OCLC Research.
- ESFRI.** (no date) www.esfri.eu. Available at: <https://www.esfri.eu/> (Accessed: 1 July 2024).
- European Commission: Directorate-General for Research and Innovation.** (2017) *Sustainable European research infrastructures – A call for action – Commission staff working document – Long-term sustainability of research infrastructures*. Brussels, Belgium: Publications Office. Available at: <https://data.europa.eu/doi/10.2777/76269> (Accessed: 29 April 2025).
- European Research Infrastructure Consortium (ERIC).** (no date) *European Commission: Research and innovation*. Available at: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures/eric_en (Accessed: 1 July 2024).
- Gabella, C., Durinx, C.** and **Appel, R.** (2018) 'Funding knowledgebases: Towards a sustainable funding model for the UniProt use case [version 2; peer review: 3 approved]', *F1000Research*, 6(ELIXIR), p. 2051. Available at: <https://doi.org/10.12688/f1000research.12989.2>
- Gates Foundation.** (no date) *Bill & Melinda Gates Foundation | Bill & Melinda Gates Foundation*. Available at: <https://www.gatesfoundation.org/> (Accessed: 5 July 2024).
- GitHub.** (no date) *GitHub: Let's build from here · GitHub*. Available at: <https://github.com/> (Accessed 5 July 2024).
- Google.** (no date) *About Google*. Available at: <https://about.google/products/> (Accessed: 5 July 2024).
- Héder, M.** (2017) 'From NASA to EU: The evolution of the TRL scale in public sector innovation', *The Innovation Journal*, 22(2). Available at: <https://innovation.cc/document/2017-22-2-3-from-nasa-to-eu-the-evolution-of-the-trl-scale-in-public-sector-innovation/>
- Hooft, R.W.W., Harrison, E.** and **Martin, C.S.** (2023) 'The road to success: Drawing parallels between “road” and “research data” infrastructures to foster understanding between service providers, funders and policymakers [version 2; peer review: 2 approved]', *F1000Research*, 12(ELIXIR), 88. Available at: <https://doi.org/10.12688/f1000research.128167.2>
- ITER.** (no date) *ITER – the way to new energy*. Available at: <https://www.iter.org/> (Accessed: 1 July 2024).
- Kaplan, R.S.** and **Bruns, W.J.** (1987) *Accounting and management: A field study perspective*. Boston, Massachusetts: Harvard Business School Press.
- Kern, F., Fehlmann, T.** and **Keller, A.** (2020) 'On the lifetime of bioinformatics web services', *Nucleic Acids Research*, 48(22), pp. 12523–12533. Available at: <https://doi.org/10.1093/nar/gkaa1125>

- Kitchin, R., Collins, S. and Frost, D.** (2015) 'Funding models for Open Access digital data repositories', *Online Information Review*, 39(5), pp. 664–681. Available at: <https://doi.org/10.1108/OIR-01-2015-0031>
- Marín de la Iglesia, J.L. and Labra Gayo, J.E.** (2009) 'Doing business by selling free services', in M. Lytras, E. Damiani and P. Ordóñez de Pablos (eds.) *Web 2.0*. Berlin, Germany: Springer Science+Business Media, LLC.
- Mayernik, M.S.** (2023) 'Toward stronger coupling between technical infrastructures and institutional processes in data-intensive science', in X. Ma, M. Mookerjee, L. Hsu and D. Hills (eds.) *Recent advancement in geoinformatics and data science*. Boulder, Colorado: Geological Society of America, pp. 45–56. Available at: [https://doi.org/10.1130/2022.2558\(04\)](https://doi.org/10.1130/2022.2558(04))
- NWO.** (2021) *Research infrastructure: Call for proposals*. Available at: <https://www.nwo.nl/sites/nwo/files/media-files/WI-nc%20call%202021%20EN%20v5.00.pdf> (Accessed: 1 July 2024).
- OECD.** (2017) *Strengthening the effectiveness and sustainability of international research infrastructures* (No. 48). Paris, France: OECD Publishing. Available at: <https://doi.org/10.1787/fa11a0e0-en>
- O'Sullivan, A. and Sheffrin, S.M.** (2007) *Economics: Principles in action*. 2nd edn. Student edn. 2003c. London, England: Pearson.
- Pringle, D.** (2018) *How the science cloud could pay its way*. Available at: <https://sciencebusiness.net/science-cloud/news/how-science-cloud-could-pay-its-way> (Accessed: 1 July 2024).
- Regulation –2022/868- EN – EUR-Lex.** (2022) *EUR-Lex.europa.eu*. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868> (Accessed: 1 July 2024).
- Slack.** (no date) *Slack is your productivity platform | Slack*. Available at: <https://slack.com/> (Accessed: 5 July 2024).
- Smith, A. and Martin, C.** (2023) 'ELIXIR's long-term sustainability plan (2023)', *F1000Research*. Available at: <https://doi.org/10.7490/f1000research.1119412.1>
- StR-ESFRI.** (2019) *GUIDELINES ON COST ESTIMATION OF RESEARCH INFRASTRUCTURES*. Available at: https://www.esfri.eu/sites/default/files/StR-ESFRI2_STUDY_RIs_COST_ESTIMATION.pdf (Accessed: 1 July 2024).
- sustainable, adj.** (2024) Oxford Dictionary. Available at: https://www.oed.com/dictionary/sustainable_adj (Accessed: 5 July 2024).
- UKRI.** (2022) *Costs you can apply for – EPSRC – UKRI*. Available at: <https://www.ukri.org/councils/epsrc/guidance-for-applicants/costs-you-can-apply-for/#contents-list> (Accessed: 1 July 2024).
- Virapongse, A., Gallagher, J. and Tikoff, B.** (2024) 'Insights on sustainability of Earth Science Data Infrastructure Projects', *Data Science Journal*, 23(14), pp. 1–27. Available at: <https://doi.org/10.5334/dsj-2024-014>

TO CITE THIS ARTICLE:

Hooft, R.W.W. and Roos, M. 2025 Financing Models for Sustainable Data Reuse Infrastructure. *Data Science Journal*, 24: 29, pp. 1–18. DOI: <https://doi.org/10.5334/dsj-2025-029>

Submitted: 28 February 2025
Accepted: 30 September 2025
Published: 15 October 2025

COPYRIGHT:

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.