Sustainable Business Models for Data Repositories

Authors: Simon Hodson, Allan Dearry, Ingrid Dillo

Conference paper

Summary

Research has benefitted enormously from the digital revolution through the ability to communicate and analyse digital data. A key enabler, a *sine qua non*, of this development is a robust and sustainable data infrastructure. The benefits of open data are widely recognized but they pose challenges for how to fund and sustain common infrastructures, particularly those that need to expand considerably to meet the increasing demand for and availability of data, as well as demand for additional services.

Whether we like it or not, for data repositories to be sustainable, they need to have a robust business model which demonstrates clear alignment between, on the one hand, their mission, activities and services and, on the other, the needs of the community they serve, the objectives and policies of funders and the willingness and ability to pay of those entities that provide income. The robustness of income streams, the diversification of sources of income and cost restraint are elements to be taken into consideration.

This paper will provide a brief introduction to the session considering the economics of sustainable data repositories from a variety of perspectives and will do so in the context of a CODATA-coordinated OECD Global Science Forum Project on 'Sustainable Business Models for Data Repositories'. The session will not be a report on this project, but will contribute substantively to the work of this project by articulating and discussing various stakeholder perspectives on sustainable business models.

How do we ensure the sustainability of data infrastructure?

In line with the OECD Principles and Guidelines for Access to Research Data from Public Funding [http://www.oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm], national and international funders are increasingly likely to mandate open data and data management policies that call for the long-term stewardship of research data created as a result public funding. Such data should be 'Findable, Accessible, Interoperable and Reusable' and is essential both for research transparency and accountability and for reuse in broader integrative and transdisciplinary research. Data themselves are increasingly viewed as a fundamental component of research infrastructure and one which carries benefits for society and the economy.

Data stewardship is performed by a wide variety of data repositories, national and international in scope, covering a variety of domains, responsible for different volumes and types of data and with mission statements and user communities that require different services and a range of 'levels of curation'. Although many established national and international data repositories have core streams of income from research funders, these sources of income are generally inelastic and may be vulnerable (whether to short-termism, ill-considered re-prioritisation or attempts to pass responsibility to other budgets). There is a pressing need, therefore, to explore new and innovative income streams and to establish sustainable business models for data repositories.

A number of data repositories are exploring means of diversifying their income streams to increase sustainability and we are also witnessing the emergence of innovative not-for-profit (Dryad, Zenodo) [http://datadryad.org/; https://zenodo.org/] and for-profit (FigShare) [http://figshare.com/] enterprises seeking to provide generic data repositories. The economics of data services more generally – including value proposition, cost recovery, stakeholders' willingness to pay and overall business model – will need to be robust in order to be sustainable. It is also important when developing sustainable business model for data infrastructure to consider the value proposition, costs and benefits for different levels of stewardship provided, from lightweight curation to value-adding services. The scope will extend beyond the diversification of income streams and development of value-added services. It will include consideration of options to enhance operational efficiency of operation, including but not limited to automation or crowd-sourcing as alternative or complementary methods of curation.

The paper will introduce the context and objectives of the session, which will take the form of short and precise witness statements followed by a structured discussion. This brief framing paper will also introduce key concepts to be addressed in the session: what data repositories are in scope, what is meant by income stream, levels of curation etc.

By using the term 'witness statement', we intend to convey that the interventions will follow a precise, short and prescribed format and will be prepared in advance. The statements will describe an existing or possible business model and offer evidence and a considered view on the sustainability of that business model in terms of stakeholder acceptance (funder, depositors, users, institutions) and in terms of the appropriateness and fit of the business model to the repository's mission and 'level of curation'. The format of the statement will allow inter-comparison. As they will be short and prepared in advance, this will leave time for a structured, moderated discussion of the evidence provided. This will provide important material for the OECD project as well as compiling substantive insights into repository sustainability of wider interest to SciDataCon participants.

Acknowledgements

Competing Interests

The authors are co-chairs of the OECD Global Science Forum project on Sustainable Business Models for Data Repositories.