Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

10-Nov-2015

_____

English - Or. English

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INNOVATION
COMMITTEE FOR SCIENTIFIC AND TECHNOLOGICAL POLICY

Global Science Forum

SUSTAINABLE BUSINESS MODELS FOR DATA REPOSITORIES

OECD Headquarters, 26-27 November 2015

Contact: Carthage Smith (E-mail: carthage.smith@oecd.org)

SUSTAINABLE BUSINESS MODELS FOR DATA REPOSITORIES

Prepared by Ingrid Dillo, Simon Hodson, Anita de Waard with input from Carthage Smith (this proposal has been prepared following the request at the 32nd meeting for the Secretariat to prepare a preliminary proposal on business models for data infrastructure).

The Project

1.1 The Challenge

1.   In line with the *OECD Principles and Guidelines for Access to Research Data from Public Funding*,[1] 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'[2] and is essential both for research transparency and accountability and for reuse in broader integrative and transdisciplinary research. Data themselves are increasingly viewed as an essential part of research infrastructure and one which carries benefits for society and the economy.[3]

2.   As a result, data infrastructure is required to curate rapidly increasing volumes of data, which are becoming more complex over time. With the volume and variety of data increasing exponentially, budgets for data stewardship will struggle to keep pace despite falling storage costs.

3.   Data stewardship is performed by a wide variety of data repositories, national and international in scope. Although many established national and international data repositories have reliable sources 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). Some 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)[4] and for-profit (FigShare)[5] enterprises seeking to provide generic data repositories. The European Open Science Cloud for Research[6] aims to provide an important shared infrastructure, including a data stewardship component, as a publicly-funded Commons. The economics of such a service – including value proposition, cost recovery, stakeholders' willingness to pay and overall business model – will need to be robust in order to be sustainable.

---

[1] http://www.oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm

[2] E.g. see https://www.force11.org/group/fairgroup

[3] See http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf and https://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20Final.pdf

[4] Dryad http://datadryad.org/; Zenodo https://zenodo.org/

[5] FigShare http://figshare.com/

[6] European Open Science Cloud for Research, announcement by EUDAT project and other stakeholders http://eudat.eu/node/612

4.	There is a pressing need, therefore, to explore new and innovative income streams and to establish sustainable business models for data repositories.

1.2 Identifying and testing sustainable business models for data repositories

5.	The project as a whole will identify a sample of emerging business models for data repositories and, through focus groups and work-shops, test the potential viability of these models with key stakeholders. The final output will be a report summarizing these findings and presenting sustainable business models on the basis of current and innovative income streams. Such a report will be of considerable value to OECD member countries and all countries struggling to understand how to establish sustainable income streams for data infrastructure and the stewardship services provided by data repositories.

6.	The focus will be on testing income streams against willingness to pay to construct viable business models. Existing work on the costs of digital curation will be explored, as will the broader value proposition of data repositories, the benefits and return on investment created. This helps provide the context for the development of sustainable business models. The prime objectives are to cast light on the following issues:

1.	How data repositories are currently funded;

2.	What additional, innovative, income streams are available to data repositories;

3.	How various income streams may fit together into a business model;

4.	How various business models match budgetary structure and the willingness and ability to pay of the various stakeholders.

7.	The project builds on the existing work of an RDA-WDS Working Group, while increasing substantially the scope of that study by including interesting case studies from GSF member countries. Above all, the OECD GSF provides the ideal setting for exploring more broadly these objectives and in particular objective 4.

1.3 Current Work on Data Repository Income Streams

8.	The project builds on the firm basis of current research into data repository income streams. The starting point is a study conducted by a Working Group co-sponsored by the Research Data Alliance (RDA) and the ICSU World Data System (WDS) and which was initiated in September 2014.[7] The group is co-chaired by Simon Hodson, executive director of CODATA, Anita de Waard, VP Research Data Collaborations at Elsevier, and Ingrid Dillo, deputy director of DANS and draws on the contributions of some 50 members.

9.	The core of the activities of the group consisted of a series of in-depth interviews with some 25 data repositories around the globe, which were analysed with the help of staff of the Center for Quantitivate Social Sciences at Harvard. The interviews were designed to build a picture of current income streams, explore what new income streams were being trialed or considered and assess the data repositories' attitudes to diversification and views on sustainability issues. The emergent findings were presented and discussed periodically in virtual meetings of the group and at the Plenary Meetings of the RDA that were well-attended by stakeholder representatives including leadership of data centres, funders,

---

[7] https://rd-alliance.org/groups/rdawds-publishing-data-cost-recovery-data-centres.html

publishers and policy makers. The group identified the most significant income streams and developed a typology of the various business models encountered and summarized this in a draft report presented at RDA Plenary 6 in Paris, September 2015.

10. The group found that roughly half the data repositories surveyed were largely funded through block grants or 'structural funding'. Such ongoing commitments from research funders have undoubted benefits but are potentially vulnerable and can be inflexible. Many data repositories were also substantially dependent on short term project funding relating to research, training or infrastructure development activities. Moreover, a significant number of data repositories were exploring or considering exploring alternative income streams. These include data deposit charges and selling curation and preservation services of various sorts to government, public and private institutions.

11. Some repositories already charge for value-added services, and a number of others are exploring ways in which this can be done while conforming to Open Access principles. A small number of repositories had completely transformed their business model. New more generic, not-for-profit and commercial data repositories have emerged and may transform the landscape. Research institutions are increasingly interested in providing data repository services and this provides a market for organizations with existing expertise in data stewardship.

12. At the Paris RDA meeting, the Group conducted a SWOT (Strength-Opportunities-Weaknesses-Threats) analysis of a number of business models. These findings, augmented with preliminary conclusions and recommendations for further work, will be incorporated into a final report published prior to the RDA Plenary 7 in Japan in March 2016.

1.4 From analysis to policy: the OECD GSF Project

13. The RDA project described above provides a firm basis for further work leading to specific policy advice. The present proposal proposes to extend the survey and analysis work that has been conducted thus far. Above all, the OECD GSF project is designed: 1) to identify further emerging and innovative income streams; 2) to test more rigorously possible business models with various stakeholders and against budgets and ability to pay; and 3) on the basis of these findings make policy recommendations to promote sustainable business models for data infrastructure.

14. Specific tasks include the following:

- Extend the number of data repositories analysed to include a broader number of OECD member countries and to increase coverage non-European data repositories.

- Conduct a focus group on innovative income streams.

- Extend the number of potential business models.

- Perform a thorough economic analysis of the business models, assessing their viability and the value of the services offered and exploring the alignment of these with stakeholder and beneficiaries' willingness to pay and budgetary authority. This task will be conducted in two stages. First, an economic analysis and presentation of business models will be conducted by appropriate experts. Second a focus group meeting will test the business models with stakeholders and refine the analysis.

- On the basis of these findings, policy recommendations to promote sustainable business models for data infrastructure will be formulated with input from the GSF.

2. Proposed participants

15.   The project will be carried out under the aegis of GSF in partnership with CODATA, the Research Data Alliance and the World Data System. A joint Expert Group, to oversee the project activities and outputs, will be established with nominations from GSF delegates and from each of the three organizations name above. The Expert Group will include representatives from the various stakeholders involved in funding decisions relating to data infrastructures. This includes representatives of funding organizations and science policy makers, data repositories, research performing organizations, and the academic community.

16.   To undertake the proposed economic analysis, experts with specific experience in the economic analysis of data infrastructures will be co-opted. Experts who may be considered to perform this role include: John Houghton (author of a number of studies on the Return on Investment of data infrastructures), Brian Lavoie, OCLC (one of the authors of the Blue Ribbon Task Force Report on Sustainable Digital Preservation and Access)[8], Micah Altman (who has experience in analyzing economic models pertaining to scholarly communications).

3. Relevance to GSF Themes

17.   This proposal is directly relevant to the GSF Strategic Priority themes 1. strengthening the science enterprise and 2. International cooperation in science. Ensuring sustainable data infrastructure is also critical for science to address global societal challenges (GSF priority theme 3).   The proposed partnership with leading international organisations to implement this project is in line with GSF's declared intention to build on other efforts and work in partnership.

18.   Funding and business models for data infrastructures were highlighted as important areas for science policy in the recent CSTP Ministerial meeting in Daejeon, 20-21 October, 2015.

19.   This study is complementary to the recently initiated GSF work on the Sustainability of International Research Infrastructures and the proposed project on International Cooperation and Coordination of Cyber-infrastructure.

4. Feasibility, Resource, Timeframes

20.   The project described here can feasibly be completed in 12 months, building on existing work and engaging with clearly identified stakeholders. The project will be conducted from 1 April 2016 to 31 March 2017, divided over quarterly reviewed tasked as described below.

21.   Q1 April-June 2016: Project set up; initiate virtual meetings of Expert Group; identification and conduct of additional data repository interviews (including those identified by GSF delegates).

22.   Q2, July-September 2016: complete additional data repository interviews; conduct and write up focus group on innovative income streams; refine presentation of business models and initiate economic analysis of these models. [Make progress report to GSF in October, 2016]

23.   Q3 October-December 2016: finalise economic analysis of business models; present and test these in focus group workshop with stakeholders (including GSF delegates).

---

[8] http://blueribbontaskforce.sdsc.edu/

24.   Q4 January-March 2017: complete draft of final report and the findings of the study and workshops; formulate policy recommendations on the basis of the project's findings and iterate these with the Expert Group.

25.   March/April, 2017: final report, with findings and policy recommendations on innovative income streams and sustainable business models for data repositories, presented to GSF for approval prior to publication.

5. Resource Requirements

26.   The resource allocations and requirements are as follows:

Item	Resource	Timing
Expert Group	In kind support	Establish April 2016
Consultant (draft texts, facilitate workshops)	~€15k	Appoint April 2016
Consultant (economics expert)	~€20k	Appoint June 2016
One-day focus group on innovative income streams [Poss hosted by GSF member country]	~€5k to cover key stakeholders, plus in kind support	July 2016
Two-day workshop to test business models with stakeholders [Poss. hosted by GSF member country]	~€10k to cover key stakeholders plus in kind support	Oct-Nov 2016

6. Novelty and relationship with other OECD activities

27.   The *OECD Principles and Guidelines for Access to Research Data from Public Funding*,[9] have had a considerable and positive impact on policy for research and data policies in North Atlantic countries and Australia. However, the funding models for data infrastructures and data repositories remain uncertain. As OECD Countries increasingly look to Open Science[10] as a means to advance knowledge more rapidly, to increase the benefits and return of investment in research and to foster innovation, the sustainable funding of the data infrastructure necessary for Open Science needs to be addressed. This work directly addresses key recommendations in the recent STI-TIP report on Open Science. It could also contribute to a cross-cutting STI activity on Digitalisation of Science and Innovation that is likely to be carried out in 2017-2018.

7. Relevance to GSF Members

28.   Policies to promote effective open science were identified by all GSF members as being of the highest priority during the GSF strategic scoping exercise that was completed in April 2015. Members specifically requested that the Secretariat develop a proposal on funding policies for data infrastructures at the 32nd GSF meeting.

---

[9] http://www.oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm

[10] https://www.innovationpolicyplatform.org/content/open-science

29.   The proposed description and economic analysis of existing and emerging business models is expected to be of direct relevance to all Member countries.

8. Planned follow-up and dissemination strategy

30.   The project could be launched with an activity at the European Conference on Open Data, sponsored by the Dutch Presidency of the EU, planned for 4 April 2016.

31.   A consultation on the interim outcomes of the project will be held at International Data Week,[11] the combined conference of CODATA, the Research Data Alliance and the World Data System in September 2016.

32.   Following approval by GSF, The final report will be presented at the RDA Plenary 9 in Barcelona in Spring 2017.

33.   Opportunities to disseminate and launch the report at meetings involving OECD, G8 and G77 science ministers will be explored.

---

[11] http://internationaldataweek.org/