



SUSTAINABLE BUSINESS MODELS FOR DATA REPOSITORIES

Ingrid Dillo, DANS
WDS Data Repositories Day, Gaborone, 09-11-2018



Sustainable business models for data repositories

Increasing need for data repositories and data stewardship.

- Increasing volume presents a challenge.
- Growing requirements for data stewardship present an even greater challenge.

Sustaining digital data infrastructure is a major issue for science policy

- current funding models will prove inelastic and not meet the growing requirements – concern on the part of repositories and funders





The cost of long-term preservation

Digital Preservation Costing Initiatives

- LIFE 1, 2 and 3. Projects to explore digital preservation costing, and develop costing models.
 - <http://www.life.ac.uk/>
- Cost Model for Digital Preservation (CMDP): Project at the Royal Danish Library and the Danish National Archives to develop a new cost model. Currently covers Planning, Migrations and Ingest
 - <http://www.costmodelfordigitalpreservation.dk>
- Keeping Research Data Safe 1 and 2 (KRDS): Cost model and benefits analysis for preserving research data
 - <http://www.beagrie.com/krds.php>
- Presto Prime cost model for digital storage
 - <http://prestoprime.it-innovation.soton.ac.uk/>
- Cost Estimation Toolkit (CET): Data centre costing model and toolkit, from NASA Goddard
 - http://www.pv2007.dlr.de/Papers/Fontaine_CostModelObservations.pdf
- Cost Model for Small Scale Automated Digital Preservation Archives (Strodl and Rauber)
 - http://www.ifs.tuwien.ac.at/~strodl/paper/strodl_jpres2011_costmodel.pdf
- APARSEN Project activity focused on digital preservation costing
 - <http://www.alliancepermanentaccess.org/index.php/knowledge-base/digital-preservation-business-models/costbenefit-data-collection-and-modelling>
- EPSRC and JISC study on Cost analysis of cloud computing for research
 - http://www.jisc.ac.uk/media/documents/programmes/research_infrastructure/costcloudresearch.pdf
- Cost forecasting model for new digitization projects (Excel and web tool under development) (Karim Boughida, Martha Whittaker, Linda Colet, Dan Chudnov)
 - http://www.cni.org/wp-content/uploads/2011/12/cni_cost_boughida.pdf
- DP4lib is developing a business and cost model for a digital preservation service
 - <http://dp4lib.langzeitarchivierung.de/downloads/DP4lib-One-Page-08-eng.pdf>
- DANS Costs of Digital Archiving Volume 2 Project, focusing on preservation and dissemination of research datasets
 - <http://www.dans.knaw.nl/en/content/categorieen/projecten/costs-digital-archiving-vol-2>
- Blue Ribbon Task Force on Sustainable Digital Preservation and Access
 - <http://brtf.sdsc.edu/>
- Economic Sustainability Reference Model (draft)
 - <http://4cproject.eu/community-resources/outputs-and-deliverables/ms9-draft-economic-sustainability-r>
- ENSURE Project - Enabling Knowledge Sustainability Usability and Recovery for Economic value
 - <http://ensure-fp7-plone.fe.up.pt/site/>
- 4C. EU funded project on costing, led by JISC.
 - <http://4cproject.eu/>
- Cost Model for Electronic Health Records (Bote, Fernandez-Feijoo, and Ruizb)
 - <http://www.sciencedirect.com/science/article/pii/S2212017312004434>
- TCP: Total Cost of Preservation (California Digital Library)
 - <https://wiki.ucop.edu/display/Curation/Cost+Modeling>
- Cost model for digital preservation (National Archives of the Netherlands)
 - http://dlmforum.typepad.com/Paper_RemcoVerdegem_and_JS_CostModelfordigitalpreservation.pdf



4C Curation Costs Exchange
Understanding and comparing digital curation costs to support smarter investments

Home About Understand costs Compare costs Read more Vendor services Help Sign Up Sign In

All about the costs of curation

What am I spending, what are they spending, what should we be spending?

Understand costs
Assessing your costs and using cost models to make smart investments

Compare costs
Add your curation costs and see how they compare with others



<http://www.4cproject.eu>



RDA Interest Group

- RDA Cost Recovery Interest Group, also supported by WDS and CODATA
- Report *Income Streams for Data Repositories*
(Feb 2016; <https://zenodo.org/record/46693#.WTUR-TOB2T8>)
- Landscape survey of 25 data repositories
- Identified major income streams and funding structure
- Typology of income streams
- SWOT analysis of income streams at RDA workshop in September

Typology of income streams



- 1) Structural (central contract)
- 2) Hosting Support (indirect or direct support through institutional hosting)
- 3) Annual Contract (from depositing institution)
- 4) Data Deposit Fee (may be paid by researcher, RPO or publisher; may originate with funder)
- 5) Access Charge (for the data or for value-adding services)
- 6) R&D Projects (to develop infrastructure or value-adding services)
- 7) Private Contracting (services to parties other than core funder)





OECD/GSF Expert Group

- OECD provided an ideal setting for this work
- Continuation of the work under the umbrella of OECD/GSF
- International expert group comprising nominees from GSF delegates and from RDA, WDS and CODATA

- Around 50 interviews in total
- Thorough economic analysis
- Innovative funding models
- Cost optimization

- Stakeholder workshops:
 - To identify and appraise possible innovative income streams and to identify approaches to cost restraint.
 - To test possible business models with stakeholders



OECD



Report with recommendations

OECD (2017), "Business models for sustainable research data repositories", *OECD Science, Technology and Industry Policy Papers*, No. 47, OECD Publishing, Paris

<https://doi.org/10.1787/302b12bb-en>.





Five high level recommendations

Recommendation 2: All research data repositories should have a clearly articulated business model.

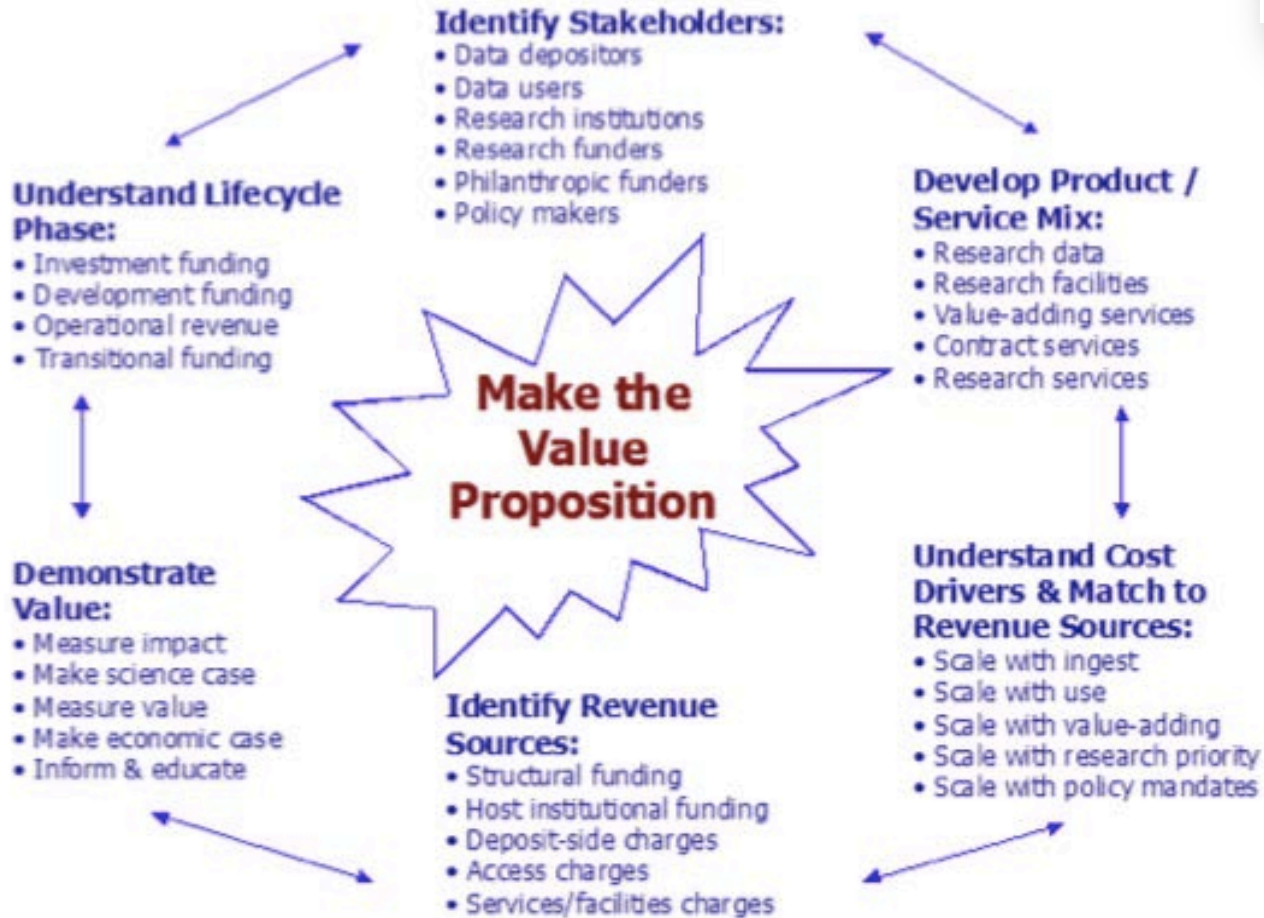
Actions needed to develop and maintain a successful business model include :

- * *Understanding the lifecycle phase of the repository's development* (e.g. the need for investment funding, development funding, ongoing operational funding, or transitional funding).
- * *Developing the product/service mix* (e.g. basic data, value-added data, value-added services and related facilities, or contract and research services).
- * *Understanding the cost drivers and matching revenue sources* (e.g. scaling with demand for data ingest, data use, the development and provision of value-adding services or related facilities, research priorities, and policy mandates).
- * *Identifying revenue sources* (e.g. structural funding, host institutional funding, deposit-side charges, access charges, and value-added services or facilities charges;
- * *Identifying who the stakeholders are* (e.g. data depositors, data users, research institutions, research funders, policy makers).
- * *Making the value proposition* to stakeholders (e.g. measuring impacts and making the research case, measuring value and making the economic case, informing, and educating).

Because the context is dynamic, these actions should be revisited regularly throughout a data repository's lifecycle.



Figure ES.1. Elements of a research data repository business model





Thank you for listening!



ingrid.dillo@dans.knaw.nl
www.dans.knaw.nl