Objective

• Originally planned for Plenary 15, the RDA Adoption week aims to **demonstrate the wide variety of RDA adoptable and adopted solutions to data sharing challenges** that people in the field encounter in their daily jobs.

Purpose of the week:

• Learn about RDA Outputs
• Converse with speakers from all around the world who have created and implemented them
• Determine how best to integrate those data sharing solutions into your own projects
• RDA Outputs are classified as RDA Recommendations (official, endorsed results of RDA Groups), Supporting Outputs (useful solutions from our RDA Working and Interest Groups) or other Outputs

• They can be searched according to their status, Data Life Cycle topics or scientific domain
• Business models for sustainable research data repositories: an OECD-CODATA Report. Adoption resulting from the IG Publishing Data Cost Recovery for Data Centres - Simon Hodson (CODATA, ICSU)

• A curriculum for foundational Research Data Science skills for Early Career Researchers - Robert Quick (Indiana University) and Sara El jadid (Ibn Tofail University)

• Learning to share: a road to national data publishing policy - Edit Görögh (University of Debrecen) and András Holl (Hungarian Academy of Sciences)

Poll

Join at

slido.com

#W350
Tell your adoption story

• Are you an adopter? RDA is actively seeking new adoption stories to inspire the further uptake of RDA outputs.

• Submit your story here: https://www.rd-alliance.org/tell-your-rda-adoption-story
CODATA CfP Data Science Journal

• RDA special collection themes:
  o Results produced by an IG or WG;
  o Description of an Adoption Case outlining how a specific recommendation or output has been implemented;
  o Other types of work related to RDA activities.

• RDA Europe 4.0 still has funds available for the publication of articles in DSJ

• Open to all interested applicants regardless of their geographical provenance.

• Deadline 17 July
OECD-CODATA Report: Business models for sustainable research data repositories

Simon Hodson, Executive Director, CODATA
www.codata.org
OECD-CODATA Report: Business Models for Sustainable Data Repositories

- Started as a WDS Working Group on Cost Models for Data Repositories
- Became an RDA-WDS Interest Group on Income Streams for Data Repositories
- Developed a proposal for a project under the OECD GSF.
- Funding from OECD GSF, additional funding and support from CODATA.
Business models for sustainable research data repositories (2017):
https://doi.org/10.1787/302b12bb-en
The Challenge: Business Models for Sustainable Data Repositories

- Sustaining digital data infrastructure is a major issue for science policy!
- Genuine concern that current funding models will prove inelastic and not meet the growing requirements – concern on the part of repositories and funders.
- Important to demonstrate the value proposition of data repositories / data services.
- Sustainability is not just about whether something is funded, but how it is funded: what are the most effective and sustainable mechanisms for funding?
- Relatively little work done on sustainability, on income streams, on business models.
Business Models for Sustainable Data Repositories: 
Central Questions

- How are data repositories currently funded?
- How do the different mechanisms for funding (income streams) affect sustainability?
- What innovative income streams are available to data repositories?
- What means of optimising costs are available?
- How do income streams match willingness/ability to pay of various stakeholders?
- How do income streams/willingness to pay fit together into a sustainable business model?
Business Models for Sustainable Data Repositories:

Project Team

- Co-Chairs of OECD Project: Ingrid Dillo, DANS; Simon Hodson, CODATA.
- Consultant and writer: Paul Uhlir, Chair of CODATA Data Policy Committee, previously National Academies, USA.
- Economics consultant: John Houghton, author of many studies of economics of data infrastructure.
- OECD GSF Expert Group.
- Invited participants of November 2016 and March 2017 Workshops.
Project Activities and Programme

- **Detailed interviews** with 47 data repositories funding and business models, conducted by members of the Expert Group and by OECD staff.
- **First workshop, Nov 2016**: SWOT analysis of different business models, discussion of cost restraint.
- **Economic analysis, Nov 2016-Feb 2017**: conducted by John Houghton and discussed with the project team, co-chairs and Expert Group. Identified strengths and weaknesses of particular revenue streams, frameworks for business models and recommendations. Draft report circulated prior to workshop.
- **Second workshop, March 2017**: presentation of economic analysis, discussion of strengths and weaknesses of different business models and of draft recommendations with stakeholders.
- **Redrafting of report, April-July 2017**: final iterations and refinement of report and recommendations.
What are we talking about?
Research Data Repositories

- Repositories preserve, manage, and provide access to many types of digital materials in a variety of formats. Materials in online repositories are curated to enable search, discovery, and reuse. There must be sufficient control for the digital material to be authentic, reliable, accessible and usable on a continuing basis.


- Repositories curate data (making it findable and usable), provide access to data and long term stewardship. Some repositories, additionally, perform activities of data refinement and enhancement or provide various analytic services over the data. **However, the sine qua non is the stewardship function.** Aggregator services that do not curate the data are out of scope.
Business Models and Revenue Sources

- **Business model**: ‘A plan for the successful operation of a business, identifying sources of revenue, the intended customer [user] base, products, and details of financing.’ [https://en.oxforddictionaries.com/definition/business_model](https://en.oxforddictionaries.com/definition/business_model)

- Some repositories are heavily reliant on a single revenue source, such that the repository’s business model can be examined largely in terms of that revenue source.

- Many repositories combine two or more revenue sources into their business models.

- The report discusses the major revenue sources and the most significant combinations, and analyse each as a business model.
Key Contributions of the Report

- Summary of findings from 47 repositories, SWOT analyses.
- Typology of revenue sources.
- Discussion of value propositions for research data repositories.
- Criteria to assess business models for research data repositories.
- Set of five detailed recommendations for policymakers and funders in OECD member states, as well as repository operators and managers.
Typology of Revenue Sources

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) Projects (to develop infrastructure or value-adding services)
7) Private Contracting (services to parties other than core funder)
Identify Stakeholders:
- Data depositors
- Data users
- Research institutions
- Research funders
- Philanthropic funders
- Policy makers

Develop Product / Service Mix:
- Research data
- Research facilities
- Value-adding services
- Contract services
- Research services

Identify Revenue Sources:
- Structural funding
- Host institutional funding
- Deposit-side charges
- Access charges
- Services/facilities charges

Understand Cost Drivers & Match to Revenue Sources:
- Scale with ingest
- Scale with use
- Scale with value-adding
- Scale with research priority
- Scale with policy mandates

Understand Lifecycle Phase:
- Investment funding
- Development funding
- Operational revenue
- Transitional funding

Demonstrate Value:
- Measure impact
- Make science case
- Measure value
- Make economic case
- Inform & educate

Make the Value Proposition
Box 4.1. Criteria to assess the sustainability of research data repository business models

There are a number of criteria by which one can assess whether a research data repository business model may be sustainable. These include:

- Does the business model make sense economically - is it a recognisable model from underlying economic principles?
- Does the model provide a source, or sources, of revenue that scale with demand and costs?
- What are the mechanisms or factors in each business model that help make it responsive, or not, to requirements?
- Which stakeholders benefit from the data being freely available, and how?
- What effects are different repository business models likely to have on data sharing and use?
- What are the mechanisms in each business model that ensure that any paying stakeholders have a clear understanding of the repository’s value proposition?
- Are some business models more suitable for certain types of repositories than others?
- Are there optimal mixes of revenue sources for various types of data and types of repositories?
- Are there ways of grouping or dividing data and related services to better match single or combinations of revenue sources?
- What are the incentives within various business models for cost constraint?
- What are the mechanisms in the business model that might encourage cost optimisation?

These criteria form the basis for the following analysis.
<table>
<thead>
<tr>
<th>Funding Source</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural funding</td>
<td>• Compatible with open data principles.</td>
<td>• Fixed, multi-year may not scale easily.</td>
</tr>
<tr>
<td></td>
<td>• Longer-term stability.</td>
<td>• Competes with research funding.</td>
</tr>
<tr>
<td></td>
<td>• Larger-scale and efficiencies.</td>
<td>• Too many eggs in few baskets.</td>
</tr>
<tr>
<td></td>
<td>• Flexible as to allocation.</td>
<td></td>
</tr>
<tr>
<td>Host or institutional funding</td>
<td>• Compatible with open data principles.</td>
<td>• Limited purview, with focus on local community.</td>
</tr>
<tr>
<td></td>
<td>• Longer-term stability.</td>
<td>• May lead to fragmentation of domain data and lower interoperability.</td>
</tr>
<tr>
<td></td>
<td>• Efficiencies through sharing services.</td>
<td>• Limited incentive to add value to data and develop related services.</td>
</tr>
<tr>
<td></td>
<td>• Close to researchers (customers).</td>
<td></td>
</tr>
<tr>
<td>Annual deposit-side contract</td>
<td>• Compatible with open data principles.</td>
<td>• Unpredictable year-to-year and involves high transaction costs (managing contracts).</td>
</tr>
<tr>
<td></td>
<td>• Demand oriented and scales with demand (data ingest).</td>
<td>• Limited engagement with, or focus on, data users.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• May lead to fragmentation and loss of scale if many players, or monopolies if few players.</td>
</tr>
</tbody>
</table>
| Data deposit fees | • Compatible with open data principles.  
• Demand oriented and scales with demand (data ingest).  
• Researchers price sensitivity ensures cost constraint.  
• Open data is part of research and its funding. |
|------------------|----------------------------------------------------------------------------------|
|                  | • Cost disincentive to depositing, so depends on strong mandates.  
• May lead to low level of curation to contain costs (price).  
• May be difficult for repository to compete for deposits with comparable repositories that do not charge. |
| Data access charges (subscriptions or use fees) | • Users pay for what they want, so funding reflects value.  
• More market-oriented approach may provide incentive for cost constraint. |
|                  | • Not compatible with open data principles and many funder mandates, limiting the potential market size.  
• Charges limit use and will reduce the value of data.  
• Revenue scales with use and not ingest or curation costs.  
• Vulnerable to funding cuts. |
| Contract services and project funding | • Compatible with open data principles.  
• Can increase contact between staff and clients.  
• Potential for innovation. |
|                  | • Short-term and not a sustainable sole revenue source.  
• High transaction costs, chasing money.  
• Revenue does not scale with data ingest or use (costs).  
• Inflexible as to allocation of funding. |
| Diversification of revenue sources | • No single source of failure.  
• Can maintain compatibility with open data principles.  
• Flexible and enables experimentation with new services. |
|                  | • May lead to higher transaction costs (managing multiple funding sources).  
• May lead to Mission drift. |
Business Models for Sustainable Data Repositories: Recommendations

The following policy recommendations primarily target policymakers and funders in OECD member states, as well as repository operators and managers.

1. All stakeholders should recognise that research data repositories are an essential part of the infrastructure for open science.
2. All research data repositories should have a clearly articulated business model.
3. Policy makers, research funders, and other stakeholders need to consider the ways in which data repositories are funded, and the advantages and disadvantages of various business models in different circumstances.
4. Research data repository business models are constrained by, and need to be aligned with, policy regulation (mandates) and incentives (including funding).
5. In the context of financial sustainability, opportunities for cost optimisation should be explored in order to be able to effectively manage digital assets over time.
Thank you for your attention!

Simon Hodson
Executive Director, CODATA
www.codata.org
simon@codata.org
@simonhodson99 ; @CODATAnews
All stakeholders should recognise that research data repositories are an essential part of the infrastructure for open science.

Research data repositories provide for the long-term stewardship of research data, thus enabling verification of findings and the re-use of data. They bring considerable economic, scientific, and social benefits. Hence, it is important to ensure the sustainability of research data repositories. Sustainability depends, inter alia, on a clearly articulated value proposition and the development of a 'business model' (See Recommendation 2).

- Policy makers and research funders should take a strategic view of the data landscape and seek to ensure the appropriate provision of repositories. They can do this by ensuring that the researchers they fund have access to suitable and sustainable research data infrastructure, so that the research community can meet expectations for data preservation and sharing, and comply with open data mandates.
- Research data repository operators and managers need to study and understand the value proposition of their repositories, and clearly articulate it for all stakeholders in the research system.
- Research data repository operators and managers should continually review their business model as the repository evolves, and revise it accordingly.

Recommendation 1: research data repositories are an essential part of the infrastructure for open science
All research data repositories should have a clearly articulated business model. Actions needed to develop and maintain a successful business model include (Figure ES.1):

- 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); and
- 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 the data repository's lifecycle.
Policy makers, research funders, and other stakeholders need to consider the ways in which data repositories are funded, and the advantages and disadvantages of various business models in different circumstances.

It is important to consider what system of allocation will best ensure that the optimal level of funding will be made available for research data repositories. For example:

- **Structural funding** typically involves a trade-off between funding for data repositories and funding for other research infrastructure or for research itself. That allocation will best be made by informed actors making choices, such as through a funding allocation process involving widespread research stakeholder participation, expert consultation, and 'road-mapping'.

- Funding models depending on deposit or access fees bring the trade-off closer to the researchers, but their success in optimising allocation will depend on the extent to which the actors are informed and on their freedom of choice. The latter may be constrained by open data mandates (regulation).

- **Host institutional funding** may divorce informed actors from the funding decisions or require additional processes to ensure greater stakeholder understanding of the value of the repository services.

**Recommendation 3 : the advantages and disadvantages of various business models to be considered**
Project funding often provides a mechanism to test the need for a data repository and the initial capacity to create one. However, as the repository matures and scales to provide an ongoing, reliable and quality service, a different funding model is likely to be needed.

- From an economic perspective, this is the distinction between *investment funding* to establish a business, and an *ongoing revenue source* during the operational phase.
- This distinction is not yet well made in the research data repository environment, but should form an important part in the design and evolution of repository business models.

Research data repository costs will change over time. As the global data repository infrastructure evolves there will be increasing learning and scale economies, which have the potential to reduce repository costs, although this needs to be balanced against increased data flows.

- Consequently, policy makers and funders should be wary of allocating a fixed percentage of research funding for research data repository infrastructure, as it would be very difficult to establish the appropriate level and very difficult to change it once established.
- The allocation of funds is likely to be better made when left up to those closest to their application (e.g. allocating funding to research and letting researchers and research managers meet open data requirements as best suits their needs).

**Recommendation 3 (cont.)**
Research data repository business models are constrained by, and need to be aligned with, policy regulation (mandates) and incentives (including funding).

- Policy makers should be cautious of 'un-funded mandates'. They should combine regulation and incentives thoughtfully to achieve best results.
- Some business models depend on willingness to fund the repository in recognition of a strong value proposition (e.g. structural or host funding). Other business models are heavily dependent on strong policy incentives and regulation (e.g. deposit-side charges). Still other business models may limit data re-use and reduce the overall benefits that could be derived from research data curation and sharing (e.g. access-side charges).
- A key issue is matching funding and revenue sources to the main cost drivers, to ensure that revenue scales with demand and repository costs. These cost drivers can relate to the level of activities (e.g. deposits, access, and use), and/or to the level of curation (e.g. basic versus enhanced).
Recommendation 5: need to explore cost optimisation

In the context of financial sustainability, opportunities for cost optimisation should be explored in order to be able to effectively manage digital assets over time. Therefore, policy makers, research funders, and repository managers should:

▪ Obtain greater clarity concerning costs, in order to fully understand and manage them.
▪ Consider cost optimisation system-wide (throughout the whole data lifecycle), rather than simply focus on cost savings at the repository level, as there is a risk that repository cost saving may only lead to cost shifting.
▪ Consider the effect a funding model has on cost constraints, as the more a funding model depends on or creates low price elasticity of demand, the lower the incentive for cost constraints will be.
▪ Monitor the research landscape for emerging opportunities. As data repository activities grow and develop, there will be increasing opportunities to buy services from specialist providers, potentially enabling greater cost optimisation.
▪ Take advantage of economies of scale. For example:
  ▪ By encouraging or funding the establishment of lead organizations at the national level, and encouraging those organizations to collaborate globally.
  ▪ By encouraging or funding collaboration and federation. Not all research data repositories need to perform specialised curation and preservation tasks. Similarly, not all institutions or organizations need to create individual repositories. Collaboration and federation can help to manage and reduce costs.
OECD Global Science Forum Project:
Business Models for Sustainable Data Repositories

- Relatively little work has been done on the economics and business models of data infrastructure.
    ➢ Need to understand value proposition for communities.
    ➢ Need to understand how repositories are funded.
A curriculum for foundational Research Data Science skills for Early Career Researchers

Sara El Jadid and Rob Quick
Slides Provided by Hugh Shanahan

rquick@iu.edu
eljadidsara@gmail.com

RDA Adoption Week
2020-06-19

@RobQuick5
Gaps

Identified gap in skills for Early Career Researchers (ECRs)

Opportunity/Impact - LMICs

Practices for what comes next (FAIR, Open Data, Open Science)

Isolated ECR’s
Data Science Skills

Responsible Research Practices

Research Data Science

CODATA-RDA Schools

Building Communities
Objectives

Short-Medium term
Run expanding series of schools in *Research Data Science* regionally across the world to ECR’s

Long term
Curriculum (or close to) gets taught by HEI’s globally to ECR’s
CODATA-RDA schools

Joint effort between CODATA and RDA

CODATA - Committee on Data of the International Science Council (ISC).

"CODATA exists to promote global collaboration to advance Open Science and to improve the availability and usability of data for all areas of research. “
CODATA-RDA schools

Two week event

Focus on ECR’s from LMIC’s

Data Science + Responsible Research skills
= Research Data Science

NOT another bootcamp!
Curriculum

- Software Carpentry
- Computational Infrastructures
- Visualisation
- Research Data Management
- Open and Responsible Research
- Author Carpentry
- Analysis
- Information Security
Principles

Reuse materials - it’s out there, use it.

Materials available under open source licence.

Mix ‘sychronised typing’ with seminars with discussion with team work.

Trieste, Italy is a test bed - other schools are regional

Use as many regional instructors as possible.

Recruit from student body.
Trieste, Aug 2016
July 2017
Aug 2018
Aug 2019
Aug 2020 (Virtual)

Kigali, Oct 2018
San José, Dec 2019

São Paulo, Dec 2017
Dec 2018
Dec 2020 (Virtual?)

Addis Ababa, Jun 2019

Brisbane, Jun 2018

Pretoria, Jan 2020

Dec 2020 (Virtual?)
State of play

By February 2020 we have delivered ....

9 Two Week Schools on 4 continents

~400 students taught from 40 countries

In British figures corresponds to ~£250K of teaching

ALL AS A VOLUNTEER NETWORK
Data Steward strand

Partnership

“The FAIRsFAIR project addresses, in a 36 months timeplan, the development and concrete realisation of an overall knowledge infrastructure on academic quality data management, procedures, standards, metrics and related matters, based on the FAIR principles.”

Run three instances of the school in Europe

Train-trainer events
Idea

Add a Data Steward strand to school.

Data Steward students work with ECR but have their own modules as well.
Remember

Most of all what’re doing here is

Changing a culture

Getting people to work together

Building communities that follow an approach

That is really, really, hard - but more effective in the long run
Task Group chairs

Louise Bezuidenhout

Raphael Cóbe

Marcela Alfaro Córdoba

Sara El Jadid

Bianca Peterson

Rob Quick

Hugh Shanahan

S. Venkataraman
Links

RDA Output for Curriculum DOI: [10.15497/rda00038](https://github.com/CODATA-RDA-DataScienceSchools/Materials)

https://github.com/CODATA-RDA-DataScienceSchools/Materials

https://github.com/CODATA-RDA-DataScienceSchools/TaskGroup
Learning to share: a road to national data publishing policy

RDA Adoption Week / June 19, 2020

Edit Görögh / University of Debrecen
András Holl / Hungarian Academy of Sciences
RDA Adoption Grant

**Grantees:** University of Debrecen and the Library and Information Centre of the Hungarian Academy of Sciences

**Time period:** Sept. 2019 – Sept. 2020

**Objective:** developing a data management package, which guides the users to RDA compliant data repositories and provides guidelines for uploading and archiving data.

**Primary tasks:**
- Uni Debrecen focuses on the data publishing workflows,
- Hungarian Academy of Sciences works on repository audit and certification processes.
Tasks and results
data publishing workflow

1. Translation of RDA/WDS Publishing Data Workflows WG Recommendations (doi: 10.15497/RDA00004)

2. Creating dissemination materials – planned to be distributed at trainings and data repository presentations.

3. Developing a data publishing workflow for Hungarian researchers and testing it through the data repository under development at Uni Debrecen.

4. Joining the national dialogue on data management.
Data repository

Test version of institutional data repository:
Dataverse based, customised platform, interoperability with international standards

User feedback:
easy uploading process, user friendly uploading page, some problems occurred with certain file formats
Data publishing guidelines

- Based on the RDA/WDS Publishing Data Workflows WG Recommendations, short guidelines are created on the opening page of the data repository.
- Translation of the recommendations is disseminated through open-access.hu website.
Joining the national dialogue

1. Joining the efforts of HRDA (Hungarian RDA Node)

2. A national survey on data management issues:
   - Based on the requirements and in cooperation with the National Research, Development and Innovation Office.
   - Primary objectives: to review
     ✓ the data management practices of Hungarian researchers,
     ✓ the willingness to open and share data.
   - 957 responders: 620 researchers, 148 PhD, 181 early career researchers, etc.
Sharing and publishing data

Willingness to share openly research data:
✓ 32% very much
✓ 6% not at all

Interest to use other researchers’ data:
✓ 32% high interest
✓ 7% no interest at all
✓ 30% use other data
Emphasis on repositories

The RDA/WDS Publishing Data Workflows Working Group advocates the implementation of existing good practices for repositories and all parts of the data publishing process and the development of new standards where necessary.
Hungarian Data Repositories

András Holl & Edit Görögh

HRDA (Hungarian National Node of RDA)
http://hrda.hu/

Library, Hung. Acad Sci. & Univ. of Debrecen
Grant – see Edit Görögh’s slides

Background

Hungarian (publication) repositories (minimal research data content), a couple of data repositories, one under construction

Repository Certification Committee

Research data culture

Financial backdrop
Adoption grant

- translation of recommendation
- discussion of recommendation
- pilot certification – two repositories of one data center
- status
- experience
- delays (pandemic, other difficulties)
- next steps (final self-assessment, formal certification, revision of the translation of the recommendation, possible feedback to RDA, certification of further repositories)
Welcome to KDK Repository

Welcome to the Research Documentation Centre (RDC) at the Centre for Social Sciences! You can browse through researches conducted at the four institutes (Institute for Legal Studies, Institute for Minority Studies, Institute for Political Science, Institute for Sociology) of the Centre for Social Sciences and download research collections. Every user can search among the topics, the researchers and other basic information about the research collections. Some of our collections are available to the public without registration.

Learn more about the RDC here: kdk.ttk.mta.hu/en.

Latest Additions

View most recently added items in the repository.

Search Repository

Search the repository using a full range of fields. Use the search field at the top of the page for a quick search.
On-boarding of existing Hungarian (publication) repositories

- general obstacles
- delays with the process

What do we expect from general (publication) repositories?

- Data supplements to deposited publications
- Quality control: the peer review process of the publication
- Simple data formats, small volumes (size and number of data files)
- Limited responsibilities - “as is”
- Goals: reproducibility rather than reuse