New business models for open research

Todd Vision

Jared Lyle

Mark Hahnel
Reliance on the perpetual generosity of a single funding source is risky

Plant genetics database at risk as funds run dry

National Science Foundation to cut support for Arabidopsis resource.

Alison Abbott

The world’s most valued plant database faces extinction because its funding is being phased out by the US National Science Foundation (NSF), and no alternative source is on the horizon.

"This is the wrong way to go," says genomics researcher Ernest Retzel of the National Center for Genome Resources in Santa Fe, New Mexico. "I believe it will set the field back."

The NSF says that it does not have a policy to support long-term, established research-infrastructure projects such as the Arabidopsis Information Resource (TAIR), which maintains a free, open-access database of genetic and molecular-biology data for Arabidopsis thaliana, or thale cress, the widely used model plant. "We
A diversity of beneficiaries allows for a diversity of revenue sources.
http://datadryad.org
Sources of revenue

- **Membership fees**
  - $500-$5000/yr
  - Stakeholder governance
  - Scale with community size

- **Data Publishing Charges** cover operating costs
  - Storage (10x current annual cost)
  - Curation (currently 1hr per data package)
  - System upkeep (software, servers, services, etc.)
  - Business administration

- **Grants for R&D**
  - Special projects
  - Technology innovation
Sources of revenue

- **Membership fees**
  - $500-$5000/yr
  - Stakeholder governance
  - Scale with community size

- **Data Publishing Charges** cover operating costs
  - Storage (10x current annual cost)
  - Curation (currently 1hr per data package)
  - System upkeep (software, servers, services, etc.)
  - Business administration

- **Grants for R&D**
  - Special projects
  - Technology innovation

*Scale linearly with deposits*
### Data Publishing Charges help sustain open data at Dryad

<table>
<thead>
<tr>
<th>Plan</th>
<th>Charge</th>
<th>Popularity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subscription</strong></td>
<td>Annual fee based on journal volume, $25-30 per research publication(^2)</td>
<td>44%</td>
</tr>
<tr>
<td><strong>Pay on submission</strong></td>
<td>$80 per data package, paid by the submitter</td>
<td>26%</td>
</tr>
<tr>
<td><strong>Deferred Payment</strong></td>
<td>$70-75 per data package(^2)</td>
<td>25%</td>
</tr>
<tr>
<td><strong>Pre-paid voucher</strong></td>
<td>$65-70 per data package(^2)</td>
<td>4%</td>
</tr>
<tr>
<td><strong>Waivers</strong></td>
<td>$0</td>
<td>1%</td>
</tr>
</tbody>
</table>

\(^1\) Additional fees for large data packages and non-integrated submissions  
\(^2\) Member discount applies
http://datadryad.org/pages/integratedJournals

### Integrated Journals and Costs for Submitters

All journals with either [integrated data submission](http://datadryad.org/pages/integratedJournals) or sponsored [Data Publishing Charges](http://datadryad.org/pages/integratedJournals) are listed below. Note that for large data packages, submitters will be asked to pay an additional $15 for the first GB beyond 10GB and $10 for each GB thereafter.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Submission</th>
<th>Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Integrated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submit: R: Prior to Review A: On Acceptance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data embargo allowed</td>
<td>Metadata hidden until publication</td>
</tr>
<tr>
<td>American Naturalist, The</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Biodiversity Data Journal</td>
<td>N</td>
<td>A</td>
</tr>
<tr>
<td>Biological Journal of the Linnean Society</td>
<td>Y</td>
<td>A</td>
</tr>
<tr>
<td>Biology Letters</td>
<td>Y</td>
<td>A</td>
</tr>
<tr>
<td>BioRisk</td>
<td>Y</td>
<td>R</td>
</tr>
<tr>
<td>BMJ Open</td>
<td>Y</td>
<td>R</td>
</tr>
<tr>
<td>Comparative Cytogenetics</td>
<td>Y</td>
<td>R</td>
</tr>
<tr>
<td>Ecological Monographs</td>
<td>Y</td>
<td>A</td>
</tr>
</tbody>
</table>

If your journal is not listed below...

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>A</th>
<th>Y</th>
<th>Y</th>
<th>$90</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Society of Naturalists</td>
<td>American Society of Naturalists</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pensoft</td>
<td>Pensoft</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$80</td>
<td>$80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Learn more about:
- [Membership](http://datadryad.org/pages/integratedJournals)
- [Submission Integration](http://datadryad.org/pages/integratedJournals)
- [Pricing plans](http://datadryad.org/pages/integratedJournals)
Some of the many ways organizations can contribute through this model

- **Any organization can**
  - become a member
  - sponsor DPCs through one of several payment plans

- **Publishers can**
  - integrate data and manuscript submission

- **Universities, publishers, & funders can**
  - sponsor DPCs. Failing that, they can facilitate pay on submission.
  - adopt open data policy policies
  - guide researchers to reputable repositories
  - fund or collaborate on R&D projects
New Business Models for Open Research: openICPSR

Jared Lyle
lyle@umich.edu

Open Repositories 2014
Helsinki, Finland
12 June 2014
What is ICPSR?

- Repository of social science data established in 1962 for data sharing and preservation
- Membership-based organization -- over 750 institutional members (colleges and universities) from around the world
- Source for training in statistics and data curation through the Summer Program
The 1915 Iowa State Census Project
(ICPSR 28501)

Principal Investigator(s): Goldin, Claudia, Harvard University, and National Bureau of Economic Research; Katz, Lawrence, Harvard University, and National Bureau of Economic Research

Summary:
The 1915 Iowa State Census is a unique document. It was the first census in the United States to include information on education and income prior to the United States Federal Census of 1940. It contains considerable detail on other aspects of individuals and households, e.g., religion, wealth and years in the United States and Iowa. The Iowa State Census of 1915 was a complete sample of the residents of the state and the returns were written by census takers (assessors) on index cards. These can... (more Info)

Access Notes

- These data are freely available.

Dataset(s)

Dataset - Download All Files (90,107 KB)

Documentation: Codebook.pdf  User guide.pdf

Data: SAS  SPSS  Stata  ASCII  Delimited
      ASCII +  SAS Setup  SPSS Setup  Stata Setup

http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/28501
Existing Funding Models

- Fee for access model (membership)
- Agency model (agency funds public access)
- Fee for deposit model (depositor funds public access)

See also:
ICPSR Funding Sources (FY14)

- **Sponsored Projects**: 48%
- **Summer Program**: 8%
- **Membership**: 20%
- **Other**: 24%

http://www.icpsr.umich.edu/files/membership/or/annualreport/AR2012-2013.pdf
Sponsored Projects
openICPSR (beta) Browse data

Three easy steps:
1. Name your project
2. Upload and describe files
3. Purchase a curation option

Watch our videos>

Share your social and behavioral science research data

Get started now »

Maximize Access
Be recognized and cited

Store Safely
Store your data with confidence

Protect Confidentiality
Ensure confidentiality and privacy

© Inter-university Consortium for Political and Social Research, 2014

About | Privacy | FAQs | Pricing | Contact us

http://openicpsr.org
Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica

Principal Investigator(s): Gertler, Paul; Heckman, James; Pinto, Rodrigo; Zanolini, Arianna; Vermeerch, Christel; Walker, Susan; Chang-Lopez, Susan; Grantham-McGregor, Sally

Title: Jamaica Earnings Analysis


Persistent URL: http://doi.org/10.3886/E2402V3

Project Description

Summary

In 1986-1987, the Jamaican Study enrolled 129 growth-stunted and disadvantaged children aged 9-24 months that lived in Kingston, Jamaica. Half of the children were randomized into a cognitive stimulation intervention which lasted 18 months (S. P. Walker, S. M. Chang, C. A. Powell, S. M. Grantham-McGregor, The Lancet 366, 1804 (2005)). A matched non-stunted group was also followed for comparison purposes. We resurveyed both the stunted and non-stunted samples in 2007-08, some 20 years after the original intervention, when the participants were approximately 22 years old. We found and interviewed 105 out of the original 129 stunted study participants. We estimated the impact of the stimulation intervention on earnings by comparing the earnings of the stunted-treatment group to those of the stunted-comparison group.
Why openICPSR?

- Distribution in ICPSR’s catalog and network
- Trusted, sustainable organization (50+ years)
- Discipline-specific metadata
- Interface
- Disseminate sensitive/restricted-use data
openICPSR Funding Model

- Fee for deposit model (depositor funds public access)
- Curation package (depositor funds curation)
- User fees (for costly services)
- Infrastructure supported by ICPSR
Other Possible Services (?)

- Bulk self-deposit
- Institutional/departmental repository
- Journal replicated data repository
Acknowledgements

• Linda Detterman
• Nathan Adams

See also:
http://www.slideshare.net/ICPSR/orientation-to-openicpsr
Thank you!

lyle@umich.edu
Extra
New Funding Models

- Commercial services
- User fees (for costly services)
- Overhead (% grants)
- Infrastructure (federal agencies support)

See also:
• What is the long-term funding stream?
• What happens if service fails?
• How to maximize discoverability?
• Reach a network of researchers?
• Abide by archival standards?
• Handle sensitive data?
Inter-university Consortium for Political and Social Research

citable | shareable | discoverable
Policy on data management and sharing

The Wellcome Trust is committed to ensuring that the outputs of the research it funds, including research data, are managed and used in ways that maximise public benefit. Making research data widely available to the research community in a timely and responsible manner ensures that these data can be verified, built upon and used to advance knowledge and its application to generate improvements in health.

We believe that success in maximising the value of research data depends crucially on fostering a culture in which both data generators and data users adopt good research practice, and act with integrity and transparency in managing, using and sharing research data. Researchers, research institutions, resource providers, funders and publishers all have important roles to play in helping to develop this enabling environment and in developing the resources and systems required.

There is a growing international consensus on the need to preserve and share research datasets in a manner that maximises their long-term value. This has been articulated in key documents such as the OECD Principles and Guidelines for Access to Research Data from Public Funding (2007); the Fort Lauderdale Principles (2003) and the Toronto Statement (2009).

Our policy on data management and sharing reflects the principles set out in these statements. It is consistent with the Trust’s position on open and unrestricted access to published research, our guidelines on good research practice and our policy on intellectual property and patenting.

The policy should be read in conjunction with the associated guidance. We will keep both our policy and guidance under review.
ABOUT US
DIGITAL SCIENCE IN A NUTSHELL

At Digital Science, we are actively working with entrepreneurs and developing software internally to make scientific research more efficient.

We are a technology company serving the needs of scientific research...

We provide software and technology to serve a community we understand – scientific researchers. Many in our team started their careers as scientists so we know the daily challenges they face. Whether at the bench or in a research setting, our products help to simplify workflows and change the way science is done.

We incubate and invest in interesting projects.

We invest in promising start-ups and small businesses that serve the needs of scientists, many with a start in academic research settings.
<table>
<thead>
<tr>
<th>Logos</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOCKSS</td>
<td>Backed up in multiple institutions around the world</td>
</tr>
<tr>
<td>DataCite</td>
<td>DOIs provided by DataCite at the California Digital Library</td>
</tr>
<tr>
<td>Committee on Publication Ethics</td>
<td>Adhere to ethics of academic publishing, as per guidelines</td>
</tr>
<tr>
<td>ORCID</td>
<td>ORCID launch partner, files to be pushed to author profiles</td>
</tr>
<tr>
<td>Amazon Web Services</td>
<td>All content hosted on AWS with triple file storage, fast load times and unbeatable uptime</td>
</tr>
</tbody>
</table>
**figshare for publishers**

Make your articles come to life

figshare enables you to host large amounts of data in your online articles with no impact on your infrastructure. We ♥ all file types and host them for you in the cloud.
**figshare** for institutions

We securely host research data privately or publicly in the cloud.

**figshare** for institutions is a simple, user-friendly and cost-effective solution for you to manage all of the research outputs of your institution. It caters to the needs of anyone who deals with research data. We ♥ all file types and safely host them for you in the cloud.
figshare technology platform

**CORE**
- Globally redundant figshare server infrastructure ensuring 99.99% availability
- 24/7 Support
- regular upgrades

**OPTIONS**
- fully client-branded experience including client-specific DOIs
- placing your institutions output among more than 1.5M datasets in figshare.com and ensuring maximal discoverability of your datasets

**PAY AS YOU GO**
prepaid data packages

**STORE ON YOUR INFRASTRUCTURE**
Take home message from each of us

• Sustainability should be the only goal in this space.

• Diversified funding is key.

• Don’t be afraid to reach out to *all* the beneficiaries of your resource.