

Open science and the evolving concept of scientific publication

Matthew B. Jones

National Center for Ecological Analysis and Synthesis (NCEAS)
University of California Santa Barbara



Open Science for Synthesis 2014
August 7, 2014

Quest for reproducible research

Scientific Publication and Communication

Vision for Reproducible Open Science

PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMP'T
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

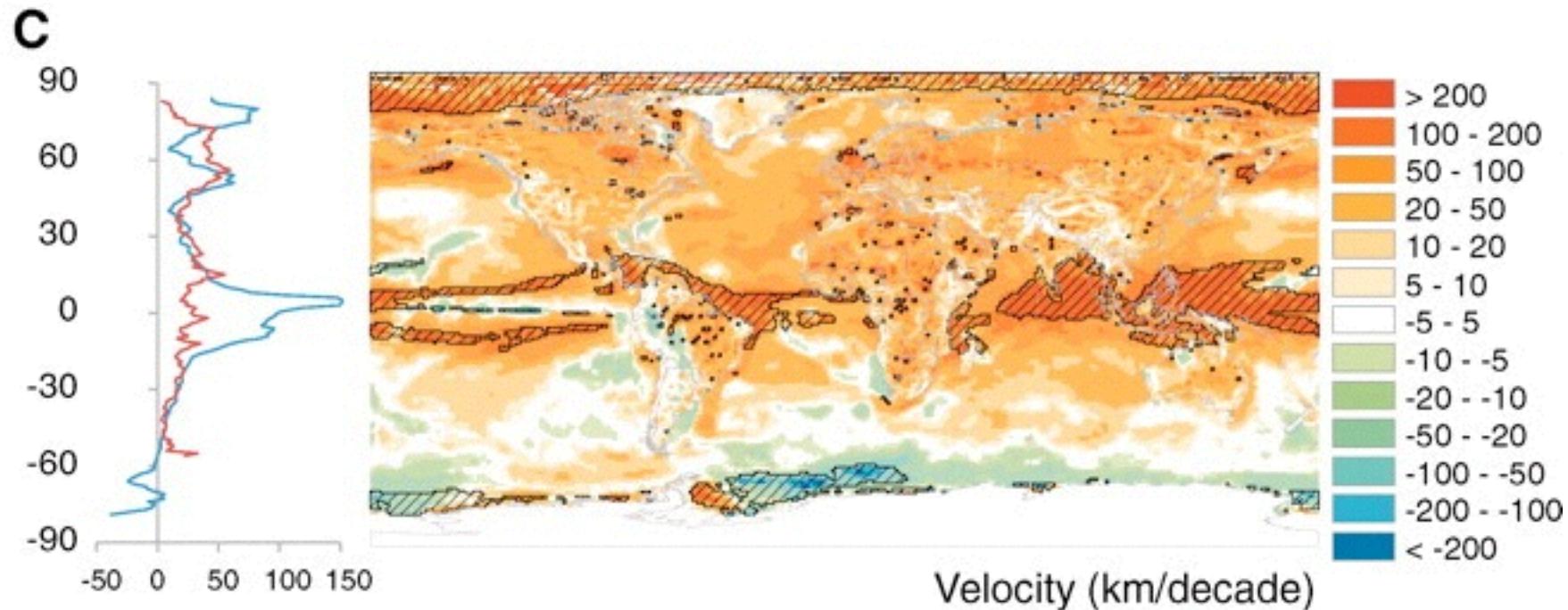
In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.

- The journal article, largely unchanged
- Philosophical Transactions of the Royal Society
 - ~350 years (1665-2014)

Pace of Shifting Climate in Marine and Terrestrial Ecosystems

Burrows et al. 2011. Science 334:652-655

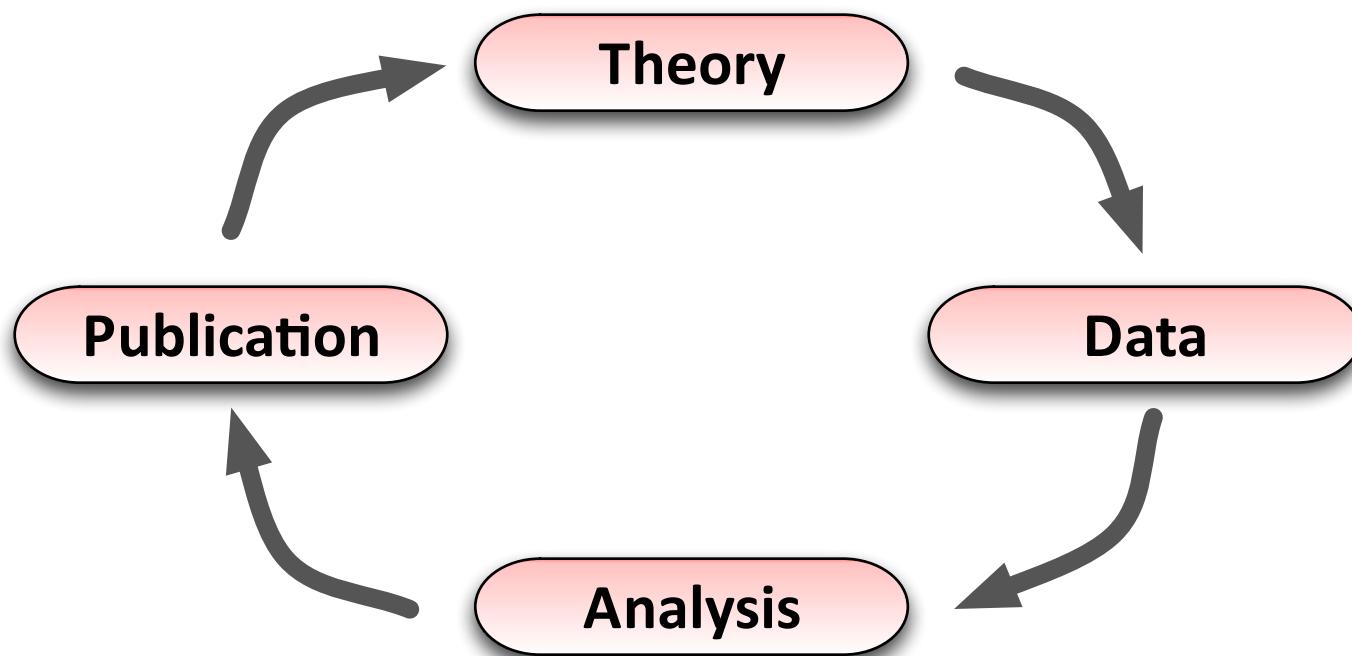
“The velocity of climate change is **two to seven times faster in the ocean** than on land in the sub-Arctic and within 15° of the equator...”



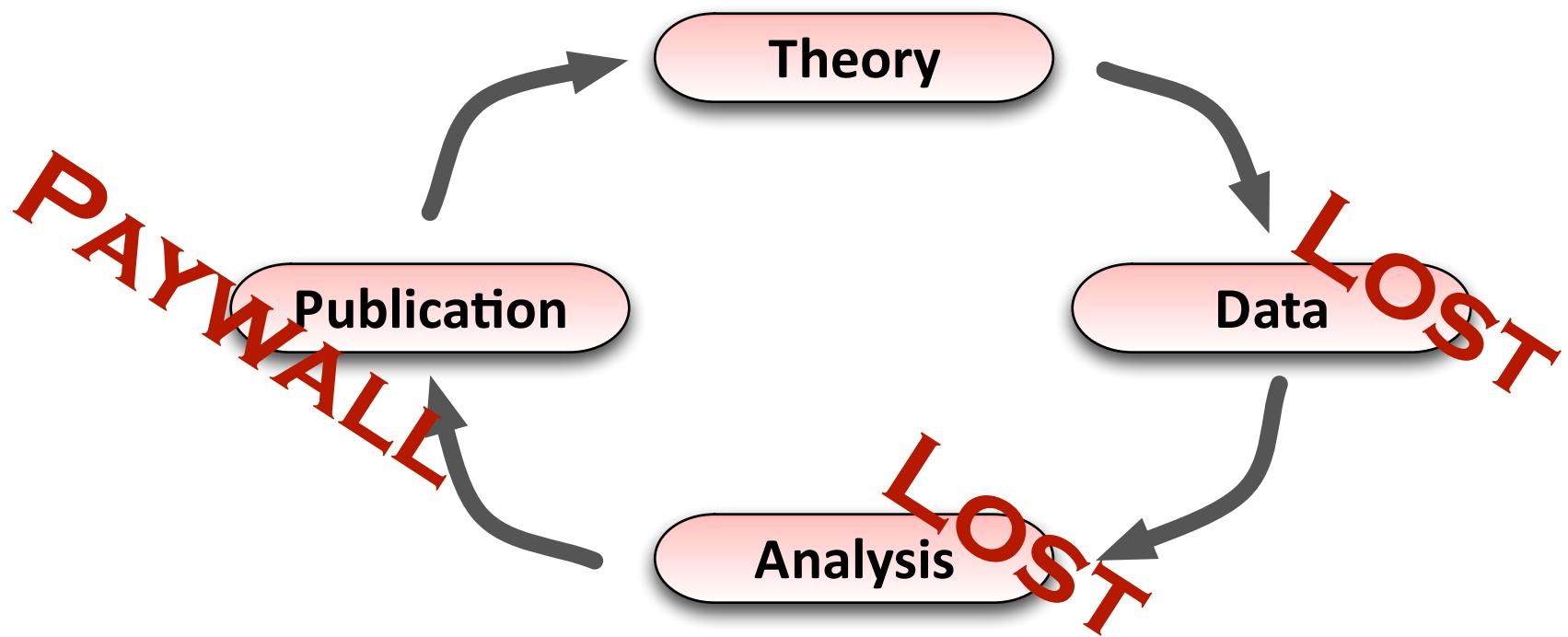
The Science Contract

- Science is a contract to share both results and methods
 - Reproducibility is key
 - Without reproducibility, it isn't really scientific
- Should be as fast as possible
- Should be as reliable as possible

State of reproducible science



State of reproducible science



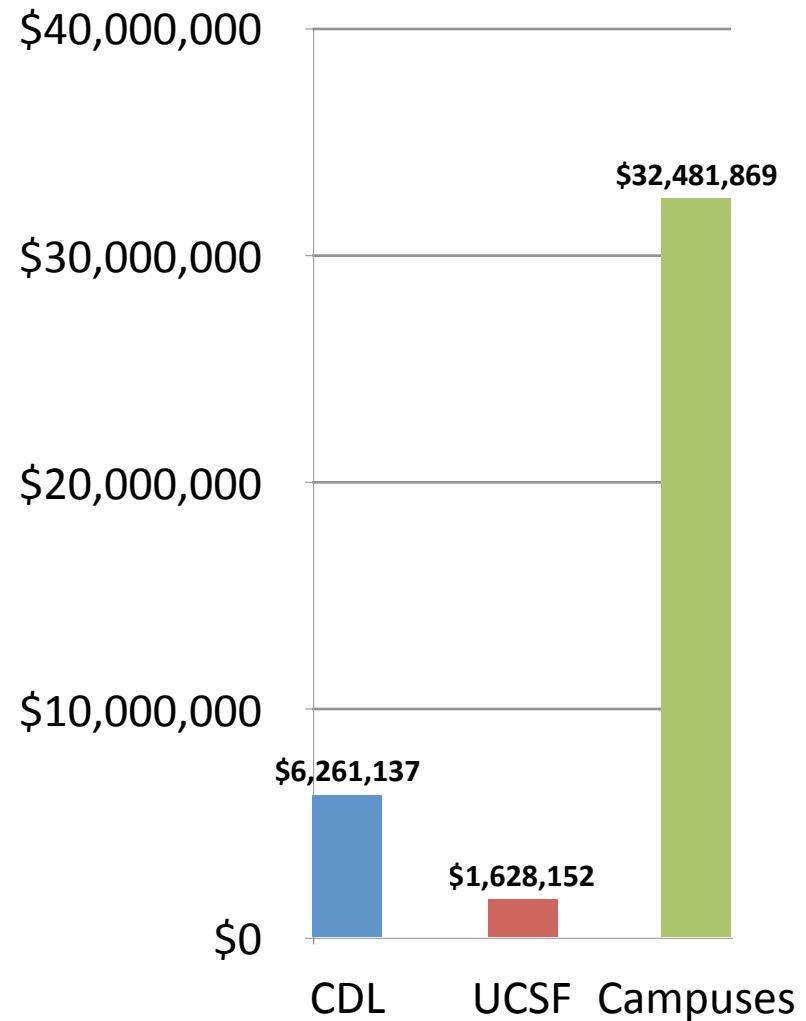
Quest for reproducible research

Scientific Publication and Communication

Vision for Reproducible Open Science

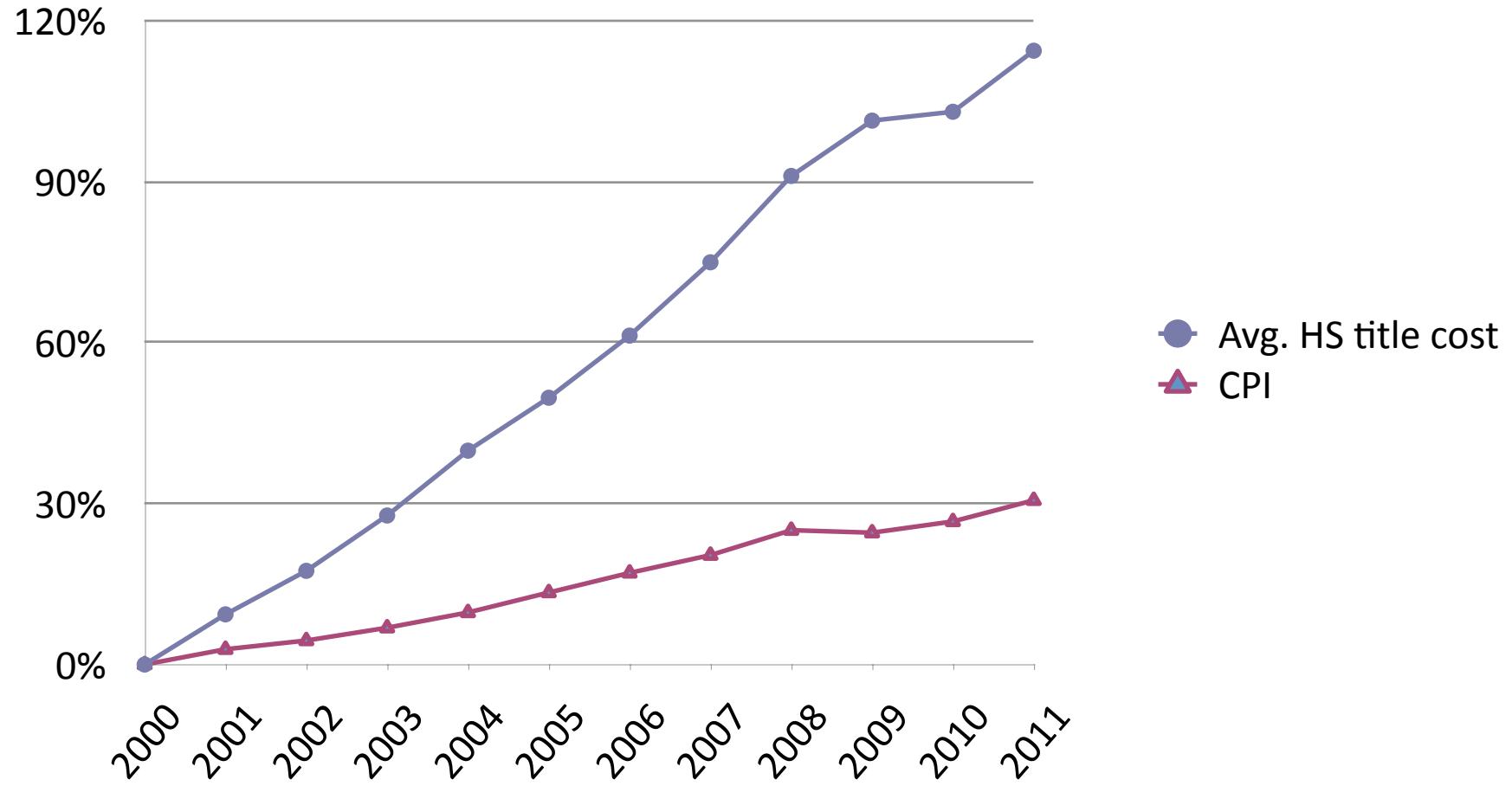
2011 UC Systemwide Subscriptions

- Total
 - \$38,743,006
- CDL
 - \$6,261,137
 - 16%
- 10 Campuses
 - \$32,481,869
 - 84%
- UCSF
 - \$1,628,152
 - 4% of total



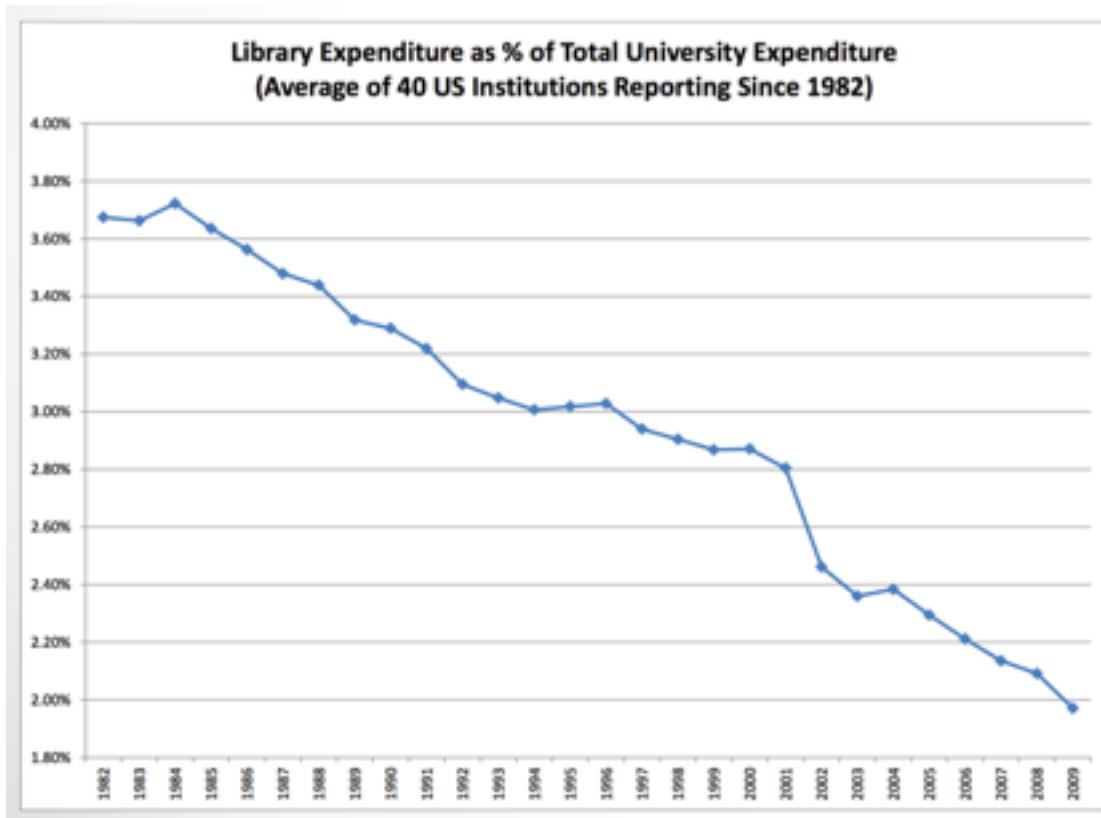
Source: Rich Schneider, UCSF Open Access Report

Percent Increase in Cost for the Average Health Sciences Journal versus the CPI



Source: Rich Schneider, UCSF Open Access Report

Libraries are losing access



Cancellations
9 database contracts cancelled since 2008.

600 journals (7.5%) cancelled in 2010-2011, including one entire contract.

More journal cancellations in 2013.

2010/2011 Profits for Commercial Publishers

	<u>Profits</u>	<u>Revenues</u>	<u>Profit Margin</u>
• Elsevier	\$1.2B	\$2B	36%
• Wiley	\$106M	\$253M	42%
• Springer	\$467M	\$1.4B	34%
• Informa	\$74M	\$230M	32%
• Apple			24%
• Google			27%

Source: Rich Schneider, UCSF Open Access Report

What's wrong with this model?

- Authors provide **ALL** of the **content** and **MOST** of the **labor** in science publishing
 - Theory, experimental design, data collection, analysis, deduction, inference
 - Authorship, formatting, graphics design
 - Editing
 - Peer Reviewing
- Scholarship is **expensive to produce**
 - Cost to scholars, universities, and funders
- Far **cheaper to distribute**
 - Cost to publishers
 - Differential is publisher profit

The Internet disrupts,
just ask the music
industry

Author, Editor, and Reviewer Value

- Reviewers: Typically 5 hours per review
 - *The STM Report, 2009*
- UC Peer review, all publishers: \$21M
 - *Rich Schneider, UCSF Open Access Report*

UC authorship contributions to:

Elsevier **articles**: 2.2% of all articles

Elsevier **revenue**: \$31M

Elsevier **profit**: \$9.8M

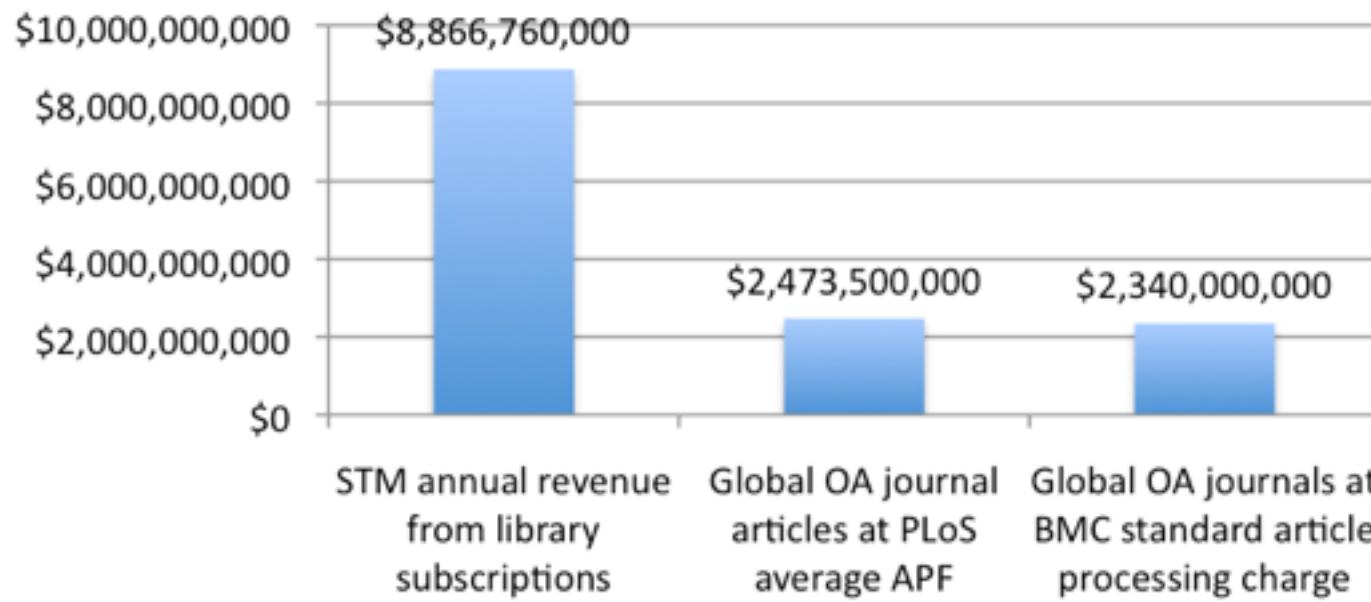
Nature **articles**: 12% of all published articles

Nature **revenue**: \$5M

Nature **profit**: \$700K

Where from here? Open Access Publication

Library savings with full flip to open access



"Using the PLoS average article processing fee of \$1,649 U.S. per article, or BMC average article processing charge of \$1,560 U.S., libraries worldwide could fund full open access to the world's estimated 1.5 million scholarly peer-reviewed journal articles produced every year at **less than 30% of current annual global academic library journal expenditures.**"

UC Open Access Mandate

- UC Authors have copyright
- Change the **default policy** to provide UC with a **non-exclusive right to distribute works** before negotiating with the publisher
- Applies to all UC scholarly works
- **Opt-out clause** for academic freedom
- Author **obligation to deposit works** with UC eScholarship

Source: [http://osc.universityofcalifornia.edu/
openaccesspolicy/](http://osc.universityofcalifornia.edu/openaccesspolicy/)

Library Support for Self-archiving



About eScholarship Publish/Deposit Browse Content Help My Account My Items (0)

Search publications...

[Advanced Search](#)
[Return to search results](#)

Narrow results:

Peer-reviewed only

▼ Discipline

- Arts and Humanities
- Life Sciences
- Law
- Social and Behavioral Sciences
- Architecture
- Engineering
- Medicine and Health Sciences
- Education
- Business
- Physical Sciences and Mathematics

▼ UC System

- UC Berkeley
- UC Davis
- UC Irvine
- UC Los Angeles
- UC Merced
- UC Riverside

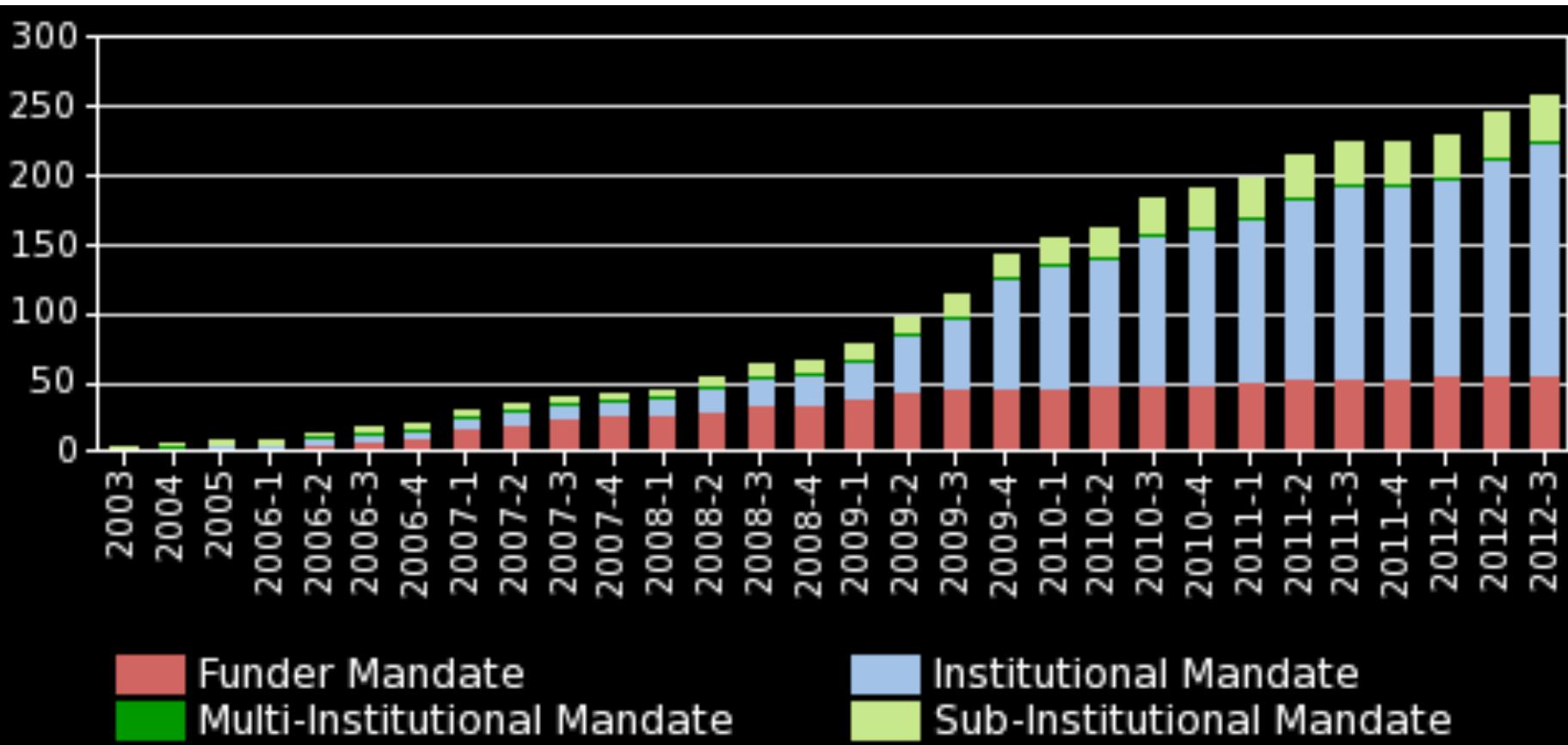
RSS for this search Results 1 - 1 of 1

Sort by: date descending Show: 10 per page

	Title:	Author:	Published:	
1	Ecoinformatics: supporting ecology as a data-intensive science	Michener, William H.; Jones, Matthew B.	Cell Press, 2012	Open Item Show Abstract Show Similar Items Save Item

Growth in Institutional Mandates

- 163 Institutional mandates Dec 5, 2012



Source: <http://roarmap.eprints.org/>

C. Kelty, June 2012

- “The overall goal of a sustainable scholarly publishing model is to move more towards **paying for services we value**, rather than paying for access to content”
 - “Preservation, findability, promotion, design, and other services that improve quality and accessibility are well worth paying for”
 - “Skyrocketing subscription costs that limit access to only the richest institutions are not”

Quest for reproducible research

Scientific Publication and Communication

Vision for Reproducible Open Science

\$ Affordable Author Plans - For Life

You're probably used to paying per paper when you publish in Open Access. Or paying for page and image charges in subscription journals. PeerJ is doing away with all of that. Pay once, publish for life.

[Individual plans](#)[Multi-author or Group plans](#)

\$0 **Free** - join the community and help us rethink scholarly communications.

You'll also get one public preprint per year in *PeerJ PrePrints*, but no peer-reviewed publications in PeerJ.

[Create my free account](#)

Basic Plan

\$99**LIFETIME**

\$99 when/before submitting
or \$129 when accepted
[See FAQ below](#)

Within PeerJ**One** publication/year

Enhanced Plan

\$199**LIFETIME**

\$199 when/before submitting
or \$239 when accepted
[See FAQ below](#)

Within PeerJ**Two** publications/year

Investigator Plan

\$299**LIFETIME**

\$299 when/before submitting
or \$349 when accepted
[See FAQ below](#)

Within PeerJ**Unlimited** publications/year

Reputation derives from review and editors, not publishers



Follow eScholarship on [f](#) [t](#)

[About eScholarship](#) [Publish/Deposit](#) [Browse Content](#) [Help](#)

[My Account](#)

[My Items \(0\)](#)

Search publications...

go

[Advanced Search](#)

frontiers of biogeography
the scientific magazine of the International Biogeography Society



RSS

[Share](#)

Select Issue

[Print Page](#)

[Search journal...](#)

go

[Journal Info](#)

Frontiers of Biogeography

Unless stated otherwise (for particular sections in the instructions for authors), enquiries should be made by e-mail to frontiersofbiogeography@gmail.com.

ISSN: 1948-6596

[About Us](#)

[Editorial Board](#)

[Contact Us](#)

[Policies](#)

[Aim and Scope](#)

[Submission Guidelines](#)

[Submit Article](#)

[Administrator Login](#)



cover: Schooling bannerfish.
Picture by Jon Hanson, creative commons license; modified from Wikipedia Commons.

Current Issue, Volume 4, Issue 3, 2012

[Cover](#)

[Schooling bannerfish](#)

[FB Information](#)

[Aims, open access, editorial board, and officers](#)

[Editorial](#)

[Research letters at the frontiers of biogeography
Field, Richard; Hortal, Joaquin; Dawson, Michael N](#)

[News & Update](#)

[update: Choosing the right path for species distribution modeling
Tessarolo, Geiziane](#)

[update: Stemming "ignorance creen" in paleoecology and](#)

Publishing and reviewing blogs

- doi:10.6084/m9.figshare.97901

Living in an Ivory Basement

Stochastic thoughts on science, testing, and programming.

misc

personal

python

science

teaching

testing

Posting blog entries to figshare

Just over a week ago, I [posted a list of wanted tech](#) that I thought would help further open science. One item that struck a chord with a number of people on Twitter and in the comments was the idea of giving blog entries a DOI:

4. An easy way to copy blog posts to figshare and give them DOIs.

Thu 22 November 2012

By [C. Titus Brown](#)

In [science](#).

tags: [blog](#)[webmaking](#)[figshare](#)

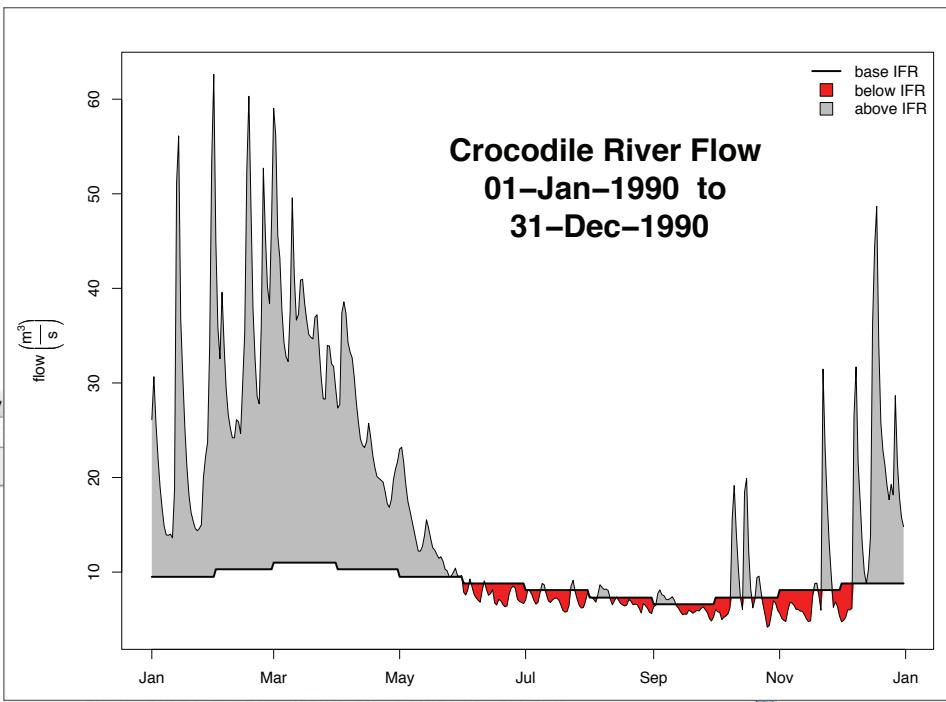
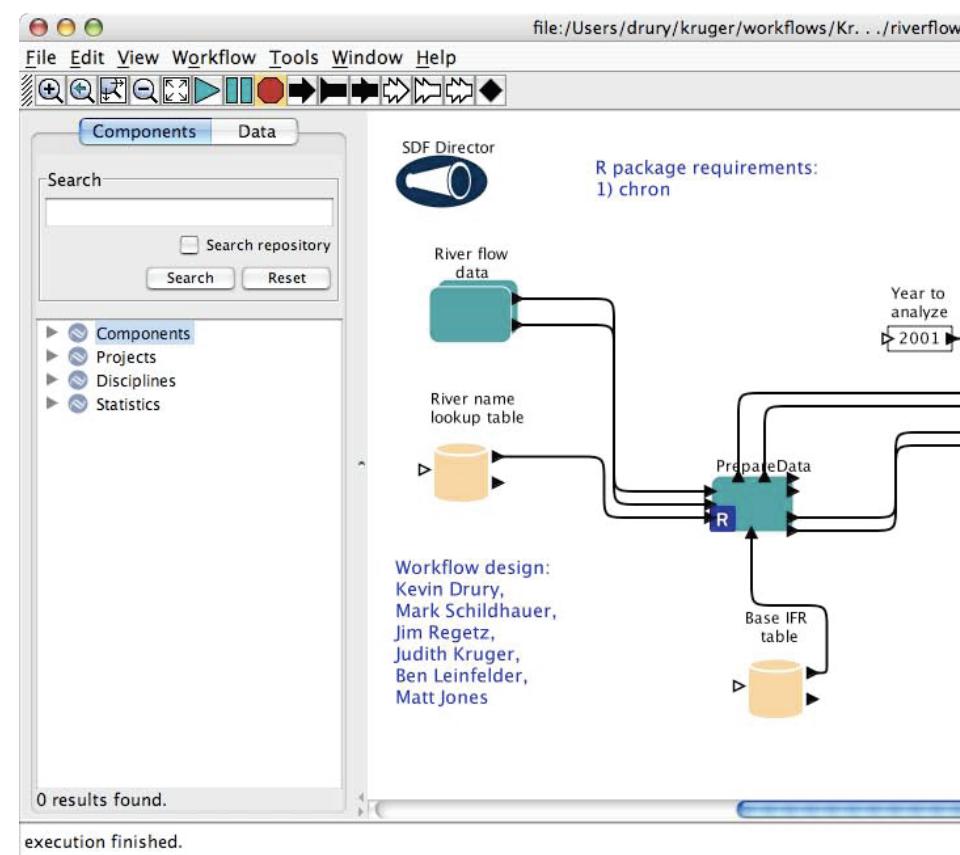
Seriously, why hasn't this already been done? It should be a short hack to make use of the figshare API to suck down a blog entry to figshare, "freeze" it, and post the DOI to the original blog entry.

This would enable a community of opinionated microcontributions, each of which can be tracked and versioned).

And once you can blog executable notebooks, well, there you go!



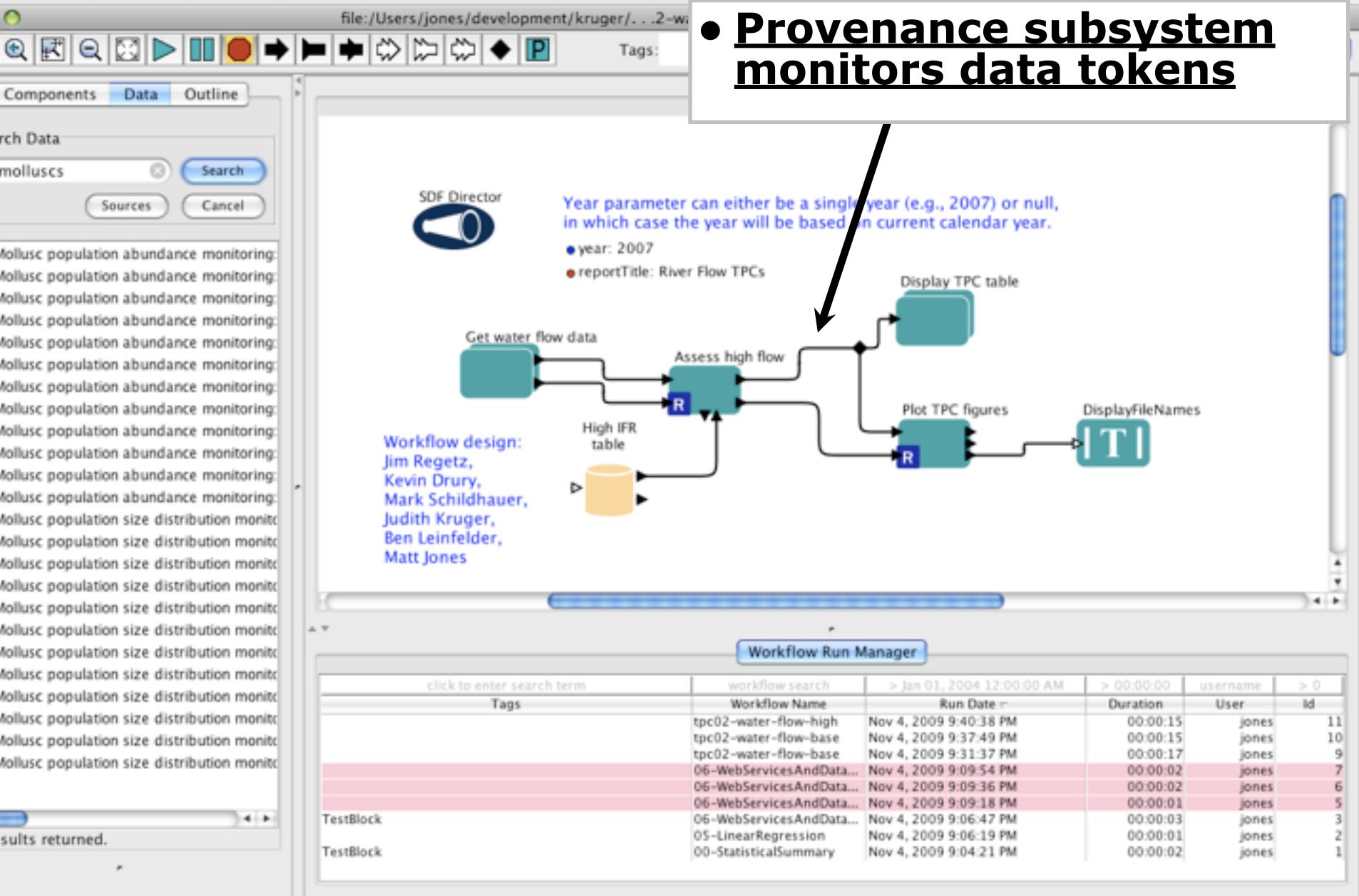
Archiving isn't just for pubs and data...





Run Management & Sharing

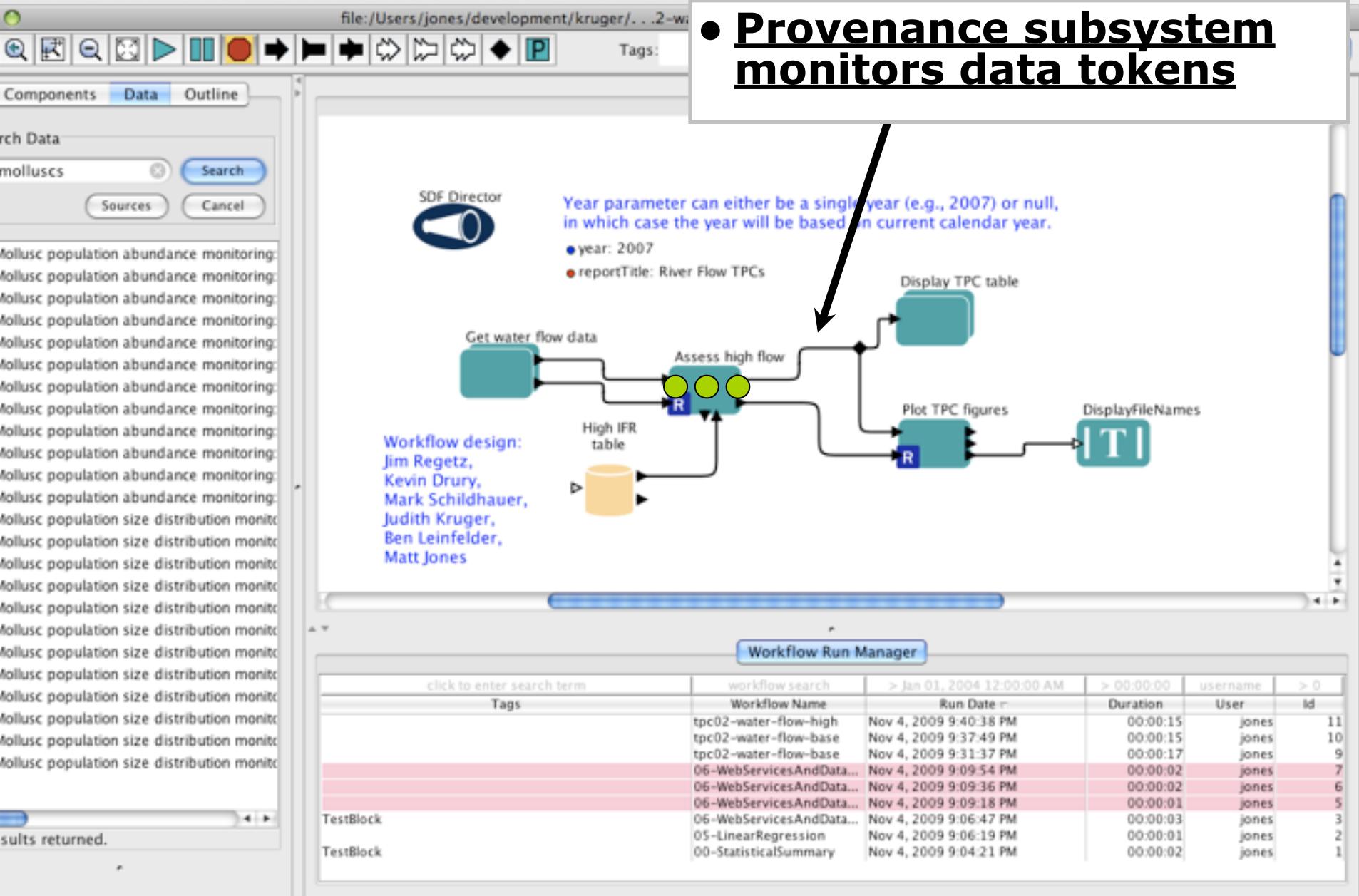
- **Provenance subsystem monitors data tokens**



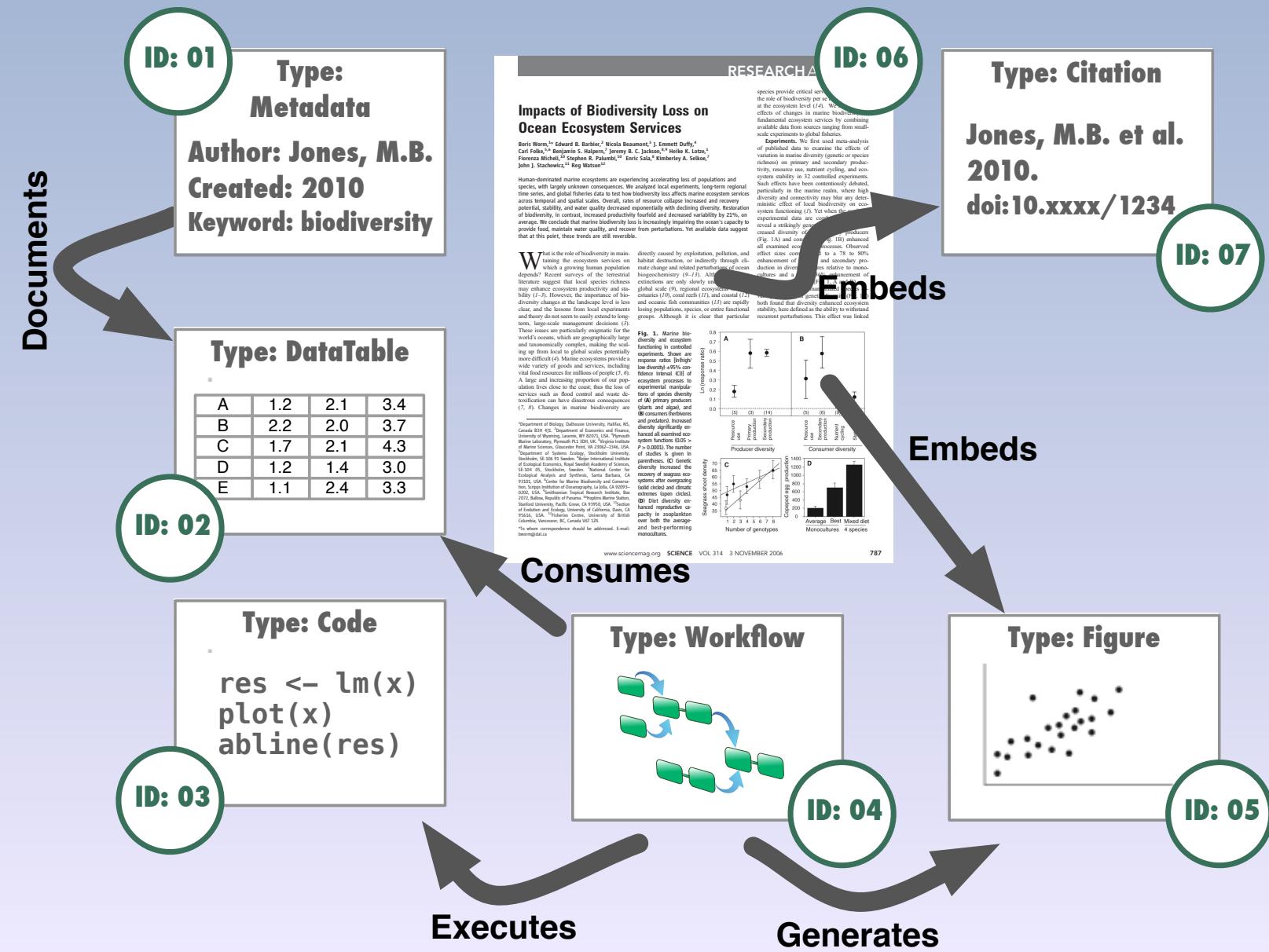


Run Management & Sharing

- **Provenance subsystem monitors data tokens**



Towards a “Living Paper”



Emulating open source culture

- Open source has a great culture of re-use
 - Re-writing code that already exists is bad practice
 - Elaborate code sharing and versioning systems for code re-use
- ‘Forking’ to minimize effort
 - Take someone else’s work
 - Modify it to suit your needs
 - Publish modifications

Emulating open source culture

- Open source has a great culture of re-use

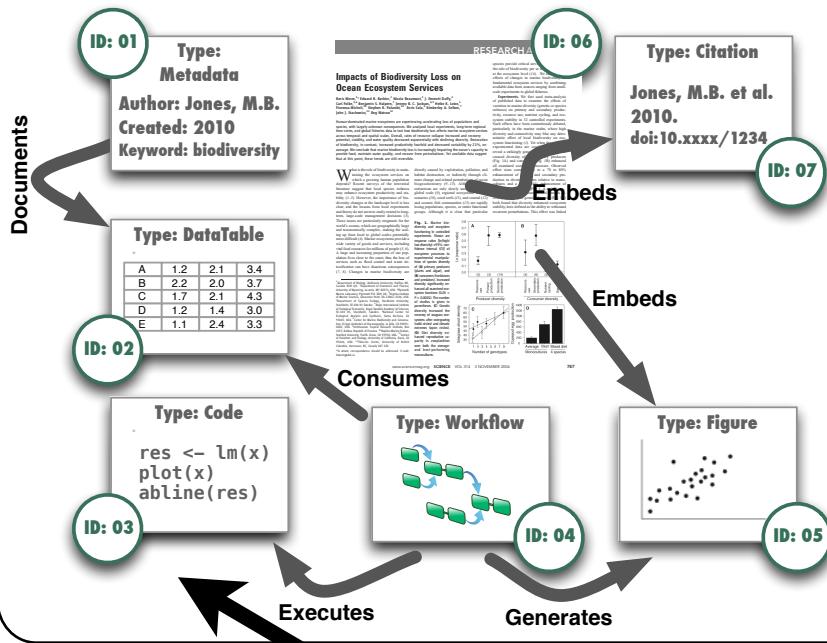
The screenshot shows a GitHub repository page for 'mbjones / ezid'. The top navigation bar includes links for Search or type a command, Explore, Gist, Blog, Help, and user mbjones. Below the header, the repository name 'mbjones / ezid' is displayed, along with options to Pull Request, Unwatch, Star, Fork, and Graphs. The main content area shows the repository's purpose: 'Implementation of the EZID identifier services in a Java library' with a 'Read more' link. It features tabs for Code, Network, Pull Requests (0), Issues (0), Wiki, Graphs, and Admin. Below this, there are buttons for Clone in Mac, ZIP, HTTP, SSH, and Git Read-Only, along with the URL https://github.com/mbjones/ezid.git and a Read+Write access icon. The 'Code' tab is selected. At the bottom, there are dropdowns for branch (set to master) and files (set to Files), along with links for Commits, Branches (1), Tags, and Downloads. The commit history for 'ezid /' shows 5 commits by mbjones, with the latest commit being '78ca84772e'. The commits listed are:

- Added multithreaded Executor service embedded in a client that can qu... [mbjones]
- src 4 months ago Added multithreaded Executor service embedded in a client that can qu... [mbjones]
- LICENSE.txt a year ago Initial design and development of an EZID service wrapper for Java. T... [mbjones]
- README.txt a year ago Initial design and development of an EZID service wrapper for Java. T... [mbjones]
- pom.xml a year ago Initial design and development of an EZID service wrapper for Java. T... [mbjones]

Forking a science publication

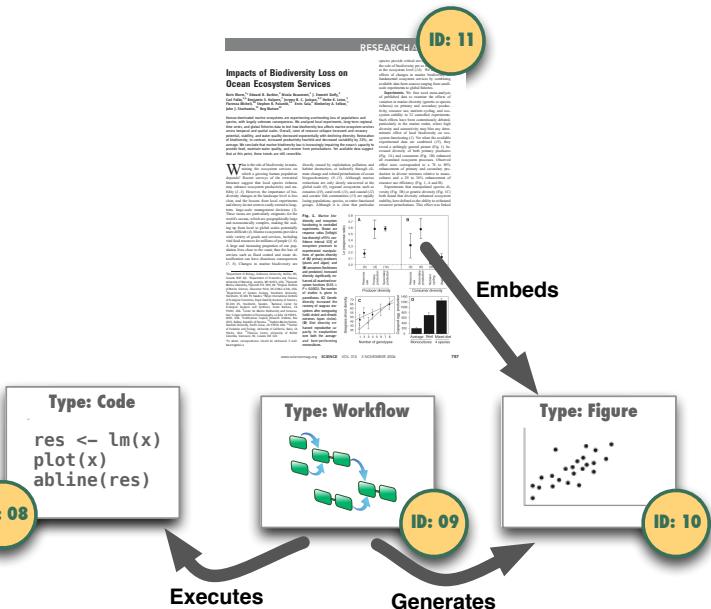
- All the results are archived and versioned
 - Open and re-run code to verify results
 - Modify data, analysis, etc as needed
 - Re-run, and archive new versions of results
 - Provenance traces show relationship to original work

Publication: version 1

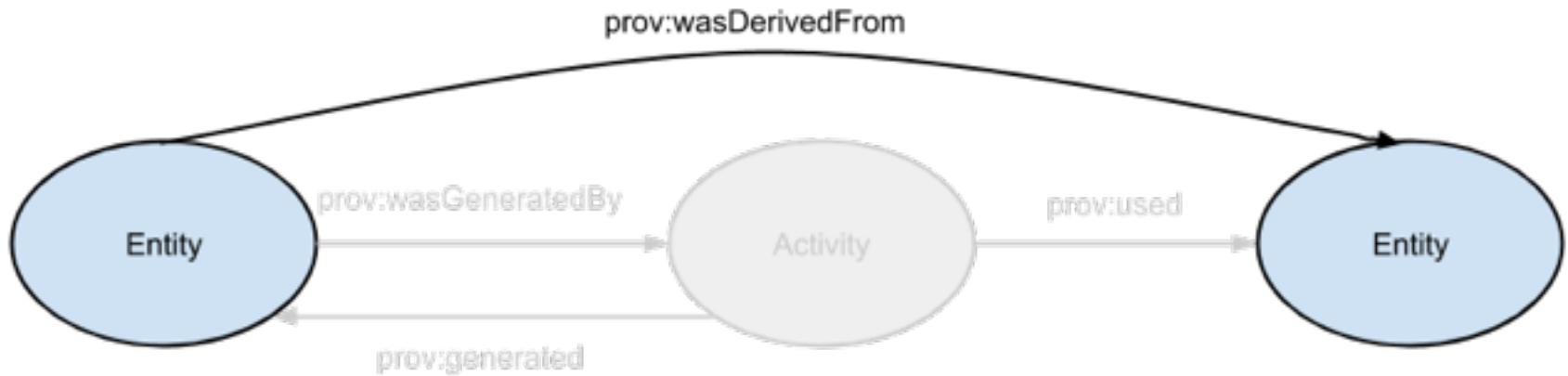


Replaces

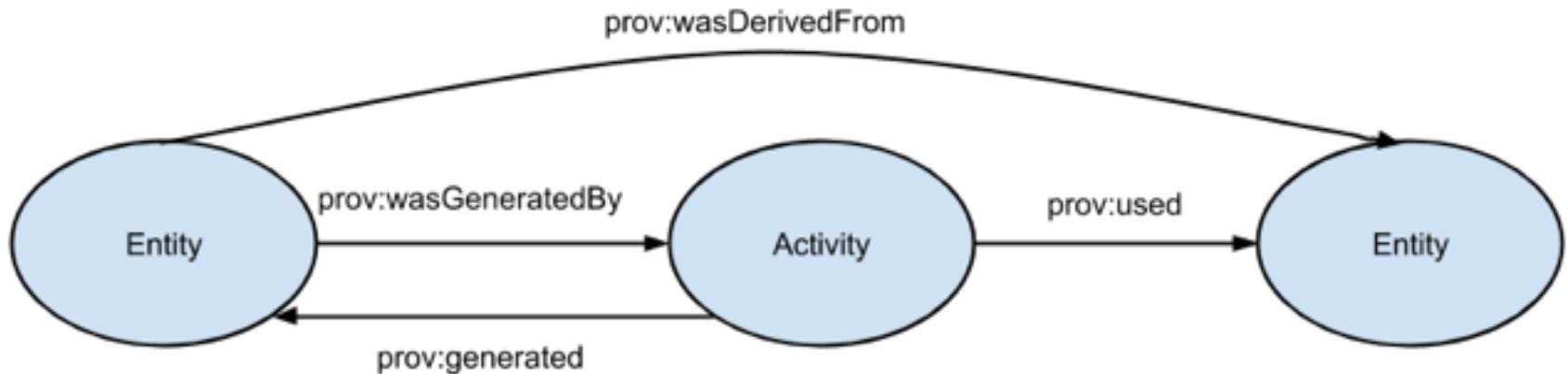
Publication: version 2



A simple PROV model



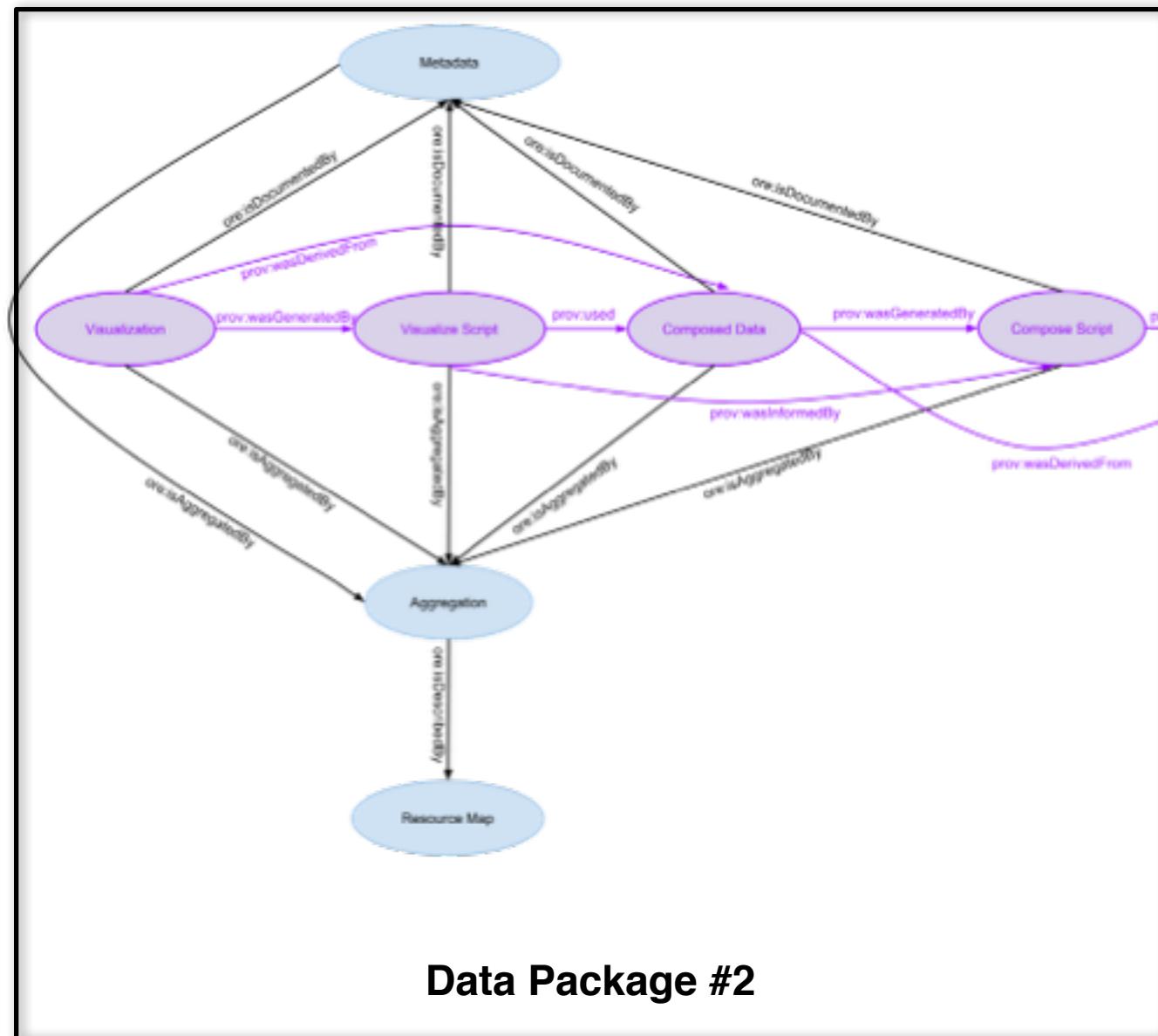
PROV with Activity



prov:wasDerivedFrom



Provenance of scientific outputs



Data Table, Image, and Other Data Details

Technical Metadata

Ecological Metadata Language (EML) File

knb

Other Entity

seabirdSurveyDiversity.png

[Download](#)



Description

Object Name [seabirdSurveyDiversity.png](#)

Size 189493

Authentication 514259fe89f514259fe89fng5142vrl...

Format Name image/png

Online Distribution Info [seabirds.5.1](#)

This image was derived from data [knb.482.1](#) using the activity
[seabirds.4.1](#)

Other Entity

Species_List

[Download](#)



Description

Object Name [Species_List](#)

Size 189493

Authentication 514259fe89f514259fe89fng5142vrl...

Format Name csv

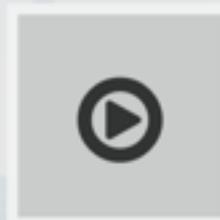
Online Distribution Info [seabirds.3.1](#)

This table was derived from [knb.485.1](#), [jstocking.4.9](#), and
[jstocking.3.4](#) using activity [seabirds.6.1](#)

Other Entity

seabirdSurveyScript

[Download](#)



Description

Object Name [seabirdSurveyScript](#)

Size 189493

Authentication 514259fe89f514259fe89fng5142vrl...

Format Name application/octet-stream

Online Distribution Info [seabirds.4.1](#)

This activity generated [seabirds.5.1](#) using data [seabirds.3.1](#)

Data Derivation Provenance

Related Datasets

This dataset is derived from:

Sir Allister Hardy Foundation, and Sonia Batten. 2014.

Marine Bird and Mammal Observations in the North Pacific Ocean.

knb.485.1

   22 views

This dataset is derived from:

Mary Anne Bishop. 2013.

Long-term monitoring of Seabird Abundance and habitat associations during late fall and winter in Prince William Sound

jstocking.4.9

   17 views

This dataset has derivation:

Mary Anne Bishop and Kathy Kuletz. 2013.

Marine Bird Observations in the Prince William Sound, AK, Nov2009 - Mar2012.

jstocking.3.4

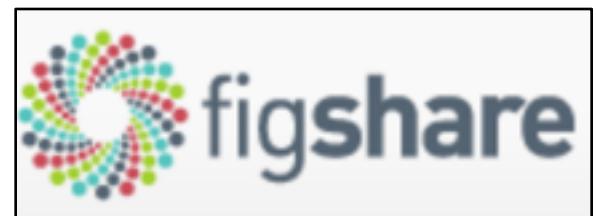
   9 views

Features of a Living Paper

- Archive all products of science, not just papers
 - Data
 - Analytical code
 - Workflows
 - Figures
- Communicate immediately and broadly
- Post-submission peer review/ranking
- **Build incrementally on past work**

Realizing the Living Paper

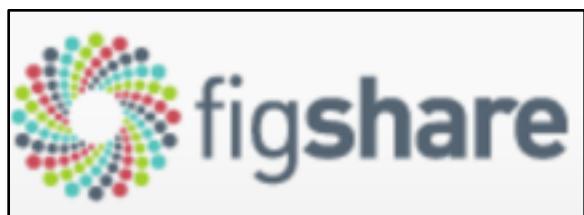
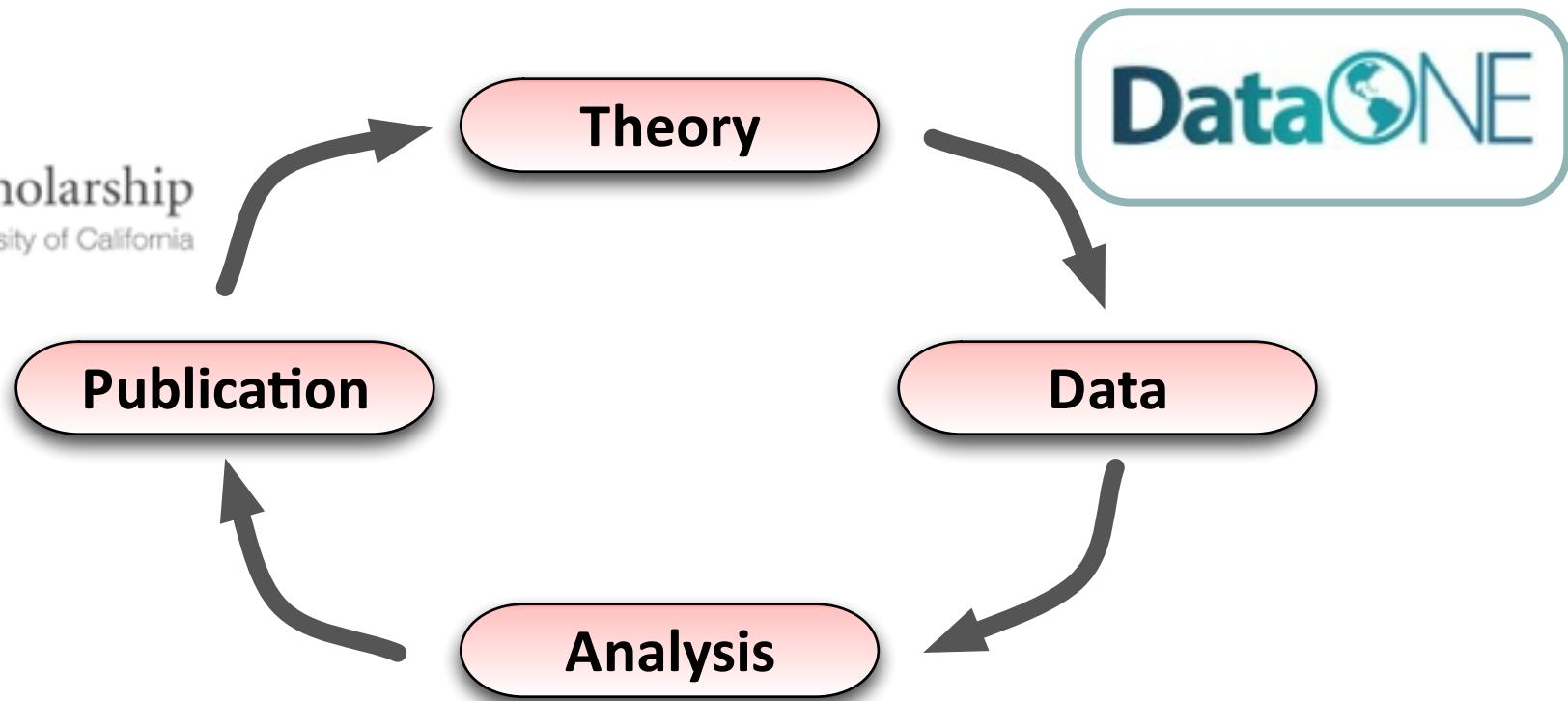
- Convergence of
 - Open access
 - Open source
 - Open data
- Facilitated by:
 - Data Federations
 - Open analysis platforms
 - Scientific workflows
 - Social media platforms
 - Digital Object Identifiers



A new vision for reproducible science



eScholarship
University of California



Questions?

- <http://www.nceas.ucsb.edu/ecoinfo/>
- <http://knb.ecoinformatics.org/>
- <http://sonet.ecoinformatics.org/>
- <http://dataone.org>