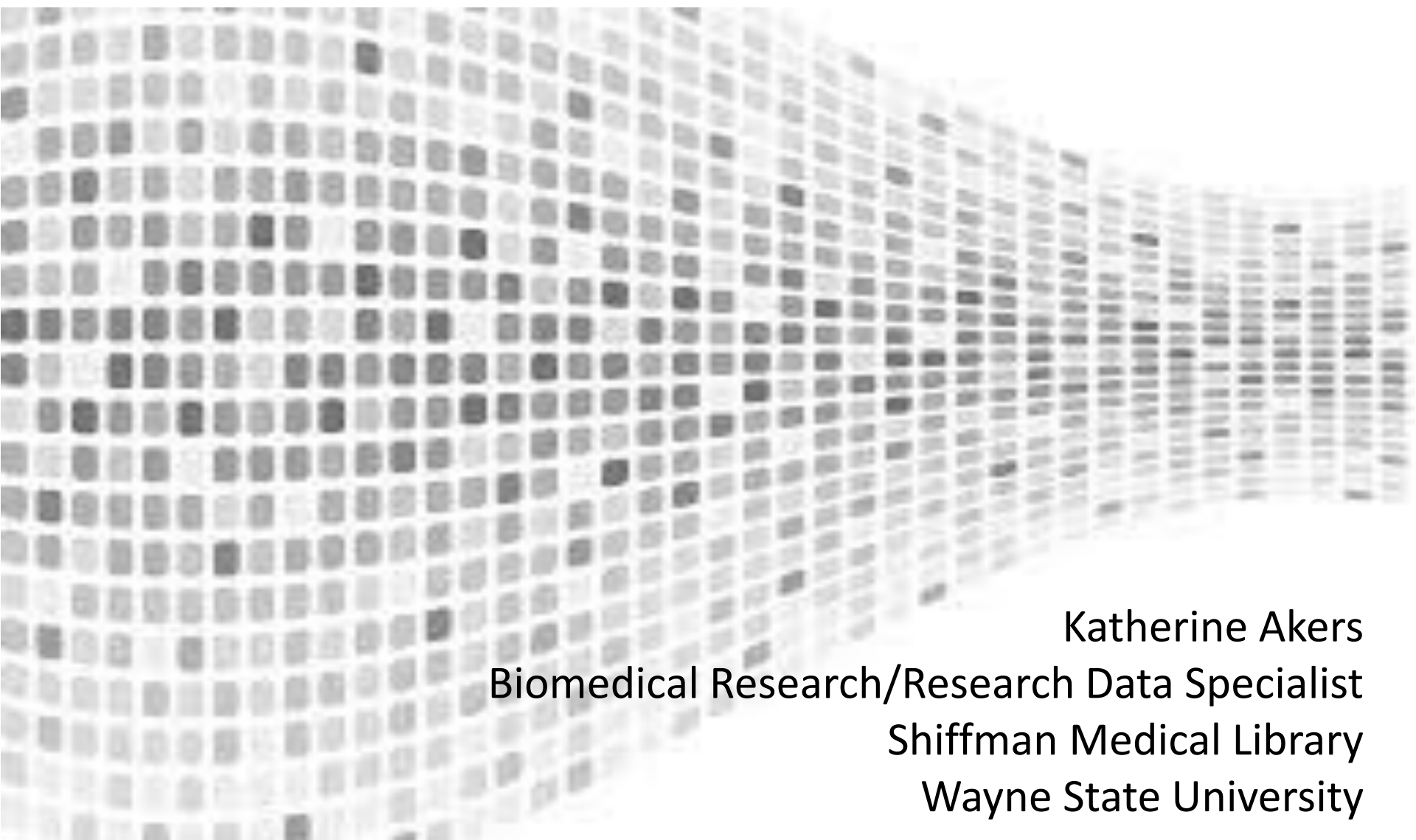


# **The library's role in promoting the sharing of scientific research data**



Katherine Akers  
Biomedical Research/Research Data Specialist  
Shiffman Medical Library  
Wayne State University

# Funding agency requirements



Starting in 2003, applications requesting \$500k/year in direct costs are required to include a **data sharing plan** (1-2 paragraphs) or state why data sharing is not possible.



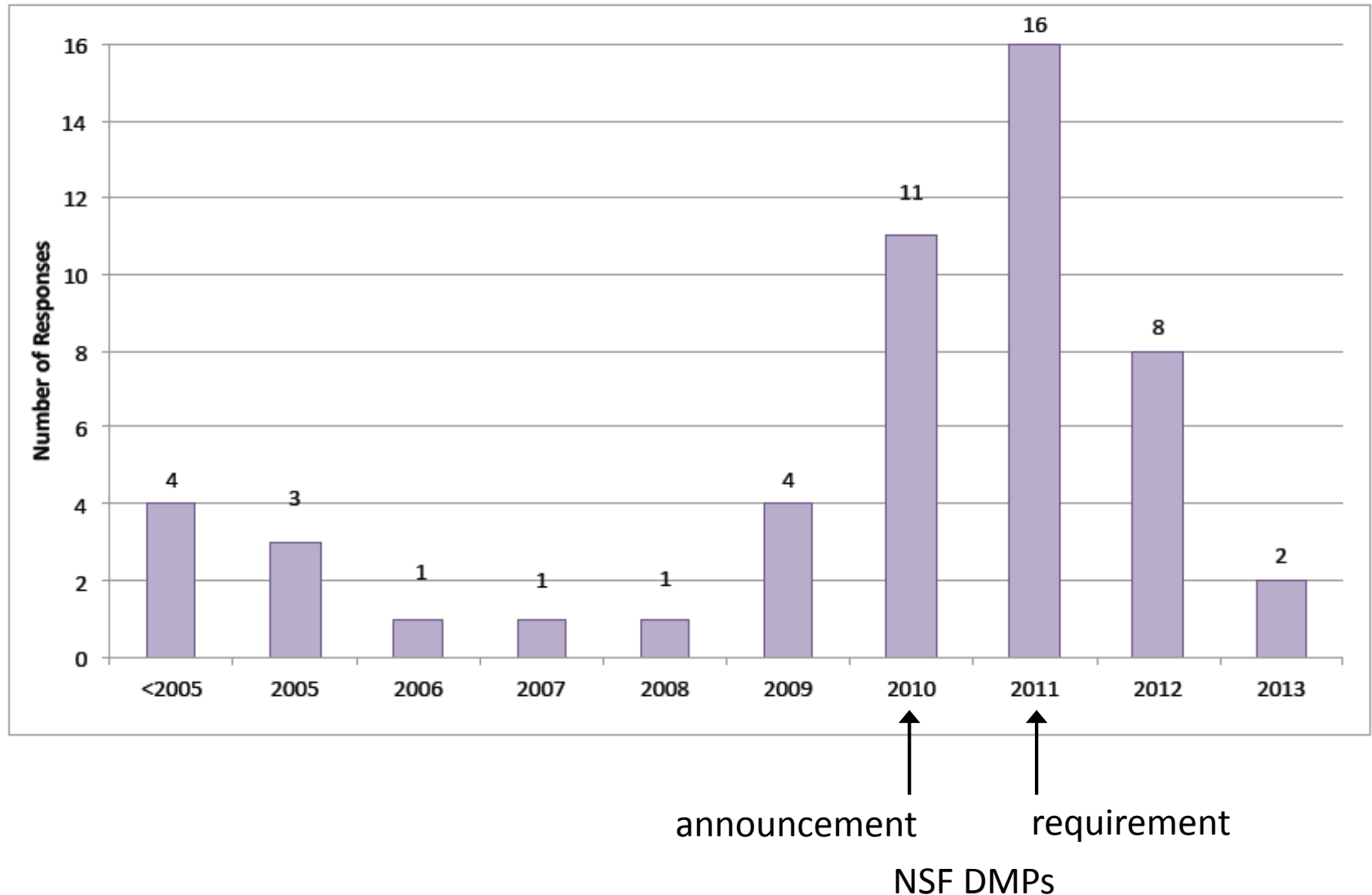
Starting in 2011, all applications are required to include **data management plan** 1-2 pages) describing how data will be disseminated and shared.

# NSF Data Management Plan (DMP): Biological Sciences Directorate



- Describe data that will be collected and metadata standards that will be used.
- Describe the resources/facilities that will be used to store and preserve the data after the grant ends.
- Describe dissemination methods that will be used to make data available to others after the grant ends.
- Describe policies for data sharing and public access (e.g., protection of privacy, intellectual property rights)
- Describe roles and responsibilities of all parties with respect to data management after the grant ends.

# Library research data services: Born out of the NSF DMP requirement

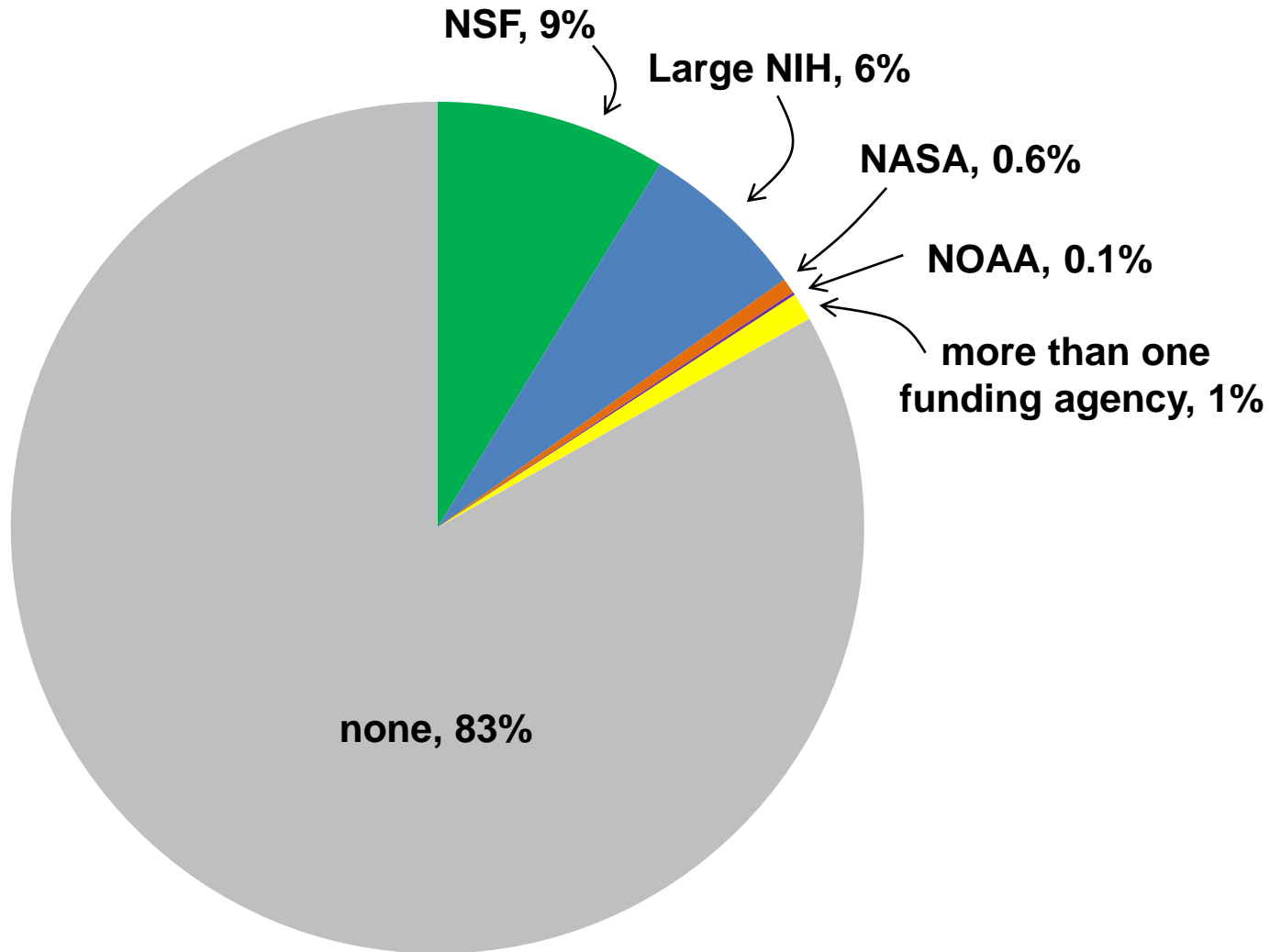


# DMP-related services

- Providing information about DMPs (libguides, etc.)
- Holding workshops on DMPs for faculty members
- Assisting with the formulation and/or writing of DMPs



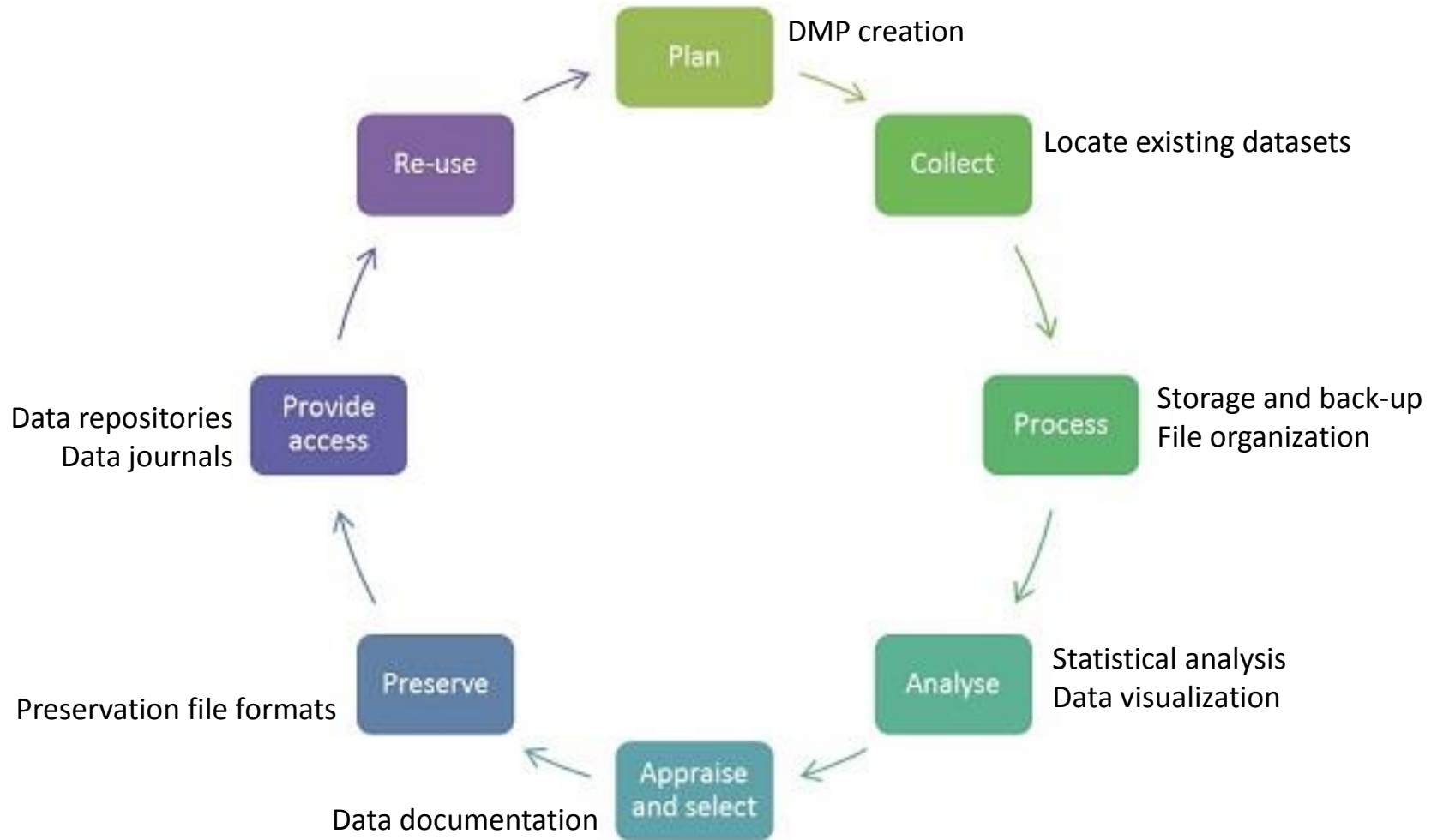
# U-M faculty with grants requiring DMPs



*“... a thin thread on which to hang an entire service.”*

—Dorothea Salo

# Consider the entire research data lifecycle



# Data sharing and management snafu in 3 short acts



<http://youtu.be/N2zK3sAtr-4>



# Data repositories

Institutional repositories



# Data repositories

## Institutional repositories



## Disciplinary repositories

GenBank

SRA

dbGaP



Influenza Research Database

# Dryad: scientific data repository

[About](#)[For researchers](#)[For organizations](#)[Contact us](#)[Login](#)[Sign Up](#)

1 DEPOSIT  
DATA



2 GET PERMANENT  
IDENTIFIER



<http://dx.doi.org/10.5061/dryad.20>

3 WATCH YOUR  
CITATIONS GROW!



4 RELAX, YOUR DATA  
ARE DISCOVERABLE  
AND SECURE

[Submit data now](#)[How and why?](#)

## Search for data

[Go](#)[Advanced search](#)

## Browse for data

[Recently published](#)[Popular](#)[By Author](#)[By Journal](#)

### Recently Published Data

Biro PA, Adriaenssens B (2013) Data from: Predictability as a personality trait: consistent differences in intraindividual behavioral variation. *The American Naturalist*  
<http://dx.doi.org/10.5061/dryad.h8c10>

Sangster G (2014) Data from: Declining rates of species described per taxonomist: slowdown of progress or a side-effect of improved quality in taxonomy? *Systematic Biology* <http://dx.doi.org/10.5061/dryad.m4g2n>

Mikheyev AS, Tin MMY (2014) Data from: A first look at the Oxford Nanopore MinION

## Be part of Dryad



Publishers, societies, universities, libraries, funders, and other stakeholder organizations are invited to become [members](#). Tap into an active knowledge-sharing network, receive discounts on submission fees, and help shape Dryad's future.

[Submission integration](#) is a free service that allows publishers to coordinate manuscript and data submissions. It makes



# **Partnerships between data repositories and academic libraries**

# Partnerships between data repositories and academic libraries

## 1. Membership

- Pay annual membership fee
- Attend annual Dryad membership meeting
- Vote on Dryad Board of Director members, bylaws, and budget
- Potential to influence the development of Dryad



**Working Together on Data Discovery, Access and Reuse:  
2014 Dryad & Dataverse Community Meeting**

**May 28, 2014: Cambridge, MA USA 8:30AM - 3:30 PM**

### **Location:**

The Institute for Quantitative Social Science at Harvard University  
CGIS South Building  
1737 Cambridge Street, Cambridge, MA 02138

### **Summary:**

Individual researchers, their institutions, publishers and funders all play a role in ensuring that research data is available for future knowledge discovery. This joint community meeting will highlight existing solutions and emerging issues in the discovery, access and reuse of research data in the social and natural sciences.

### **Program**

#### [Session 1 video](#)


- |       |   |
|-------|---|
| 08:30 | <b>Opening Reception, light breakfast provided</b>  |
| 09:00 | <b>Welcoming remarks</b>  |
| 09:10 | <b>Update on the Dryad Digital Repository</b>   |
| 09:35 | <b>Update on the Dataverse Network</b>  |
| 10:00 | <b>The Future of Open Data: What to Expect from Funders I</b><br>Peter McCartney (Program Director, Division of Biological Infrastructure, NSF) |
| 10:25 | <b>Break</b>  |

#### [Session 2 video](#)

- |       |  |
|-------|--|
| 10:45 | <b>The Future of Open Data: What to Expect from Funders II</b> |
|-------|--|

# Partnerships between data repositories and academic libraries

## 2. Provide financial assistance for data deposit



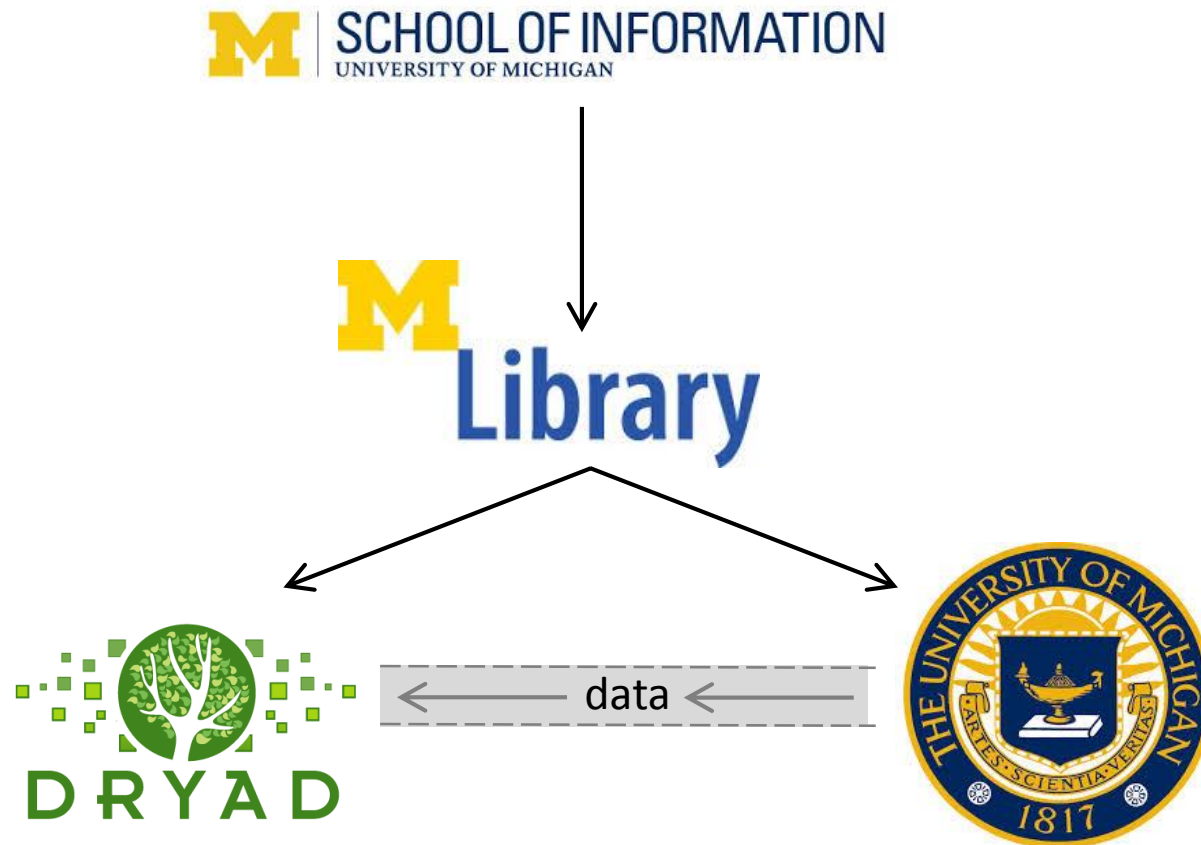
**Data Publishing Charges to help sustain open data at Dryad**

[Learn More »](#)

Payment Plan	Member	Non-member	Minimum purchase
1. Voucher Plan	\$65 per data package	\$70 per data package	25 vouchers
2. Deferred Payment Plan	\$70 per data package	\$75 per data package	1 yr contract
3. Subscription Plan	annual fee based on \$25 per published research article	annual fee based on \$30 per published research article	2 yr contract
For individuals: Pay on submission	NA	\$80 per data package, payable by the submitter	1 data package

# Partnerships between data repositories and academic libraries

## 3. Local data curator



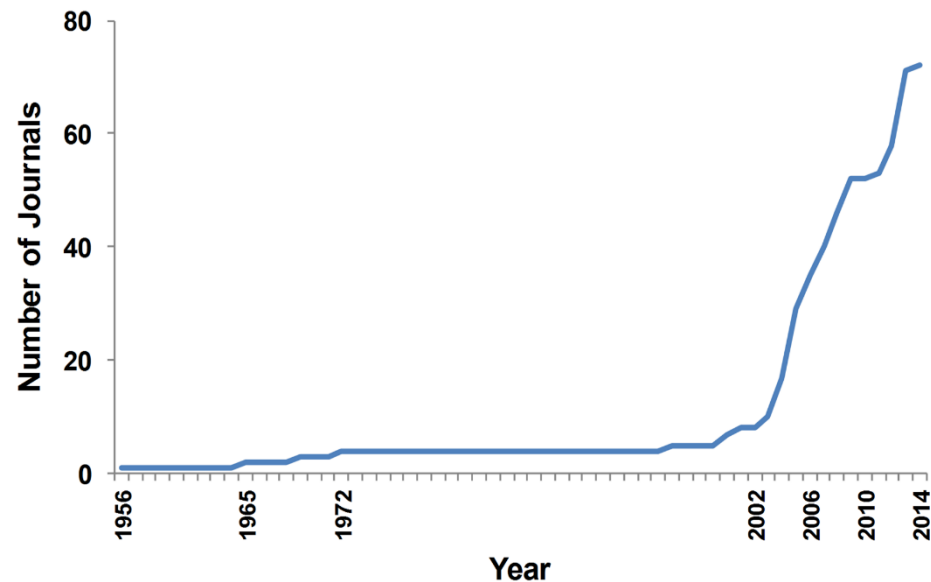
# Barriers to research data dissemination

- Data sharing takes time
- Lack of incentive





# Journals that publish “data papers”



# Nature Publishing Group's *Scientific Data*

# SCIENTIFIC DATA

110110  
0111101  
1101110  
011101101

[Advanced search](#)

[Home](#) | [Archive](#) | [About](#) ▼ | [For Authors](#) ▼ | [For Referees](#) | [Data Policies](#) ▼ | [Collections](#) ▼

## Featured Data Descriptor



### A Southern Indian Ocean database of hydrographic profiles obtained with instrumented elephant seals

Roquet *et al.*

Data Descriptor | 2nd September 2014

Our understanding of Southern Ocean currents is limited by a lack data, particularly in seasons and locations that are hard for ships to access. By affixing sensors to wild seals, these researchers have built a large database of temperature and salinity profiles, extending our knowledge of this key component of our planet's oceans.

photo by C. McMahon

Also this week: [Our first Data Descriptor Collection, from the National Evolutionary Synthesis Center \(NESCent\)](#)

## Latest content

Data Descriptor | 02 September 2014

### Artificial selection on anther exertion in wild radish, *Raphanus raphanistrum*

Jeffrey K. Conner, Cynthia J. Mills [...] Keith Karoly


Data Descriptor | 02 September 2014

### A Southern Indian Ocean database of hydrographic profiles obtained with instrumented elephant seals

Fabien Roquet, Guy Williams [...] Mike Fedak

## About *Scientific Data*

*Scientific Data* is an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

 [E-alert](#)

 [RSS](#)

 [Facebook](#)

 [Twitter](#)

 [Submit manuscript](#) ►

## Announcements

### Scientific Data Updates

a blog by *Scientific Data*

### Data Matters

[Interview with Gavin Simpson](#)

[Interview with Stephen Friend](#)

### Other blogs

Nearly 50 years of Lemur data now available online – Sarah Zehr speaks about her work @ Duke Lemur Centre

Open access is not enough on its own – data must be

# Nature Publishing Group's *Scientific Data*

The image displays handwritten mathematical derivations, likely for a scientific publication. The work is organized into several horizontal sections, each containing complex formulas and diagrams.

**Top Section:** Features the inequality  $a^2 + b^2 \leq (\bar{N} + k_0)$  and the integral  $\int \frac{d\theta}{1 + \theta^2} = \tan^{-1} \theta + C$ . It also includes the condition  $X = M$  for  $N < \bar{N} < K_a$ . A diagram of a star-like graph with six nodes is present.

**Middle Section:** Contains the formula  $V_1 = 2 \sum_{K \geq 0} \frac{(-1)^K (6K + 72)}{\sqrt{(2xy^2)(3ab + 3x)^2}}$  and the condition  $X = 0$  for  $N \geq \bar{N} + k_0$ . It also includes the formula  $X = M$  for  $N < \bar{N} < K_a$ . A diagram of a star-like graph with six nodes is present.


**Bottom Section:** Contains the formula  $V_1 = 2 \sum_{K \geq 0} \frac{(-1)^K (6K + 72)}{\sqrt{(2xy^2)(3ab + 3x)^2}}$  and the condition  $X = 0$  for  $N \geq \bar{N} + k_0$ . It also includes the formula  $X = M$  for  $N < \bar{N} < K_a$ . A diagram of a star-like graph with six nodes is present.

The derivations involve complex mathematical symbols, including integrals, sums, and square roots, and are accompanied by diagrams of star-like graphs with six nodes.

<http://youtu.be/hrHM3bUym3g>

# Example “Data Descriptor”

[www.nature.com/scientificdata](http://www.nature.com/scientificdata)

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» High-throughput screening

» Apoptosis

» Cell biology

» Cancer genetics

» RNAi

Genome-wide functional genomic and transcriptomic analyses for genes regulating sensitivity to vorinostat

Katrina J. Falkenberg<sup>1</sup>, Cathryn M. Gould<sup>2</sup>, Ricky W. Johnstone<sup>2,3</sup> and Kaylene J. Simpson<sup>2,3</sup>

Identification of mechanisms of resistance to histone deacetylase inhibitors, such as vorinostat, is important in order to utilise these anticancer compounds more efficiently in the clinic. Here, we present a dataset containing multiple tiers of stringent siRNA screening for genes that when knocked down conferred sensitivity to vorinostat-induced cell death. We also present data from a miRNA overexpression screen for miRNAs contributing to vorinostat sensitivity. Furthermore, we provide transcriptomic analysis using massively parallel sequencing upon knockdown of 14 validated vorinostat-resistance genes. These datasets are suitable for analysis of genes and miRNAs involved in cell death in the presence and absence of vorinostat as well as computational biology approaches to identify gene regulatory networks.

Design Type(s)	genome-wide siRNA screen • genome-wide miRNA screen • synthetic lethality screen • compound treatment design • strain comparison design • transcription profiling design
Measurement Type(s)	apoptosis assay • transcription profiling assay
Technology Type(s)	RNAi screening • cell culture method
Factor Type(s)	replicate • cell line • compound • assay method • treatment condition
Sample Characteristic(s)	Homo sapiens • colon • colonic cancer cell line

<sup>1</sup>Cancer Therapeutic Program, The Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002, Australia. <sup>2</sup>Victorian Centre for Functional Genomics, The Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002, Australia. <sup>3</sup>Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Victoria 3052, Australia.

Correspondence and requests for materials should be addressed to K.J.S. (email: kaylene.simpson@petermac.org)

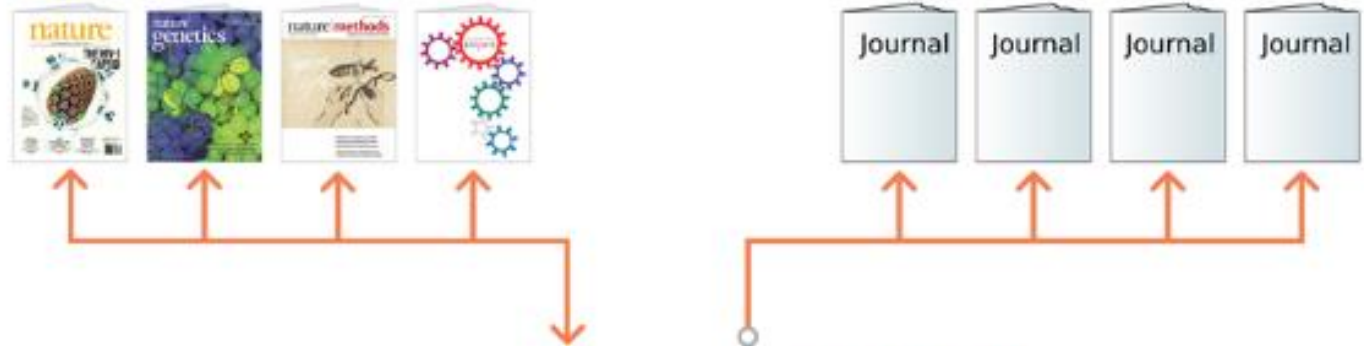
Received: 16 April 2014  
Accepted: 12 June 2014  
Published: 8 July 2014

SCIENTIFIC DATA | 1:140017 | DOI: 10.1038/sdata.2014.17

- **Background:** motivation for data collection
- **Methods:** data collection and processing procedures
- **Data Records:** data location, description of data file contents
- **Technical Validation:** quality-control measures
- **Usage Notes:** suggestions for data use

# Data journals complement use of data repositories

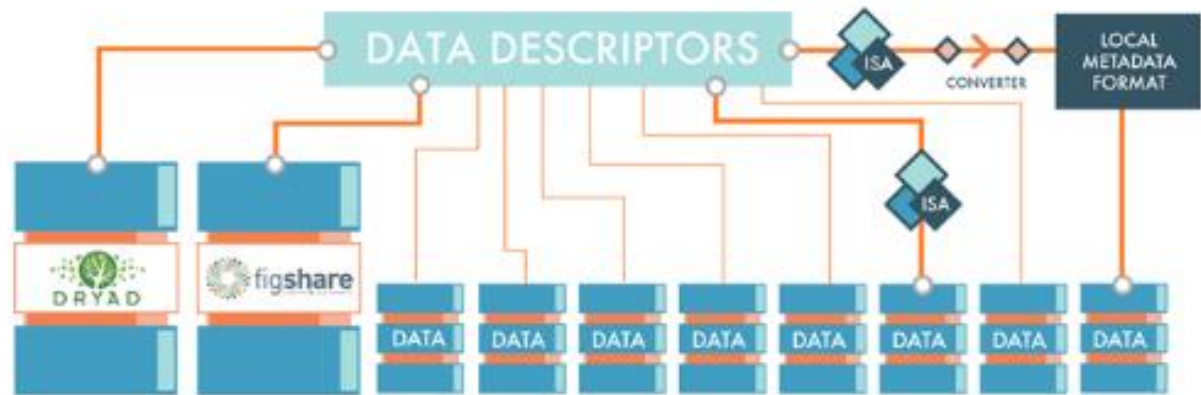
Traditional  
journal  
articles



Data  
papers

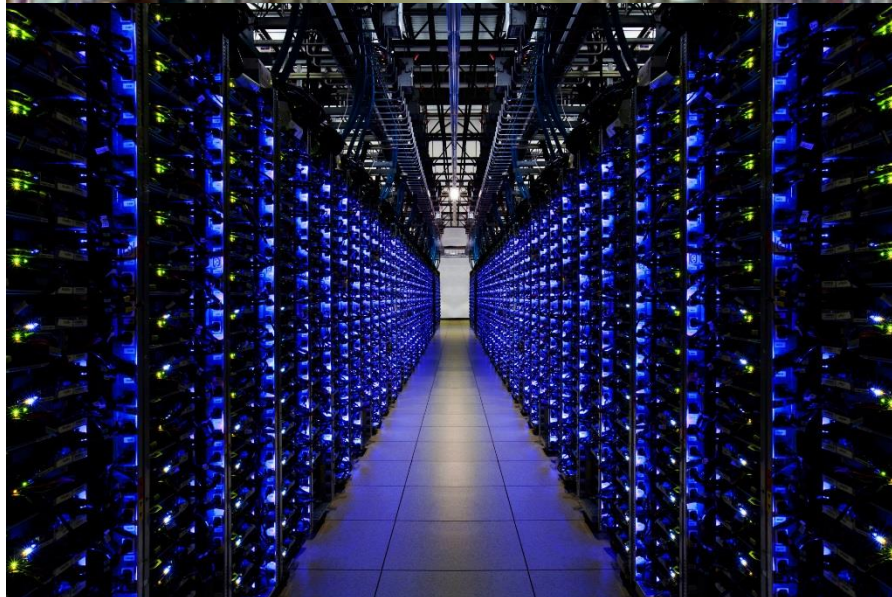
SCIENTIFIC DATA

Data  
repositories





# Data training for librarians



# Data training for librarians



# Tiered approach to data training for librarians



## 1. Research Data Concepts (basic training)

### Working with data

- File formats, naming, versioning
- Database design
- Metadata
- File organization
- Data storage
- Data security

### Sharing and preserving data

- Social barriers to data sharing
- Data repositories
- Data journals
- Data citation
- Data management planning



# Tiered approach to data training for librarians



## 1. Research Data Concepts (basic training)

## 2. Deep Dives into Data (advanced training)

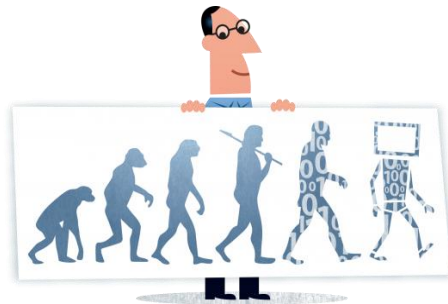
- Ecology (example science discipline)
- Psychology (example social science discipline)
- Health Sciences
- Arts & Humanities

# Deep Dives into Data (for specific subject areas)



What are the data sharing policies of the subject area's major funding agencies, academic societies, and journals?

What subject-specific data repositories are available?



What is the culture of data sharing in the subject area?

# Tiered approach to data training for librarians



1. Research Data Concepts (basic training)

2. Deep Dives into Data (advanced training)

**3. Data-related seminars (continuing education)**

- Evaluating DMPs
- Data storage options on campus
- Text mining
- Data interviews with faculty members

# At Wayne State University . . .

- Connect and collaborate with other research units across campus
- Develop workshops on research data management, scientific publishing, and research impact.
- Build research-oriented libguides for specific populations of campus researchers.
- Plan “Science Boot Camp” for librarians in the Midwest
- Establish formal *Research Data Services* program within the library system

