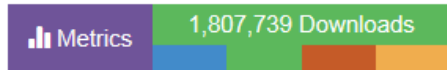




Harvard Dataverse

A collaboration with Harvard Library, Harvard University IT, and IQSS



Share, publish, and archive your data. Find and cite data across all research fields.

◀



World Agroforestry Centre  
- ICRAF Dataverse



Population Services International (PSI)  
Dataverse



INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE  
IFPRI  
International Food Policy Research Institute (IFPRI)  
Dataverse



Henry A. Murray Research Archive  
at Harvard University  
Murray Research Archive  
Dataverse

▶

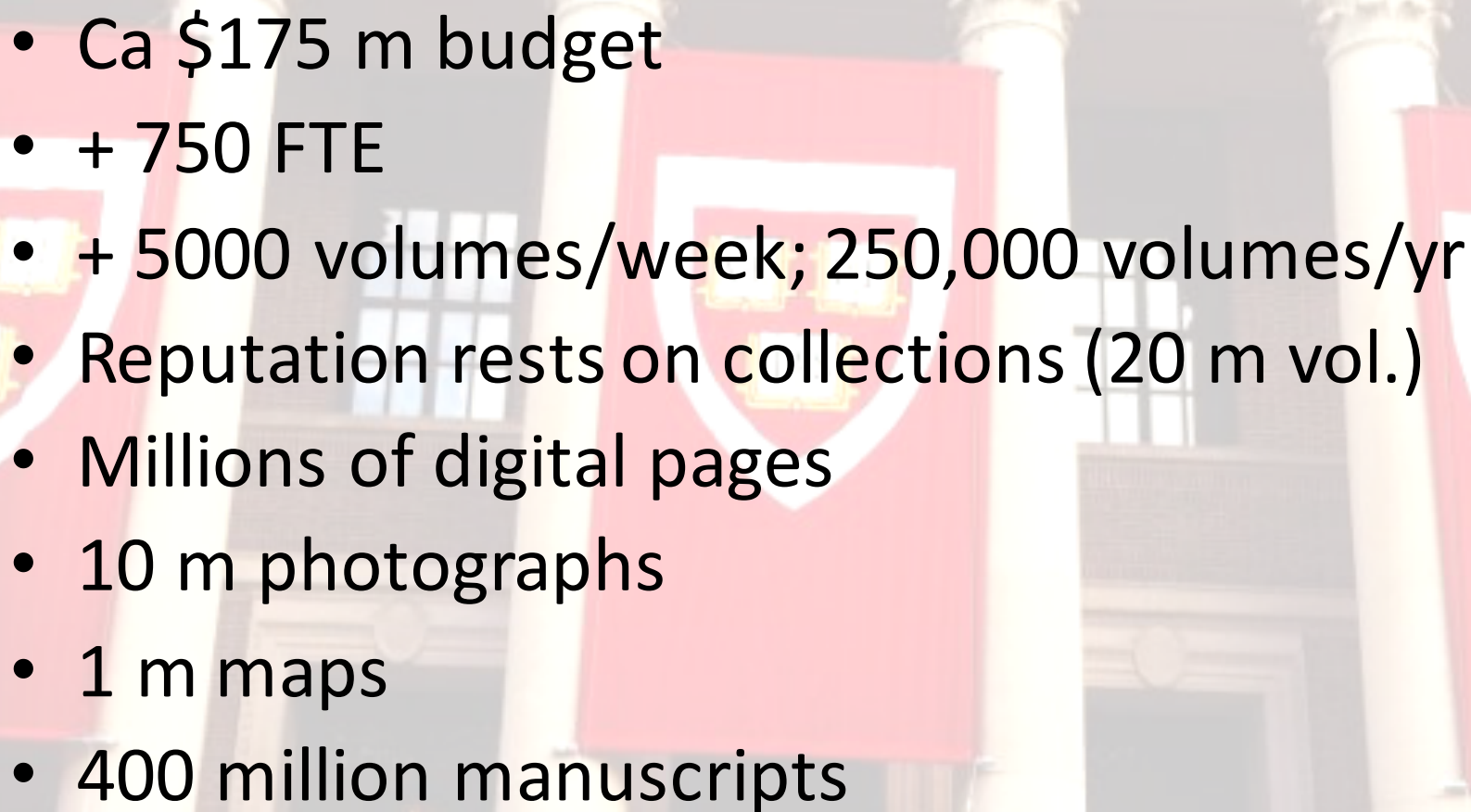
# Libraries & Data

Sarah E. Thomas

Vice President for the Harvard Library

&

University Librarian


- 
- Ca \$175 m budget
  - + 750 FTE
  - + 5000 volumes/week; 250,000 volumes/yr
  - Reputation rests on collections (20 m vol.)
  - Millions of digital pages
  - 10 m photographs
  - 1 m maps
  - 400 million manuscripts

# But What about Data?



Harvard Dataverse

A collaboration with Harvard Library, Harvard University IT, and IQSS

 Metrics

1,807,739 Downloads



Share, publish, and archive your data. Find and cite data across all research fields.



World Agroforestry Centre  
- ICRAF Dataverse



Population Services  
International (PSI)  
Dataverse



INTERNATIONAL  
FOOD POLICY  
RESEARCH  
INSTITUTE

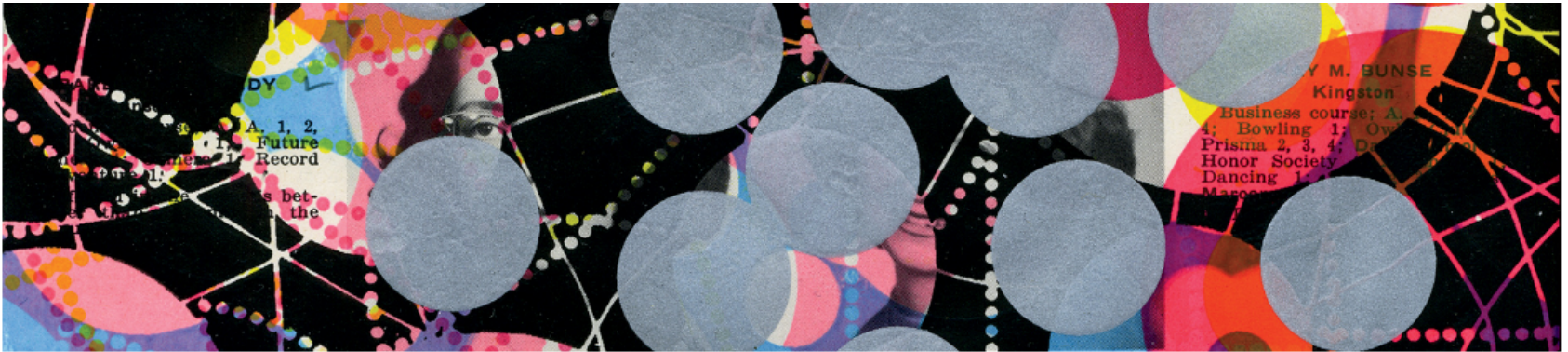
International Food Policy  
Research Institute (IFPRI)  
Dataverse



Murray Research Archive  
Dataverse



# Data, it's not just important, it's sexy.... Harvard Business School



ARTWORK: TAMAR COHEN, ANDREW J BUBOLTZ, 2011, SILK SCREEN ON A PAGE FROM A HIGH SCHOOL YEARBOOK, 8.5" X 12"

DATA

## Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

WHAT TO READ NEXT



**Big Data: The Management Revolution**

# Harvard Library Digital Strategy

- **Digitization**
- Develop a program for Pan-Harvard digitization projects.
- Update policy guiding vendor partnership for digitization from our collections.
- Create a comprehensive digital library for our reformatted collections.
- **Create sustainability plan for digital storage costs.**
- 
- **Digital Scholarship**
- Define needs for medium term storage of digital information.
- Formalize and increase library support for digital scholarship.
- Formalize and increase library support for **data research management.**
- Create a program for publishing via the Harvard Library.

Overview

Harvard Guidelines and Policies

Funding Agencies Guidelines and Policies

Writing a Data Management Plan

File Naming & Formats

Sharing Your Data

Citing Your Data

File Storage & Costs

Additional Resources @ Harvard & Beyond

## Harvard Guidelines and Policies

### At Harvard



For any research project involving data, it is important you understand your rights and responsibilities. The following links will assist you as you consider issues such as intellectual property, security and privacy.

#### [Office of Vice Provost for Research](#)

Lists research policies and contacts, as well as a [research checklist](#) that will direct you to the relevant policy or to the appropriate office you will need to contact.

#### [Office for Sponsored Programs](#)

Lists additional University policies, such as [Retention of Research Data and Materials](#), that correspond to a particular phase of the award life-cycle.

#### [General Records Schedule](#) (University Archives)

Regulates how long to keep different types of records, and whether to send records to an archive or destroy them when they are no longer needed; it covers common types of records found across the University in all formats – including electronic records and email.

#### [Office of Technology Development](#)

Includes helpful information on licensing, intellectual property, material transfer along with other guidelines, policies, forms and agreements.

### Useful Contacts



Research Development (**FAS**):

[Susan Gomes](#) Director of Research Development and Strategy  
phone: (617) 496-9448

[Erin Cromack](#), Research Development Officer  
phone: (617) 496-5252

Sponsored Research Administration, Social Sciences (**FAS**):  
[Pamela Baker-Webber](#) Sponsored Research Administrator  
phone: (857) 488-6890

[Harvard Medical School Research Administration](#)

[Sponsored Research Administration \(HMD/HSDM\)](#)

[Research Records Management \(University Archives\)](#)

Serves Central Administration, the graduate and professional schools (other than those listed below), research

# Data Symposium



[Search & Find](#)

[Research Support](#)

[Home](#) > [Harvard Purdue Data Management Symposium](#)

## Harvard Purdue Data Management Symposium



- [Symposium Home](#)
- [Joint Introduction](#)
- [Harvard Policy for Data](#)
- [Research Data Management: Our Collaborative Future](#)
- [The Research Data Revolution](#)
- [The Researcher's Perspective](#)
- [Internal Drivers: Researchers and Data](#)
- [External Drivers: Funders and Publishers](#)
- [Repository Services](#)



#hpd2015  
[@HarvardLibrary](#)  
[@PurdueLibraries](#)

This symposium was generously sponsored by Arcadia and Purdue University.

**JUNE 16-17, 2015**

This symposium on data management incorporated visionary ideas, new concepts, and inspirational speakers.

# Data Science for Librarians

 DIGITAL SCHOLARSHIP @ HARVARD

Resources

Blog

Events

Discuss (HL only)

About

Search

[Home](#) > [Resources](#) > [Data Management](#) > [Data Scientist Training For Librarians: Course Blog](#)

## Resources

> [All Resources](#)

### Category

> [Data Management](#)

> [Open Access and Open Source](#)

> [Curation & Annotation](#)

> [Library Services](#)

> [Tools](#)

> [Communities & Groups](#)

> [Preservation](#)

DATA MANAGEMENT

## Data Scientist Training For Librarians: Course Blog

Last updated: May 15, 2013



URL <http://altbibl.io/dst4l/blog/>

People Involved Chris Erdmann, Jennifer Prentice, Louise Rubin

Departments Involved CFA, Wolbach Library

Contact Info Chris Erdmann

Data Scientist Training for Librarians course has created many valuable learning outcomes that can help anyone interested in becoming a data scientist/data librarian. Tools for analyzing and cleaning data, software for visualizing data and more at <http://altbibl.io/dst4l/blog/>



## Financial Administration

# Office for Sponsored Programs

[Blog](#) [Contact OSP](#) [Policies](#) [Feedback](#)

Search

[HOME](#) [The Award Lifecycle](#) ▾ [Training](#) ▾ [Policies & Guidance](#) [Resources](#) ▾ [About Us](#) ▾[HOME](#) / [ABOUT US](#) / [BLOG](#) /

## Launch of Data Management Planning Tool

April 26, 2016

Author: Eleni Castro

As a result of collaborations with the Office of the Vice Provost for Research, Harvard University Information Technology, and IQSS, Harvard Library has launched a customized version of DMPTool, an [online data management planning tool](#), for Harvard University. Data management plans—documents that outline what researchers will do with data during and after a project—are becoming increasingly required by funding agencies such as the National Institutes of Health and the National Science Foundation. The online tool provides step-by-step guidance for

---

### Welcome to the OSP Blog!

Please note that policies are subject to revision, and specific policies referenced in past blog posts may have been updated since the posts' original publication dates. If you have any questions regarding whether a policy referenced in an older blog post is still in effect, please visit the [Policies and Guidance](#) section of our website or [contact](#) your OSP representative.

---

[Subscribe to the OSP Blog](#)

# DMPT

## Data Management

Data Management Plans

### DMPTool

WHAT IS DMPTOOL?

USING DMPTOOL AT HARVARD

BENEFITS OF USING DMPTOOL (VIDEO)

Guidelines and Policies

File Formats and Naming

Repositories/Storage

Citing Your Data

Help

## What is DMPTool?

**DMPTool** is an online tool available to help you create and share your data management plans to meet funder requirements and as a best practice for managing your data. **DMPTool** provides step-by-step guidance for creating your own DMP and includes templates and sample plans to help you address requirements specific to Harvard and your funding sources.

## Using DMPTool at Harvard

Harvard is a DMPTool partner institution. Log in using your Harvard credentials.

- Go to [DMPTool](#). Click Get Started.
- Select **Harvard University** from the menu of DMPTool partner institutions, then click **Next**.
- Enter your Harvard credentials when prompted.

## Benefits of Using DMPTool (video)



# Data Management

## Data Management Plans

- WHY MANAGE DATA?
- WHAT IS A DATA MANAGEMENT PLAN?
- WHAT DO I INCLUDE IN A DMP?
- SAMPLE DMPS
- HARVARD LIBRARY'S ROLE IN DATA MANAGEMENT

- DMPTool
- Guidelines and Policies
- File Formats and Naming
- Repositories/Storage

## Why manage data?

- **Organize your research so it can be reproduced.** Managing your data makes it easier to understand the details and procedures relating to your data and data collection throughout the lifecycle of the project.
- **Preserve and share your data to get recognition.** The data you collect are the basis of your research. They are your unique contribution, and preserving them means that your work will be recognized by others. It also ensures that your work can support future research and facilitate new discoveries.
- **Satisfy funding requirements and Harvard research policies.** The number of granting bodies requiring that data be preserved and shared is growing. A good data management plan will help you meet the requirements of your funding agency and address preservation, documentation and verification issues. It helps reviewers understand the characteristics of your data.

## What is a data management plan?

A **data management plan**, or **DMP** (sometimes also called a *data sharing plan*), is a formal document that outlines what you will do with your data during and after a research project. Most researchers collect data with some form of plan in mind, but it's often inadequately documented and incomplete. Many data management issues can be handled easily or avoided entirely by planning ahead. With the right process and framework it doesn't take too long and can pay off enormously in the long run.

Many funding agencies, especially government funding sources, require a DMP as part of their application processes. Even if you are not seeking funding for your research, documenting a plan for your data is a best practice and will help your data comply with Harvard's policies

# Social Science and Visualization



HARVARD LAW SCHOOL  
LIBRARY

[Hours](#) [Ask](#)

## Visualization Tools

An introduction to tools for creating infographics, timelines and other data visualizations.



LAST UPDATED JUN 9, 2016

[Jess Rios](#) | [✉ Email](#)

Home

[Infographics](#)

[Timelines](#)

[Word Clouds](#)

[Charts & Graphs](#)

[Maps](#)

[Other Tools](#)

[Books on Visualization](#)

[Getting Help](#)

HOME

## Welcome!

Particularly when we talk about large numbers, it can be difficult to fully understand their impact. With an ever-increasing amount of data and information available to us, data visualization is becoming more important to help people truly understand the meaning of the information that is collected. Whether you are teaching in a classroom or presenting in front of clients, the ability to distill and contextualize data is one that will set you apart and the tools in this guide will help you to do just that.

If you are new to the world of data visualization and want an overview, try



College of Communication

Why Data Storytelling

What You Will Learn

Instructors

Schedule

Registration

JOIN OUR EMAIL LIST

# Storytelling with Data

WORKSHOPS

STORYTELLING WITH DATA  
AT BOSTON UNIVERSITY

JUNE 6-10

JUNE 13-17



At **Research Data Services**, researchers across disciplines benefit from expert consultation and training in acquiring, collecting, wrangling, analyzing, visualizing, sharing, and preserving research data. Our data professionals, in collaboration with our research librarians, work to advance data science, using data to answer scientific questions, making analysis reproducible, and promoting open data to enable long-term discovery and knowledge.



The **Social, Natural, & Engineering Science** team supports the research needs across scientific disciplines, from finding materials and managing information, to understanding the changing scholarly landscape and promoting information literacy. Information experts, together with our data team, work with students and faculty to enhance their scholarship and navigate resources across the University.

Data Sources

Research Data Management

Research Software

StatLab: Data Analytics

Social, Natural, Engineering Sciences

Workshops

People

## MANAGING YOUR DATA

### MANAGING YOUR DATA

Home

Our Services

1. Before Your Research

2. During Your Research

3. After Your Research Ends

Training and Workshops

About Us

Got data? We're here to help you manage, share, and preserve your research data. In addition to our [Data Repository for the U of M](#) curation services, the Libraries will help you navigate available campus resources throughout the data lifecycle:

### [Before Your Research Begins](#)

- Schedule a data management plan (DMP) consultation ([Request Form](#)) or use our [DMP templates](#)
- Explore [funding agency requirements](#) for data and learn best practices for getting [IRB approval](#) for sharing data.
- See [more tools for planning for data management](#)

### [During Your Research](#)

### Need help?

[Contact Us](#) with your questions and we will consult with you or point you to the right person, resource, or service on campus.

# Center for Open Science Workshop at UCLA

## UCLA Library Events

[< Events](#)



### **Increasing Openness and Reproducibility in Quantitative Research**

A Workshop by the Center for Open Science

*Tuesday, May 3, 2016 - 9:00am to 12:00pm*  
[Library Conference Center](#)

PLEASE NOTE VENUE CHANGE: The workshop will be held in the Main Conference Room (not the WEC) in the Research Library. Signs will be posted to direct attendees.

#### Audience

General Public  
Registered Attendees  
UCLA Library Staff





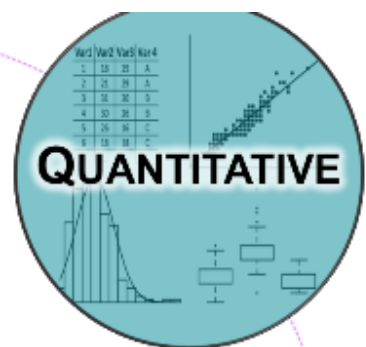
Data Services on Twitter

## Tweets by @nyudataservices

Data Services Retweeted

 **Michelle Thompson**  
@mich\_GIS

Met this #esri rockstar,  
@josephkerski , in person at  
#EsriUC2016 #EsriUC.  
@nyudataservices: wish you  
were here!



Staff member

Position

Expertise

# Data Services Specialist

Stephen Balogh

Data Services Specialist

QGIS, ArcGIS, Ruby,  
JavaScript

Andrew Battista

Librarian for Geospatial Information Systems

Google Fusion Tables,  
Google Earth, ArcGIS

# GIS

Sarah DeMott

Qualitative Data Analysis & Survey Design Specialist  
ATLAS.ti, Qualtrics, NVivo

# Qualitative Data Analysis & Survey Design

Denis Rubin

Senior Academic Technology Specialist (Quantitative Services)

MATLAB, SPSS, SAS,  
Getting Started with HPC  
Qualtrics, Python

# Quantative Services

Victoria Steeves

Research Data Management and  
Reproducibility Librarian

Research Data Management  
Plans, Data Management  
Workflow Funders

# Research Data Management and Reproducibility Librarian

Matthew Sumner

NYU Abu Dhabi Data Services Librarian

# NYU Abu Dhabi Data Services Librarian

Michelle Thompson

Geographic Information Systems Specialist

ArcGIS, Google Earth,  
Google Fusion Tables

# GIS

# Is Research Reproducibility the New Data Management for Libraries?



Science > General & Introductory Computer Science > Bulletin of the Association for Information Science and Technology

RDAP Review

## Is Research Reproducibility the New Data Management for Libraries?



Cynthia R.H. Vitale

Version of Record online: 24 FEB 2016

DOI: 10.1002/bul2.2016.1720420313

Copyright © 2016 American Society for Information Science and Technology

Issue



Bulletin of the Association for Information Science and Technology

Volume 42, Issue 3, pages 38–41, February/March 2016

# Research Data Roadshow

## Call for Two Curriculum Designer/Presenters

---

### for Research Data Management (RDM) Roadshow

ACRL is accepting applications from prospective curriculum designer/presenters to create professional development on Research Data Management. A new, day-long workshop will provide professional development support to librarians who wish to know more about data management. The target audience for this workshop is library administrators, subject liaison librarians, and other non-data specialists. Online content will be created to complement the roadshow and may include stand alone webinars.

The curriculum designers will plan and develop learning experiences for face to face and virtual learning intended to build librarians' capacity as leaders on issues/projects of campus-wide interest. Potential topics could include:

- the role of data information literacy vis-à-vis scholarly communication
- exploring the role of the liaison within the data management conversation
- how to assist with data management plans
- organization, documentation, storage, and security best practices
- discipline-specific data concerns
  - metadata standards
  - finding data
- administrative strategies
  - defining scope, setting vision
  - creating partnerships
  - resourcing data management
  - creating a program statement for the role or set of roles responsible for data management

As they develop the curriculum, a small number of ACRL member leaders, including presenters for other related workshops, will act as advisors. The curriculum designer/presenters will build on the experience of a preconference workshop offered at ACRL 2015 on research data management.



EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF MANAGEMENT AND BUDGET  
WASHINGTON, D. C. 20503

THE DIRECTOR

May 9, 2013

M-13-13

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM:

Sylvia M. Burwell  
Director

Steven VanRoekel  
Federal Chief Information Officer

Todd Park  
U.S. Chief Technology Officer

Dominic J. Mancini  
Acting Administrator, Office of Information and Regulatory Affairs

SUBJECT: Open Data Policy—Managing Information as an Asset

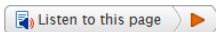
Information is a valuable national resource and a strategic asset to the Federal Government, its partners, and the public. In order to ensure that the Federal Government is taking full advantage of its information resources, executive departments and agencies (hereafter referred to as “agencies”) must manage information as an asset throughout its life cycle to promote openness and interoperability, and properly safeguard systems and information. Managing government information as an asset will increase operational efficiencies, reduce costs, improve services, support mission needs, safeguard personal information, and increase public access to valuable government information.

# Open Government Data (2016)

The screenshot shows the CONGRESS.GOV website interface. At the top, there are navigation links for Legislation, Congressional Record, Committees, and Members, along with a Sign In button. Below this is a search bar with options for Quick, Advanced, and Browse searches, and links for Search Tools, Glossary, Resources, Help, and Contact. The main content area displays the title "S.2852 - OPEN Government Data Act" for the 114th Congress (2015-2016). A "BILL" label is visible, along with a "Hide Overview" button. The page includes sections for Sponsor (Sen. Schatz, Brian [D-HI]), Committees (Senate - Homeland Security and Governmental Affairs), Latest Action (05/25/2016 Committee on Homeland Security and Governmental Affairs), and a Tracker showing the bill's progress: Introduced, Passed Senate, Passed House, To President, and Became Law. On the right side, there are links for "More on This Bill" (CBO Cost Estimates [0]), "Subject — Policy Area" (Government Operations and Politics), and "View subjects". At the bottom of the main content area, there are tabs for Summary (1), Text (1), Actions (2), Titles (3), Amendments (0), Cosponsors (2), Committees (1), and Related Bills (1).

## Summary: S.2852 — 114th Congress (2015-2016)

[All Bill Information](#) (Except Text)



There is one summary for S.2852. [Bill summaries](#) are authored by [CRS](#).

### Shown Here:

Introduced in Senate (04/26/2016)

### Open, Public, Electronic, and Necessary Government Data Act or the OPEN Government Data Act

This bill requires government data assets made available by federal agencies (excluding the Government Accountability Office, the Federal Election Commission, and certain other government entities) to be published as machine-readable data. When not otherwise prohibited by law, the data must be available: (1) in an open format that does not impede use or reuse and that has standards maintained by a standards organization; and (2) under open licenses with a legal guarantee that the data be available at no cost to the public with no restrictions on copying, publication, distribution, transmittal, citing, or adaptation.

If published government data assets are not available under an open license, the data must be considered part of the worldwide public domain. Agencies may engage with outside organizations and citizens to leverage public data assets for innovation in public and private sectors.

The Office of Management and Budget (OMB) must oversee the completeness and public availability of an enterprise data inventory that agencies must develop to account for any data assets that they create, collect, control, or maintain.



Research Article

# Approaches to Data Sharing: An Analysis of NSF Data Management Plans from a Large Research University

**Authors:** Carolyn Bishoff , Lisa Johnston

## Abstract

**INTRODUCTION** Sharing digital research data is increasingly common, propelled by funding requirements, journal publishers, local campus policies, or community-driven expectations of more collaborative and interdisciplinary research environments.

However, it is not well understood how researchers are addressing these expectations

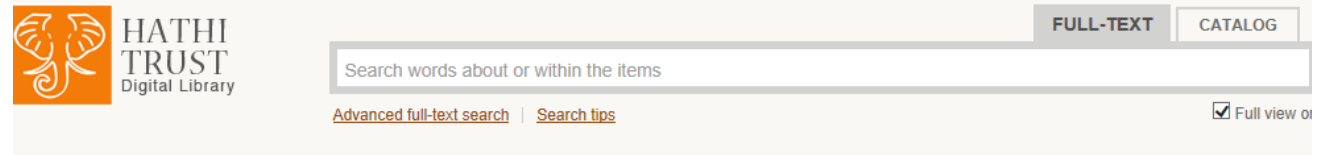
JUMP TO 0 COM

Abstract

Downloads

# What's the future?

- Expand staffing with research data skills
- Partner with others: Research, IT, faculty
- Think of digital libraries as data
- Create easy-to-find integrated and coordinated services



[Our Partnership](#) [Our Digital Library](#) [Our Research Center](#) [News and Publications](#) [Help](#)

[About](#) › One of the world's largest digital libraries opens doors to text mining scholars

## One of the world's largest digital libraries opens doors to text mining scholars

### HathiTrust Research Center now helps researchers analyze 14 million books

BLOOMINGTON, Ind./CHAMPAIGN, IL—Who influenced Charles Darwin when he was writing his pioneering theory of evolution, *On the Origin of Species*? Indiana University professor Colin Allen wants to know, and the [HathiTrust Research Center](#) may now hold the answer.