

# Research Data Services

**Harvard DataFest, January 19, 2021**

**Mercè Crosas, Ph.D., Harvard University**

University Research Data Management Officer, HUIT

Chief Data Science and Technology Officer, IQSS

[scholar.harvard.edu/mercecrosas](https://scholar.harvard.edu/mercecrosas) @mercecrosas



The Institute for Quantitative Social Science



**HARVARD UNIVERSITY**  
Information Technology

# A Recent US Report



ITHAKA S+R

[Our work](#)

[Publications](#)

[Blog](#)

RESEARCH REPORT

November 18, 2020

# Research Data Services in US Higher Education

Jane Radecki, Rebecca Springer

DOI: <https://doi.org/10.18665/sr.314397>

**Topics:** [Digital scholarship and data management](#), [Libraries](#), [Research practices](#)

**Tags:** [Data services](#)



Jane Radecki



Rebecca Springer

# Study Design and Definitions

- Review 120 US universities
- Three groups:
  - **R1:** doctoral universities: very high research activity (e.g. Harvard)
  - **R2:** doctoral universities: high research activity
  - **SLACs:** Baccalaureate colleges
- Consider research data and computing services from:
  - **Libraries**
  - **IT/Research Computing**
  - **Research center and facilities**
  - **Professional Schools** (e.g., Medical School, Business School)

**“we defined research data services as any concrete, programmatic offering intended to support researchers in working with data.”**



# Key Findings

- **Libraries** are important providers of research data services
- **IT/research computing** provide fewer research data services than libraries, but are also an important provider
- A wide variety of services are provided by **academic departments, research centers and facilities, and professional schools**
- **High performance computing** offered: 100% R1, 60% R2, 24% SLACs

# Types of Research Data Services

## Within Libraries and IT

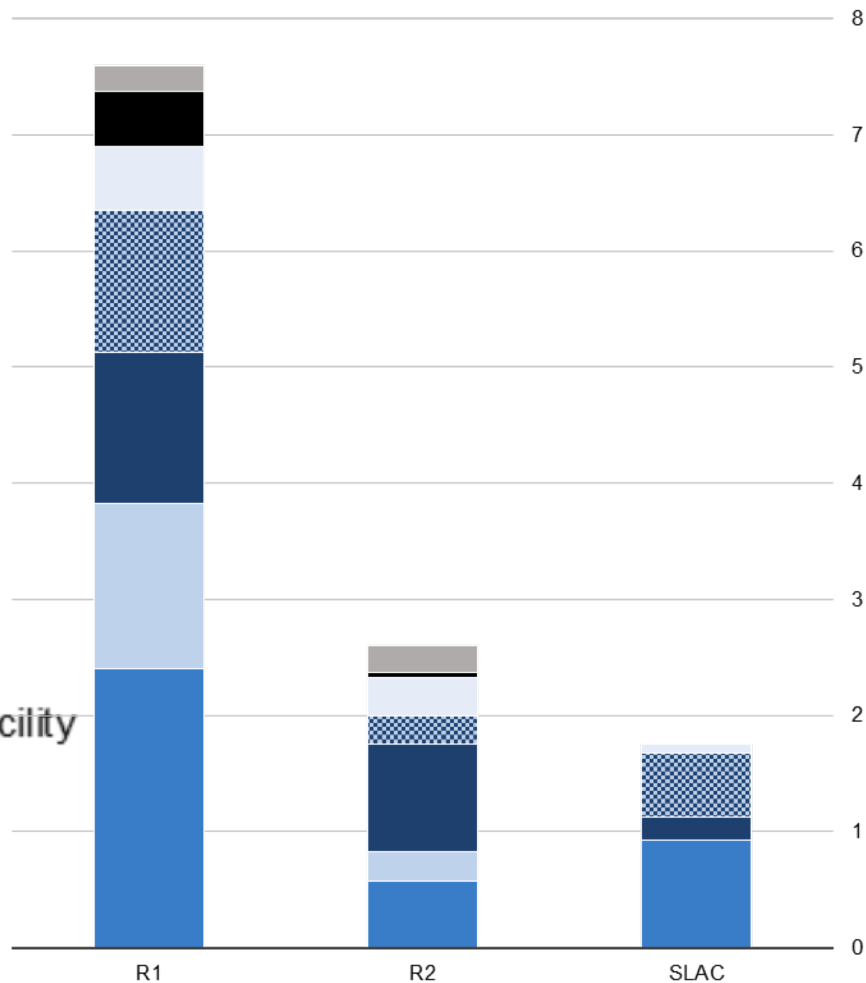
- Consulting
- Training events
- Backend work (data architecture, metadata design)
- Front end work (web development, data visualizations)

## Outside Libraries and IT

- Statistics
- Bioinformatics
- Geospatial
- Clinical data
- Business
- Social Science
- Visualizations

## Libraries are the largest contributors to Research Data Services:

- 32% in R1s
- 53% in SLACs
- Only 22% in R2s

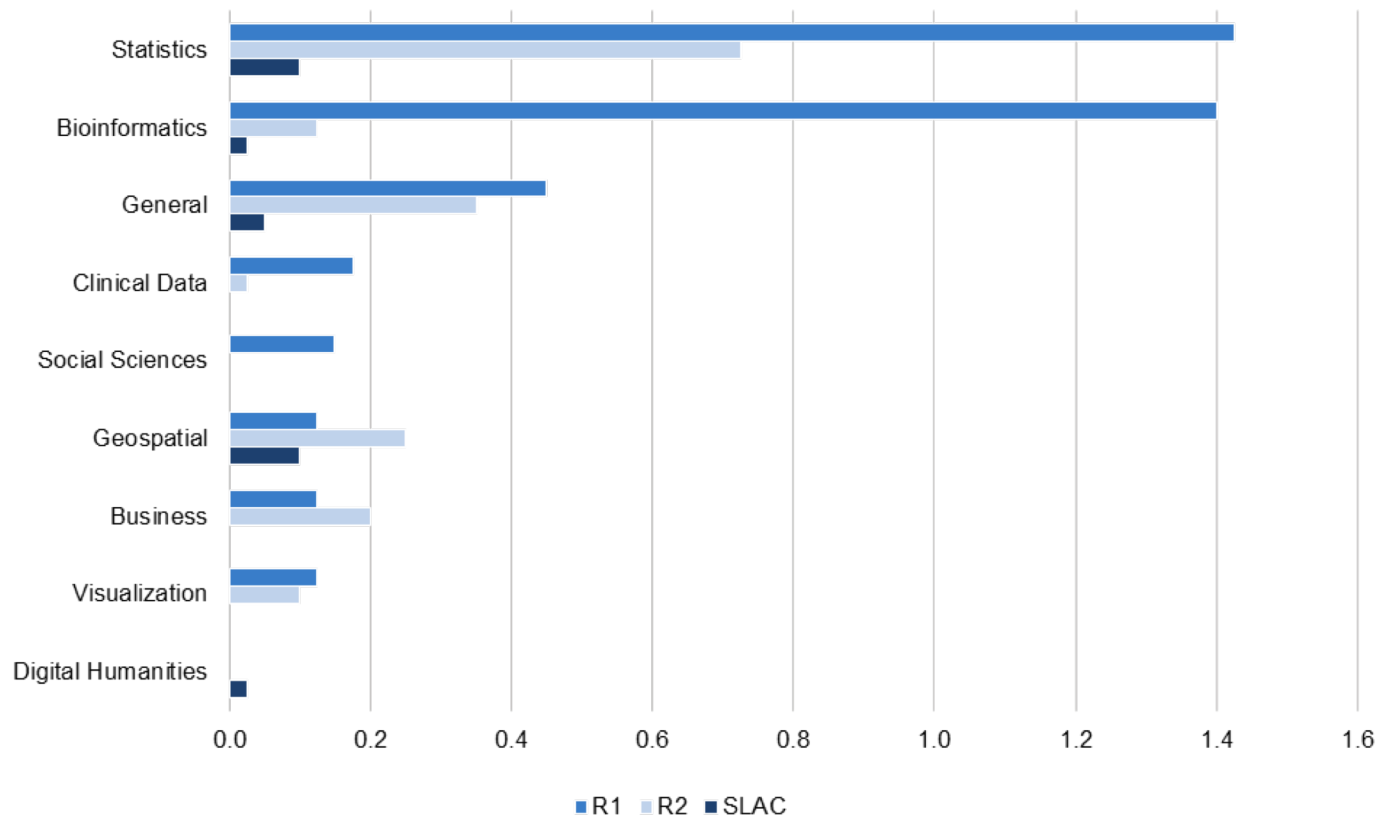


# Profile of types of Library Data Services

**Generalist consultation** is the most common service offered by the libraries

	Consulting	Training Events	Front End Work	Back End Work	Total
<b>General</b>	35.9%	16.0%	3.2%	2.6%	<b>57.7%</b>
<b>Geospatial</b>	16.7%	9.0%	0.0%	0.0%	<b>25.6%</b>
<b>Statistics</b>	7.1%	1.3%	0.0%	0.0%	<b>8.3%</b>
<b>Digital Humanities</b>	2.6%	1.3%	0.0%	0.0%	<b>3.8%</b>
<b>Social Sciences</b>	0.6%	0.6%	0.0%	0.0%	<b>1.3%</b>
<b>Health Sciences</b>	0.6%	0.0%	0.0%	0.0%	<b>0.6%</b>
<b>Other</b>	1.3%	1.3%	0.0%	0.0%	<b>2.6%</b>
<b>Total</b>	<b>64.7%</b>	<b>29.5%</b>	<b>3.2%</b>	<b>2.6%</b>	<b>100%</b>

# Average number of research data services per institution offered by centers and facilities, departments, and schools



# Report Conclusions

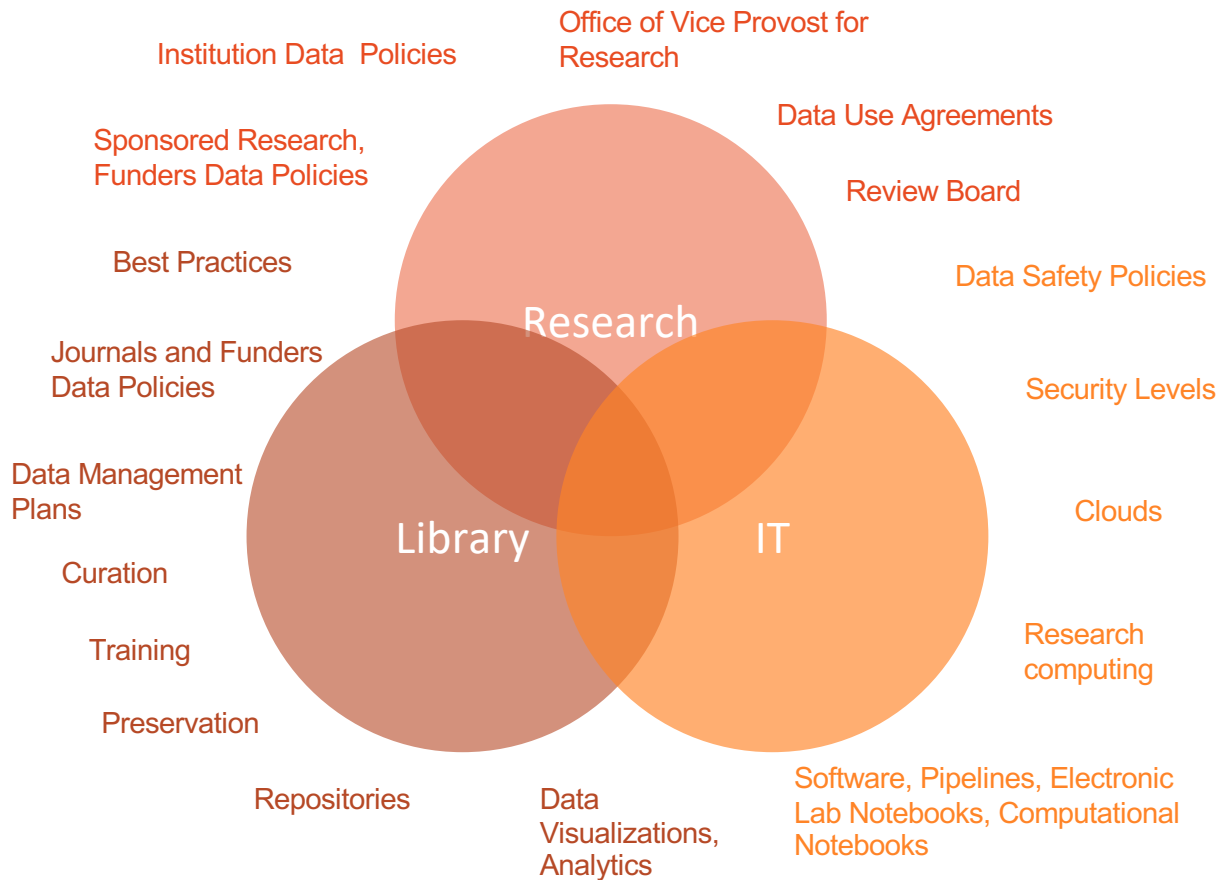
- Research Data Services in US universities are **decentralized** and **uneven**
- **Data-driven research** is increasing, but services not funded at same pace
- Should **duplicated** services merge for efficiency?
- Should **collaborative models** across universities be used to share expertise?

**“As centralized points of contact on campus, libraries, IT departments, and research offices may be particularly well positioned to act as dispatchers, connecting scholars to the services that best meet their needs”**

**Our efforts at Harvard**

**A wide variety of research services across Harvard.**

**Collaboration between Research offices, the Library, and IT is key.**





# A new inventory of research support services and a single resource to find them

- A collaboration on building a research support services catalog and a website with a **common vision**:
  - *“To help faculty, researchers, and those who work with them to advance their research by easily finding and browsing the University’s breadth of resources and services”*
- Sponsored by the Library, HUIT, and Office of Vice Provost for Research
- Initial launch planned for: Early 2021

[SUGGEST A RESOURCE](#)

# Research Support at Harvard

[Services ▼](#) [Lifecycle ▼](#) [Research Remotely](#) [About](#) [Get Help](#)

Explore the breadth of the  
University's resources and services

<https://researchsupport.harvard.edu/>

# Build inventories of existing services

- **Collect data:** Phase I, 2018-2020
  - Harvard Library (11 libraries/units)
  - Research Computing (4 schools/units)
  - Research Administration and Compliance (2 units)
  - Ethics Board (IRB), Data Use Agreements, Data Safety (3 units)
- **Create a catalog** of a total of 34 service offerings
- Focus on the **service function**, not on the service provider

# Standardize services information across units

- Classify and describe the services offerings in a **unified and uniform** way
- Three main **Services**:
  - **Research Data and Scholarship Services**: 22 service offerings
  - **Research Computing**: 6 service offerings
  - **Research Administration and Compliance**: 6 service offerings
- Three phases of the **Research Lifecycle**:
  - **Planning**: 12 service offerings
  - **Active Research**: 18 service offerings
  - **Dissemination and Preservation**: 4 service offerings

# Coordinate through working groups

- In decentralized universities, **working groups** can help establish and **achieve unified goals** and **communicate** across schools and units

In the last couple of years, we created the following groups relevant to this project:

- Working group to help **coordinate research data management** efforts
- Working group for **building the research support website**

# Browse service offerings by **three main services**

## Services

### Welcome

Harvard researchers have access to a wide range of service offerings across the University. From planning a project or study, to sharing and archiving methods or findings, our services span the entire research lifecycle. This website brings together Harvard's offerings across central units and schools, including support for research administration and compliance, data management and scholarship, and research computing.

### Browse by Services

#### Research Administration and Compliance →

Harvard offers a full spectrum of resources to support and facilitate research and researcher compliance with internal and external regulations and policies. Services available

#### Research Computing →

Research Computing at Harvard facilitates the advancement of research by providing leading-edge computing services including cluster computing, storage, software licenses, virtual instances, and

#### Research Data and Scholarship →

Researchers at Harvard generate data and scholarship that changes the world. Services across the University are available to support data creation, curation, and transformation, as well as research publishing and

# Uniformity across research support offerings

Same fields for each service offerings: Audience, Provider, Fee, Site, Contact

SERVICES
• Research Administration & Compliance
• Research Computing
• <b>Research Data and Scholarship</b>
Archiving Faculty Research Data and Archiving Data
<b>Buying and Licensing Data</b>
Copyright and Intellectual Property
Data Cleaning
Data Curation
Data Deposit
Data Handling
Data Retrieval
Data Security Support
Data Sharing and Publishing

HOME / SERVICES / RESEARCH DATA AND SCHOLARSHIP /

## Buying and Licensing Data

Consultations and instruction associated with obtaining, buying, and licensing research data.

### Details by Provider

- Harvard College, Services for Academic Programs
- Harvard Law School
- Harvard Kennedy School
- Gutman Library
- Baker Library

SERVICES
• Research Administration & Compliance
• Research Computing
• <b>Research Data and Scholarship</b>
Archiving Faculty Research Data and Archiving Data
<b>Buying and Licensing Data</b>
Copyright and Intellectual Property
Data Cleaning
Data Curation
Data Deposit
Data Handling
Data Retrieval
Data Security Support
Data Sharing and Publishing

HOME / SERVICES / RESEARCH DATA AND SCHOLARSHIP /

## Buying and Licensing Data

Consultations and instruction associated with obtaining, buying, and licensing research data.

### Details by Provider

#### • Harvard College, Services for Academic Programs

Harvard College Library, Services for Academic Programs (SAP) offer consultations to help researchers and other library staff identify data and coordinate with bibliographers and collection development librarians on possible purchases.

#### Audience

All Harvard community; focus on FAS undergraduates, graduate students, and faculty

#### Service Provider

Harvard College Library, Services for Academic Programs

#### Service Fee

None

#### Service Website

<https://library.harvard.edu/collections/data-and-government-information-collections>

#### Contact Information

Hugh Truslow and Diane Sredl [govdocs@fas.harvard.edu](mailto:govdocs@fas.harvard.edu)

# Browse services offerings by **research lifecycle phases**

## Research Lifecycle



The research lifecycle refers to the (often iterative) process of conducting research, from the initial planning, funding, and research project design to publishing and disseminating the conclusions or work of scholarship. Although the research process varies across disciplines and research domains, it often includes validating a model or hypothesis by using information and data. In turn, the results from the data help improve the model and thus, gather additional data to validate the new model. On this site, we refer to data in the broadest sense of the word, including experimental, observational, acquired, and simulated data, as well as any relevant information, artifacts, and original sources. In recent years, the research lifecycle has also included publishing

the data, code, and workflows to facilitate the reproducibility of the published results.

### Browse by Research Lifecycle

#### Planning →

[Buying and Licensing Data](#)

[Data Retrieval](#)

#### Active Research →

[Cluster Computing](#)

[Data Cleaning](#)

#### Dissemination & Preservation →

[Archiving Faculty Research Data and Archiving Data](#)

### **Planning:**

Access & Reuse

Plan & Design

### **Active Research:**

Collect & Create

Analyze & Collaborate

### **Dissemination &**

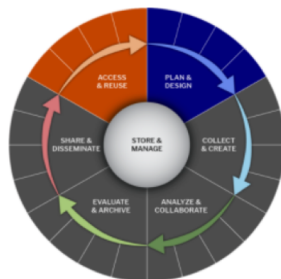
### **Preservation:**

Evaluate & Archive

Share & Disseminate



# Planning



Research planning concerns all aspects of preparing for a research project. It includes seeking funding, awareness of University and sponsor requirements, and the organization of data, records, tools, and/or resources needed to conduct the research and disseminate and archive valuable results.

## Animal Research Resources →

The University has established a number of useful resources to support animal research...

## Buying and Licensing Data →

Consultations and instruction associated with obtaining, buying, and licensing research data...

## Data Retrieval →

Consultation on how to acquire free data or retrieve data provided by a source (e.g. Library subscriptions...

## Data Safety & Regulated Data →

The University's researchers and administrators are responsible for properly managing and securing research data...

## Data Use Agreement Processing →

The transfer of data between organizations is common in the research community...

## Finding Data →

Consultation, full service (HLS, Baker), and referrals for locating sources of research data (e.g. Library subscriptions, government sponsor, repository).

## Human Subjects Research Resources →

The University has established a number of useful resources to support human research...

## Longwood Health Informationist →

Some researchers may wish to embed a data services librarian as a health informationist in their projects...

## Pre- & Post-Award Resources →

Resources and systems for research administrators, compliance officers, and researchers to support the University's research enterprise...

## Research Data Management Lifecycle →

Consultation and support for Research Data Management lifecycle activities...

## Research Design →

Full support and consultations on the design of research projects to streamline the research process...

## Training, Workshops & Capacity Building →

Ongoing training and workshops are available across the University online, and in person when available...

# Planning: Access & Reuse Plan & Design

## 12 service offerings:

- Buying and Licensing Data
- Data Retrieval, Finding Data
- Data Safety and Regulated Data
- Data Use Agreement Processing
- Human Subjects
- Animal Research Resources
- Pre- & Post-Award Resources
- Research Data Management Lifecycle
- Research Design
- Training, Workshop, Capacity Building
- Project Health Informationist

# Active Research



The active research phase of a project may include collecting or acquiring data, information, or sources, conducting quantitative or qualitative analysis, and/or using computation resources, data storage, quantitative or qualitative tools, visualizations, or information exploration.

## Cluster Computing →

Doing computations at scale allows a researcher to test many different variables at once, thereby shorter time to outcomes, and also provides the ability to ask...

## Data Cleaning →

Data Cleaning services and consultation support for cleaning, reformatting, merging, and scraping data for analyzing, visualization and reporting.

## Data Curation →

Specialists throughout Harvard Library are available to consult about data curation, organization, and integration. In order to maintain the availability...

## Data Handling →

Consultation, instruction, and support for practices and procedures involving data (e.g., reformatting).

## Data Science and Research Computing Facilitation →

A research team can often benefit from incremental help to expand their knowledge and skills, to augment their collective skill set...

## Data Science and Research Software Engineering Collaboration →

Data Science and Software Engineering play an important role in research by creating new...

## Data Security →

Consultations and/or instruction on ensuring data security during the research lifecycle, including compliance with University policies.

## Data Visualization →

Data visualization creation and support (i.e. specialized referrals) for research projects.

## Database →

In a data analysis environment, organized collections of data need to be hosted and access granted to set of researchers. A database service provides an interface to...

## Dataset Creation →

Across the University, experts are available to consult on creating data and datasets using tools like mturk, qualtrics, and other surveys and field experiments...

## Geospatial Data →

Experts are available to consult with researchers on finding, preparing, creating, and/or analyzing geospatial data...

## Lab and Biological Safety Resources →

University-wide tracking of lab safety.

# Active Research: Collect & Create Analyze & Collaborate

## 18 service offerings

- Cluster Computing, Virtual Instances
- Research Data Storage, Database, Data Security
- Software and Platforms
- Research Computing Consulting & Facilitation
- Data Science and Research Software Engineering, Statistical Analysis, Text Analysis
- Dataset Creation, Data Cleaning, Data Curation, Data Handling, Metadata creation
- Data Visualization
- Geospatial data
- Qualitative Data Support
- Lab and biological Safety

## Dissemination & Preservation



Dissemination and preservation are increasingly important parts of the research lifecycle. Sponsors, journals, and publications often require that all inputs, outputs, how research was conducted, and what tools, data, and code were used be available and accessible, alongside results and conclusions.

List of resources for dissemination and preservation below:

### Archiving Faculty Research Data and Archiving Data →

Full service options, consultation, and instruction for faculty who need to archive their research data...

### Copyright and Intellectual Property →

Consultations and/or instruction on a wide variety of topics relating to copyright and intellectual property concerns...

### Data Sharing and Publishing →

Harvard offers consultation and instruction for researchers looking to publicly share their data and research products...

### Harvard Dataverse Repository →

Harvard Dataverse is a free, self-service data repository open to all researchers provided by any discipline, both inside...

# Dissemination & Preservation: Evaluate & Archive Share & Disseminate

## 4 service offerings:

- Copyright and Intellectual Property
- Archiving data
- Data Sharing and Publishing
- Harvard Dataverse Repository

Services	Planning	Active Research	Dissemination & Preservation
<b>Research Administration &amp; Compliance</b>	<ul style="list-style-type: none"> <li>• Data Safety and Regulated Data</li> <li>• Data Use Agreement Processing</li> <li>• Human Subjects</li> <li>• Animal Research Resources</li> <li>• Pre- &amp; Post-Award Resources</li> </ul>	<ul style="list-style-type: none"> <li>• Lab and Biological Safety</li> <li>• Data Security</li> </ul>	
<b>Research Computing</b>		<ul style="list-style-type: none"> <li>• Cluster Computing</li> <li>• Virtual Instances</li> <li>• Research Data Storage, Database</li> <li>• Research Computing Consulting &amp; Facilitation</li> <li>• Data Science and Research Software Engineering</li> </ul>	
<b>Research Data &amp; Scholarship</b>	<ul style="list-style-type: none"> <li>• Buying and Licensing Data</li> <li>• Data Retrieval</li> <li>• Finding Data</li> <li>• RDM Lifecycle</li> <li>• Research Design</li> <li>• Training, Workshop, Capacity Building</li> <li>• Project Health Informationist</li> </ul>	<ul style="list-style-type: none"> <li>• Text Analysis</li> <li>• Data Cleaning, Data Handling</li> <li>• Dataset Creation, Metadata creation, Data Curation,</li> <li>• Data Visualization</li> <li>• Geospatial data</li> <li>• Qualitative Data</li> <li>• Software and Platforms</li> </ul>	<ul style="list-style-type: none"> <li>• Copyright and Intellectual Property</li> <li>• Archiving data</li> <li>• Data Sharing and Publishing</li> <li>• Harvard Dataverse Repository</li> </ul>

**What's Next**

# What we are learning

- Increase in **data science** and **data-centric research** is transforming the way we need to provide research services to our universities
- Data science and data handling are becoming an integrated part of the **education**
- Research support services are often **distributed across schools, centers and facilities**, but they might be duplicative or depend on each other
- Some services benefit from being **centralized**, but others work better close to subject expertise
- **Collaboration and communication** are key and must be constant

# Towards an integrated solution

- Research data, computing, and compliance services should be **integrated** to each other and to the research work
- We need **integrated technology and research tools** to support the services
- Whenever possible, we should **automate and streamline** the steps.  
*For example: machine-actionable Data Management Plans and Data Use Agreements; Electronic Lab Notebooks, Computational Notebooks, and Workflows integrated with repositories*

**A Harvard Data Commons can be part of the solution by providing the interoperability and tooling needed and connecting the services with the technology.**





Thank you

Mercè Crosas, Ph.D., Harvard University  
[scholar.harvard.edu/mercecrosas](https://scholar.harvard.edu/mercecrosas) @mercecrosas