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 @mercecrosas

IQSS
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

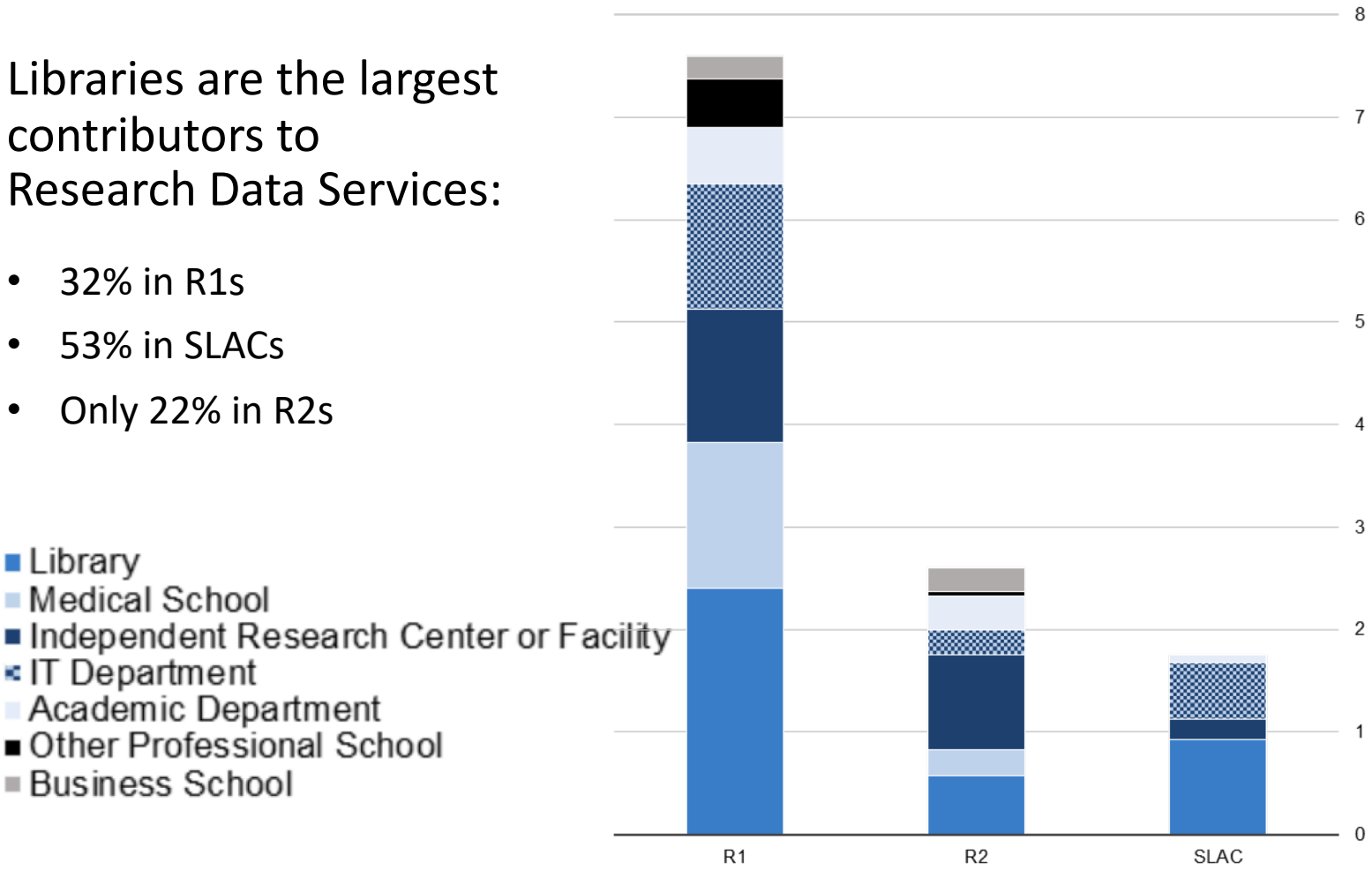
- Consulting
- Training events
- Backend work (data architecture, metadata design)
- Front end work (web development, data visualizations)

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

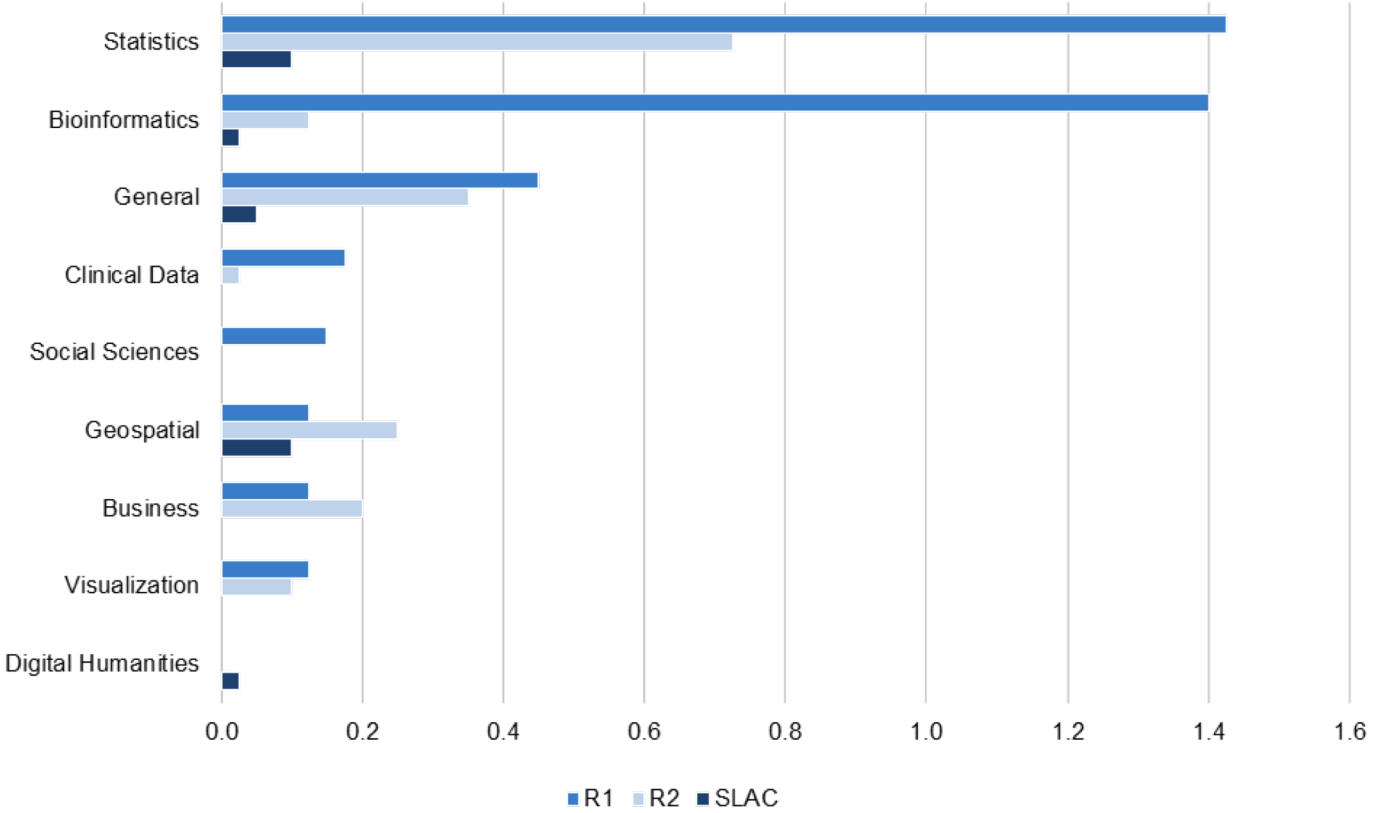


Profile of types of Library Data Services

Generalist consultation is the most common service offered by the libraries

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

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



Report Conclusions

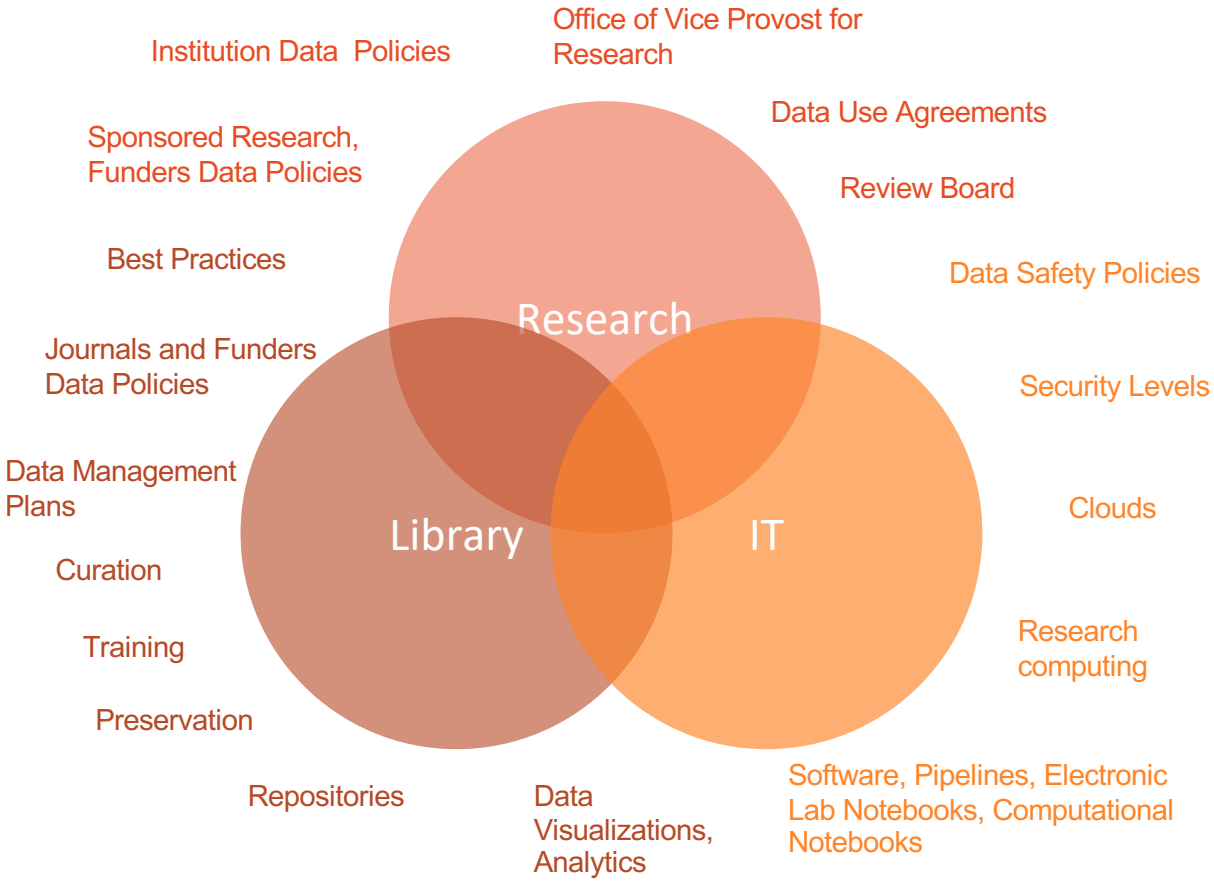
- Research Data Services in US universities are **decentralized** and **uneven**
- **Data-driven research** is increasing, but services not funded at same pace
- Should **duplicated** services merge for efficiency?
- Should **collaborative models** across universities be used to share expertise?

"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

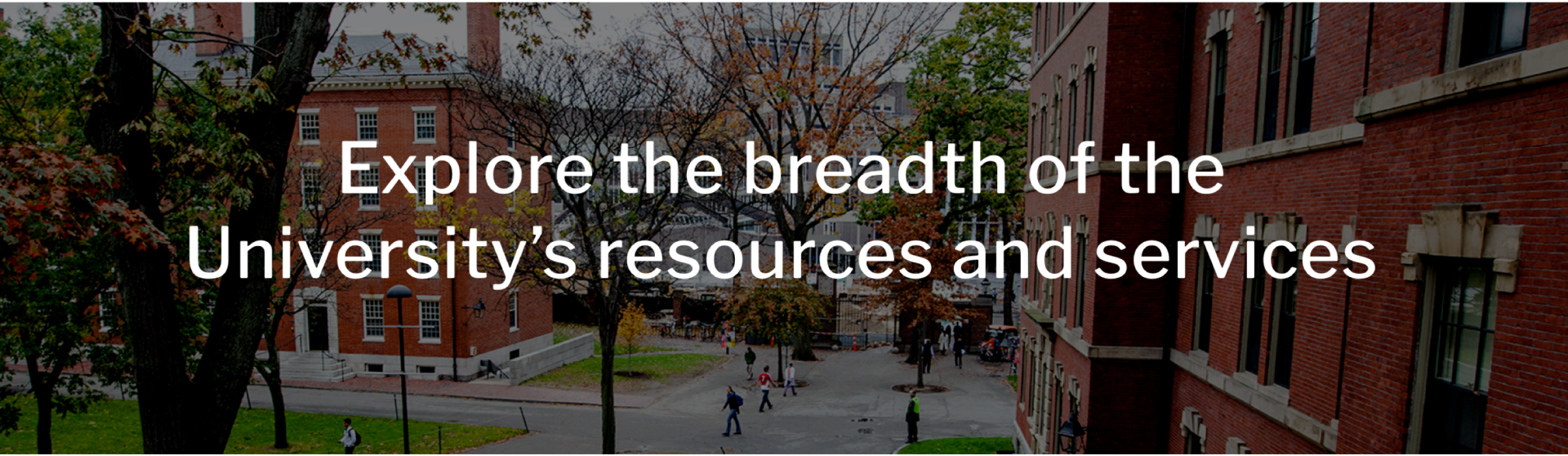
- A collaboration on building a research support services catalog and a website with a **common vision:**
  - *"To help faculty, researchers, and those who work with them to advance their research by easily finding and browsing the University's breadth of resources and services"*
- Sponsored by the Library,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

HARVARD UNIVERSITY HARVARD.EDU

SUGGEST A RESOURCE

Research Support at Harvard

Services ▾ Lifecycle ▾ Research Remotely About Get Help



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

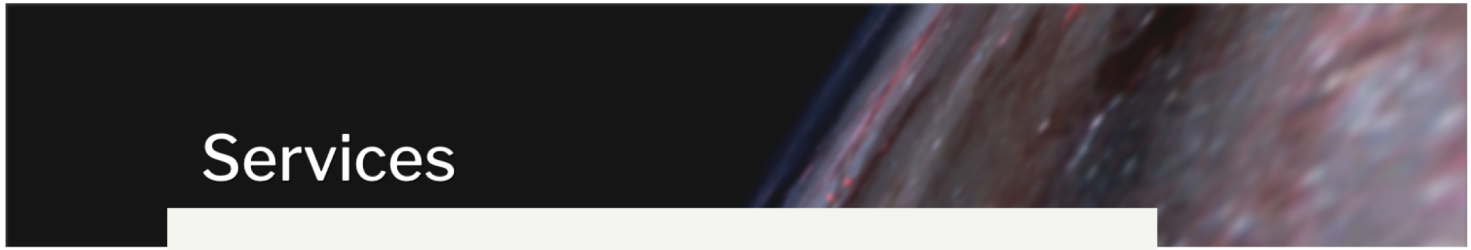
Coordinate through working groups

- 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



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 →

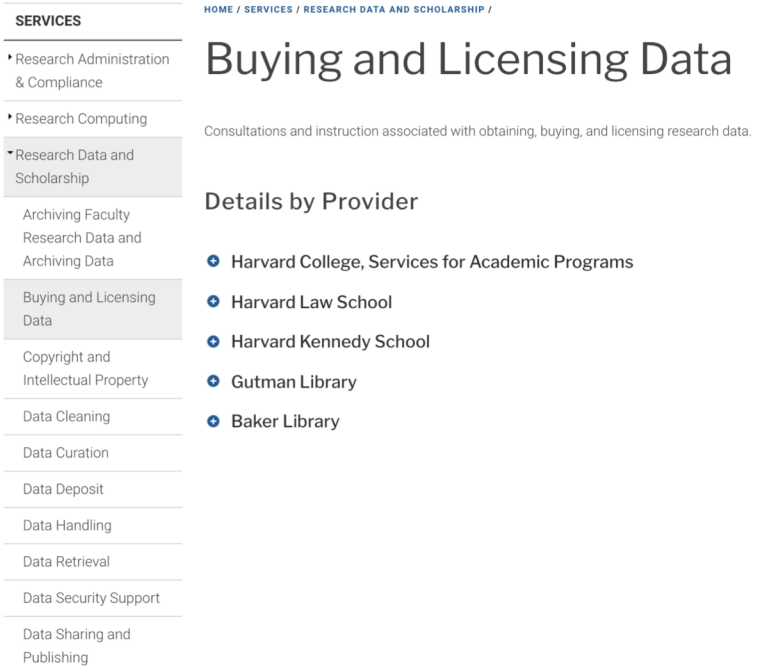
Research Computing at Harvard facilitates the advancement of research by providing leading-edge computing services including cluster computing, storage, software licenses, virtual instances, and

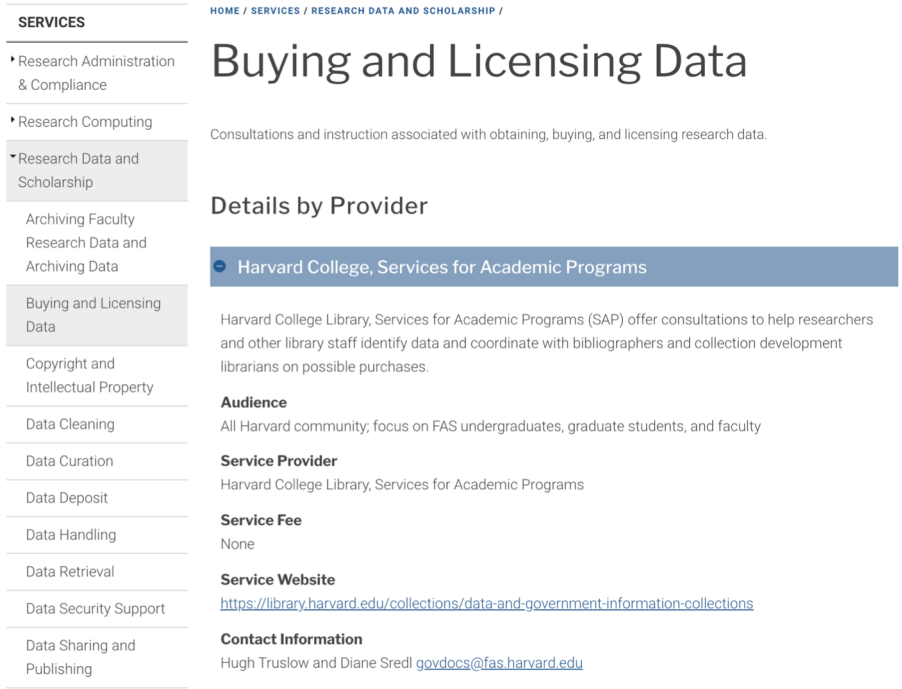
Research Data and Scholarship →

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

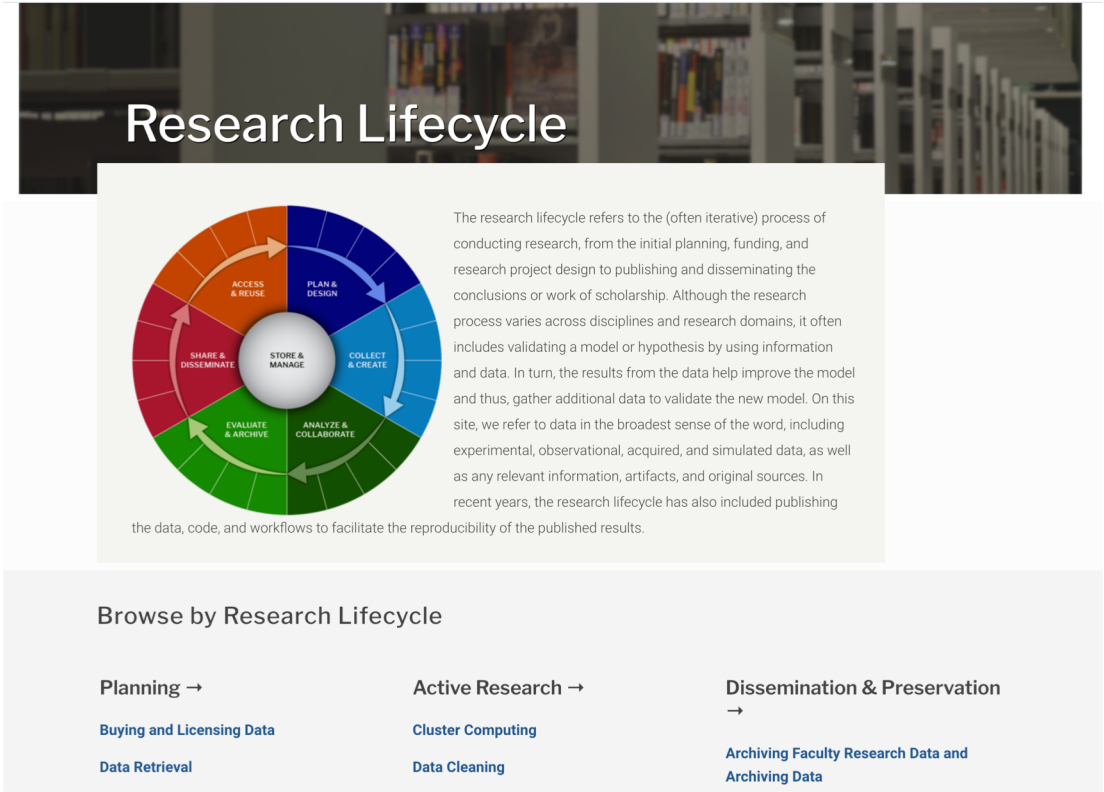
Uniformity across research support offerings

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





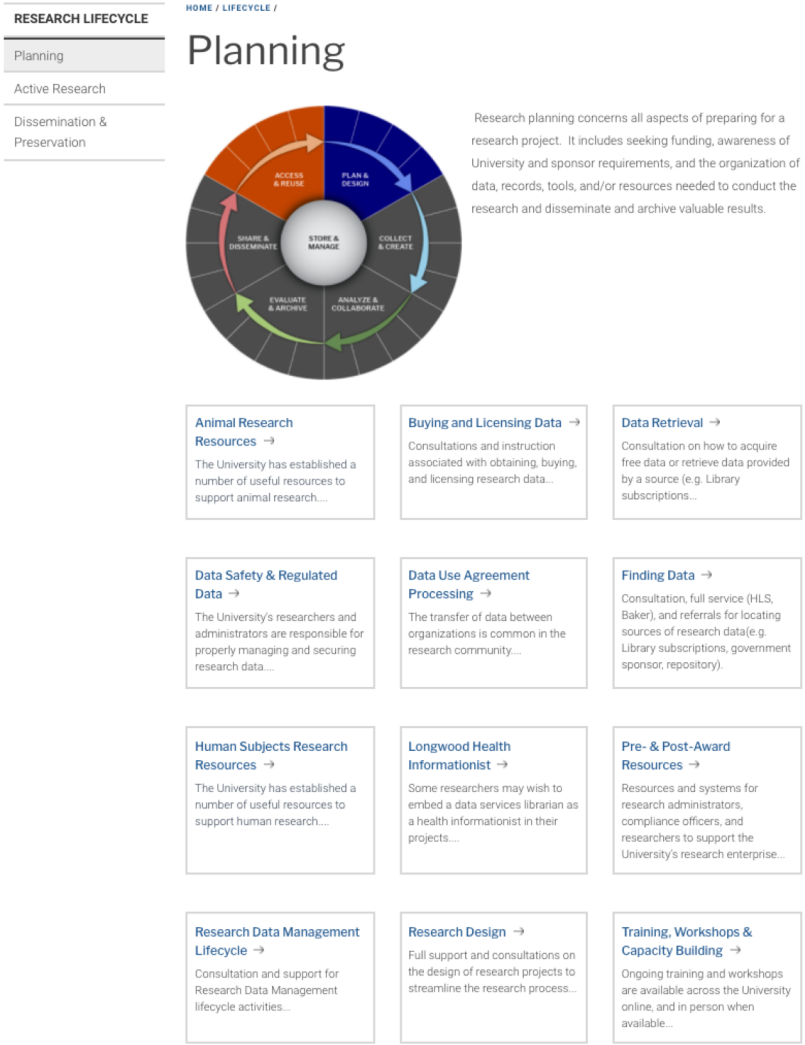
Browse services offerings by research lifecycle phases



Planning:
Access & Reuse
Plan & Design

Active Research:
Collect & Create
Analyze & Collaborate

Dissemination & Preservation:
Evaluate & Archive
Share & Disseminate



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

RESEARCH LIFECYCLE

- Planning
- Active Research
- Dissemination & Preservation

HOME / LIFECYCLE /

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....

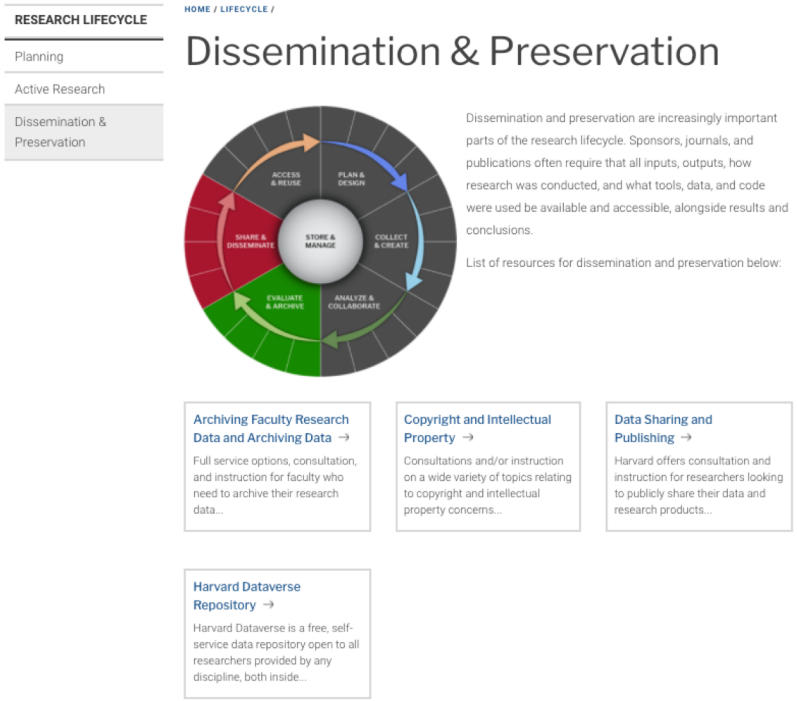
Geospatial Data →
Experts are available to consult with researchers on finding, preparing, creating, and/or analyzing geospatial data...

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:
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
Research Administration & Compliance	• Data Safety and Regulated Data • Data Use Agreement Processing • Human Subjects • Animal Research Resources • Pre- & Post-Award Resources	• Lab and Biological Safety • Data Security	
Research Computing		• Cluster Computing • Virtual Instances • Research Data Storage, Database • Research Computing Consulting & Facilitation • Data Science and Research Software Engineering	
Research Data & Scholarship	• Buying and Licensing Data • Data Retrieval • Finding Data • RDM Lifecycle • Research Design • Training, Workshop, Capacity Building • Project Health Informationist	• Text Analysis • Data Cleaning, Data Handling • Dataset Creation, Metadata creation, Data Curation, • Data Visualization • Geospatial data • Qualitative Data • Software and Platforms	• Copyright and Intellectual Property • Archiving data • Data Sharing and Publishing • Harvard Dataverse Repository

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 @mercecrosas