Integrating Services to Support Research Computing and Data:

The Harvard use case

RDA 17th Plenary, April 21, 2021

Defining, selecting and implementing interoperable and FAIR research data services

Mercè Crosas, Ph.D., Harvard University
University Research Data Management Officer, HUIT
Chief Data Science and Technology Officer, IQSS
scholar.harvard.edu/mercecrosas @mercecrosas

HARVARD UNIVERSITY

IQSS
The Institute for Quantitative Social Science

1. A unified catalog of Research Computing and Data Services

2. Towards a Data Commons to integrate services, computing, and repositories

The challenge

- **A recent increase** of research data and computing services:
  - Along with increase in data-centric and data science research
  - To support funders and journals requirements
- **Uncoordinated growth** resulting in services distributed across units and schools often disconnected

Review of Research Data Services in U.S. Universities

- Ithaka S+R report (Radecki & Springer, 2020, https://doi.org/10.18665/sr.314397):
  - Reviewed research data services from 120 U.S. Universities and Colleges
- A growing number of research data services distributed across various university units:

Within Libraries and IT (main providers)

- **Consulting (~65%)**
- **Training events (~35%)**
- 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

The Harvard use case

- **Establish a collaboration** between Research, Library, and IT/Research Computing
- **Build a catalog** to learn what services are provided across units:
  - Standardize the information
  - Unify services when possible
- **Create a research support site** to find all service offerings in a common way
  (to be announced in mid 2021; currently incorporating user testing feedback)
- **Find gaps and connect** services, tools, and teams
- **Foster a community** of research computing and data teams at the University (working groups, events)

Services offerings throughout the research lifecycle

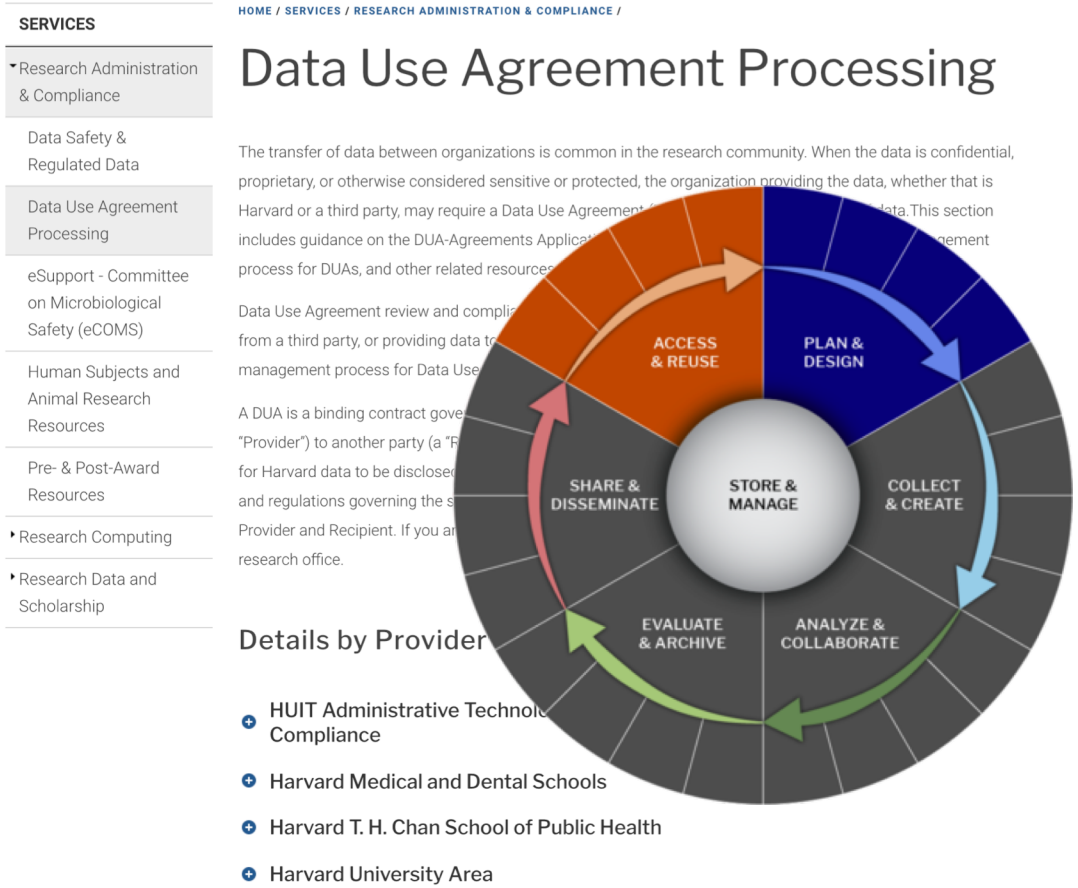
The research lifecycle refers to the (often iterative) process of conducting research, from the initial planning, funding, and research project design to publishing and disseminating the conclusions or work of scholarship. Although the research process varies across disciplines and research domains, 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.

https://researchsupport.harvard.edu/ (to be launched in mid 2021)

Planning:
Access & Reuse
Plan & Design
(14 service offerings)

Active Research:
Collect & Create
Analyze & Collaborate
(22 service offerings)

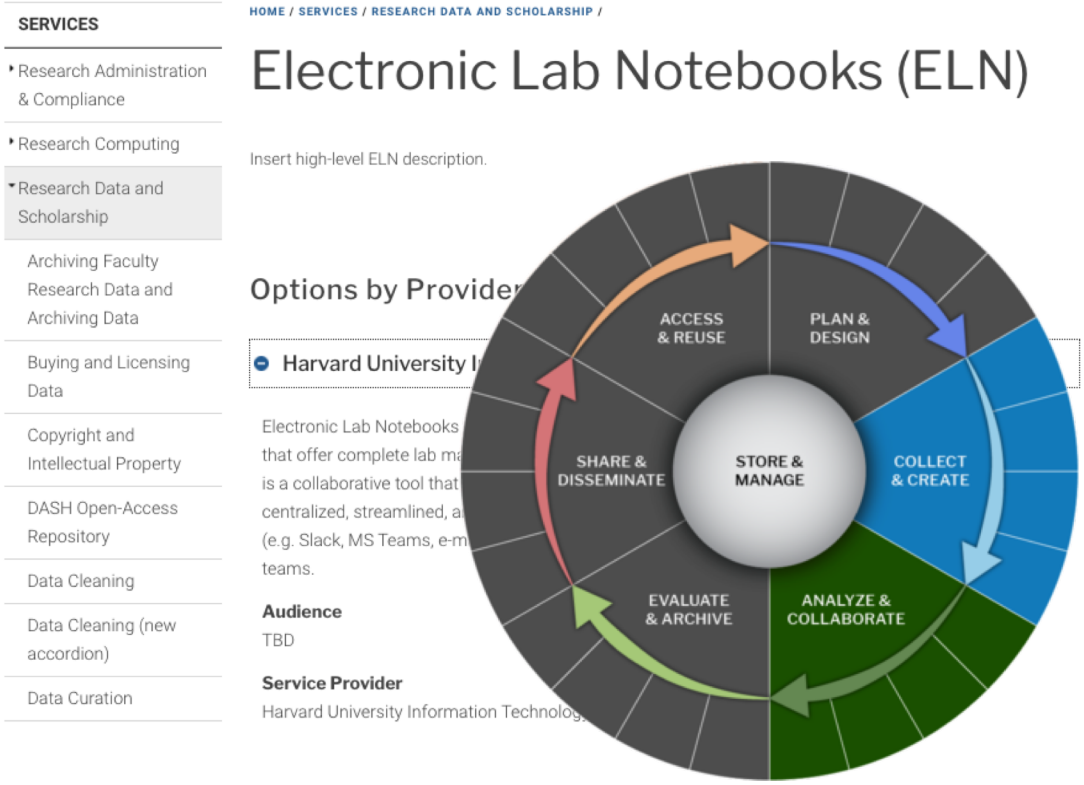
Dissemination & Preservation:
Evaluate & Archive
Share & Disseminate
(5 service offerings)



Example of Service offering in Planning Phase:

DUA Processing

- DUA and Safety System to:
  - **Track** all DUAs for incoming and outgoing compliant data
  - **Manage** DUA while data are used for research
- Assistance with DUA negotiation
- Connect process with IRB and Security officers



(under development)

Example of Service offering in Active Research phase:

Electronic Lab Notebook

- Provided through IT and supported in **collaboration** with the Library, research computing, and local labs
- **RSpace** offered University-wide; in the process of being rolled out
- **Integrates** with other services offered by the University

SERVICES

- Research Administration & Compliance
- Research Computing
- Research Data and Scholarship
  - Archiving Faculty Research Data and Archiving Data
  - Buying and Licensing Data
  - Copyright and Intellectual Property
  - Data Cleaning
  - Data Curation
  - Data Deposit
  - Data Handling
  - Data Retrieval
  - Data Security Support
  - Data Sharing and Publishing
  - Data Visualization
  - Dataset Creation
  - Finding Data
  - Geospatial Library, Data Analysis, Creation, Visualization

HOME / SERVICES / RESEARCH DATA AND SCHOLARSHIP /

Harvard Dataverse Curation

The Harvard Dataverse data curation team, staffed by member of IQSS and the Harvard Library (and separately, the Harvard Kennedy School Library), provides fee-based curation sa ions to researchers around the world who are depositing data into the Harvar

Research data replication datasets, data for relat the Harvard Dataverse. Through this engage are discoverable, accessible, interoperable

Details by Provider

Institute for Quantita

Audience

- All Affiliates
- All Faculty
- All Graduate Students
- All Undergraduate Students
- Public

Service Provider

Institute for Quantitative Social Sciences (IQ

Service Fee

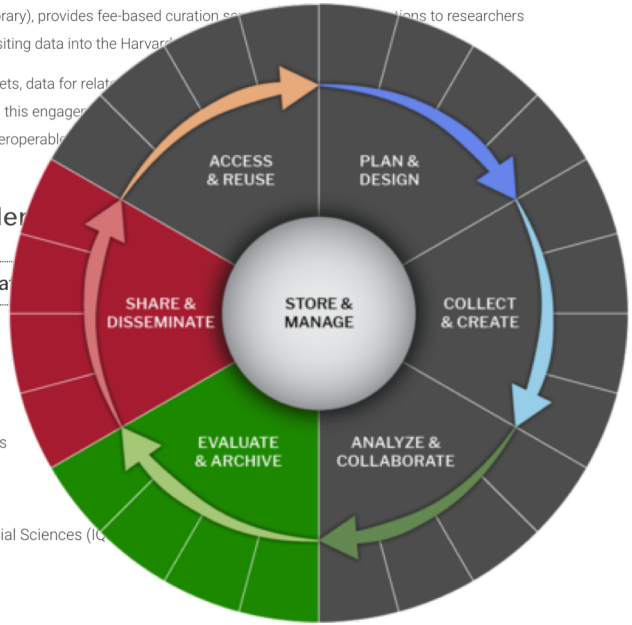
Yes

Service Website

https://support.dataverse.harvard.edu/curation-services

Contact Information

support@dataverse.harvard.edu



Example of Service Offering in Dissemination phase:

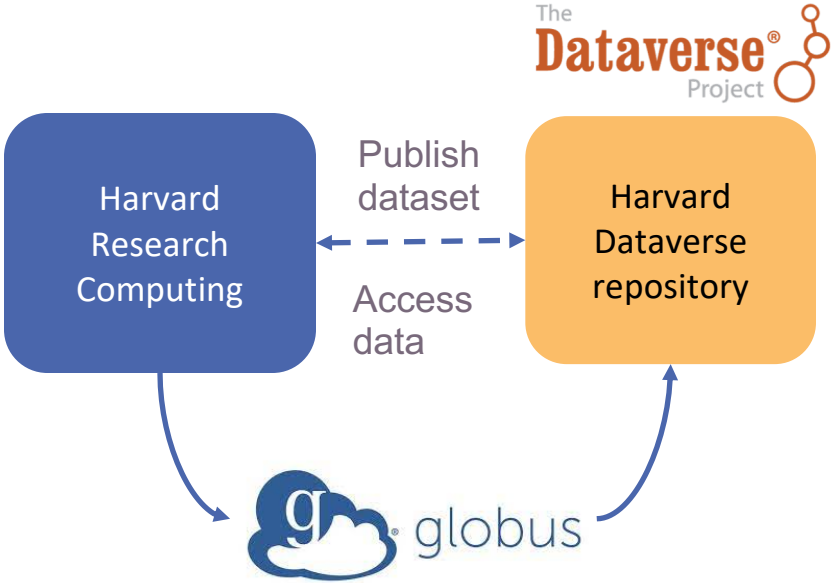
Dataverse Curation services

- A collaboration between IQSS and the Library
- Tiered service offerings
- Coming soon a new service for supporting “managed collections” interested in receiving Core Trust Seal certification.

1. A unified catalog of Research Computing and Data Services

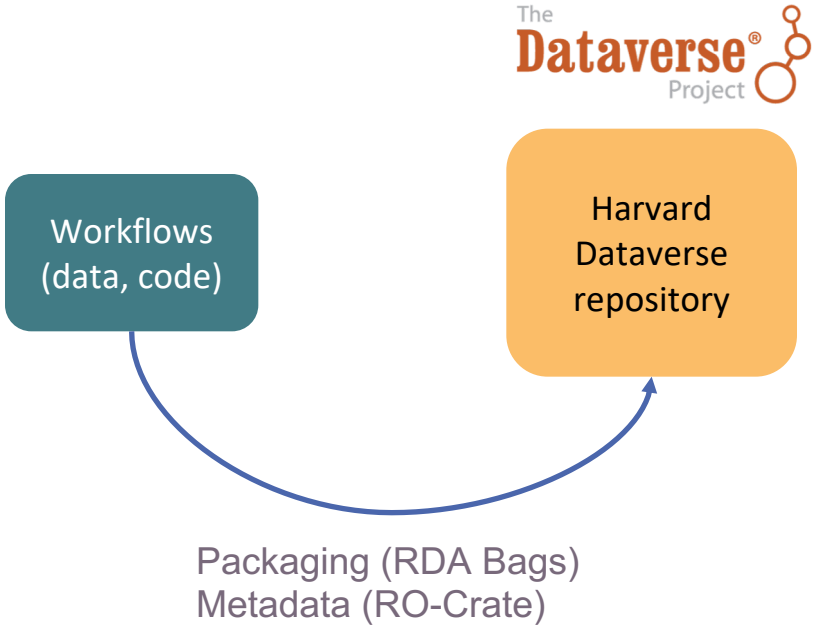
2. Towards a Data Commons to integrate services, computing, and repositories

Integrate Research Computing with Dataverse Repository



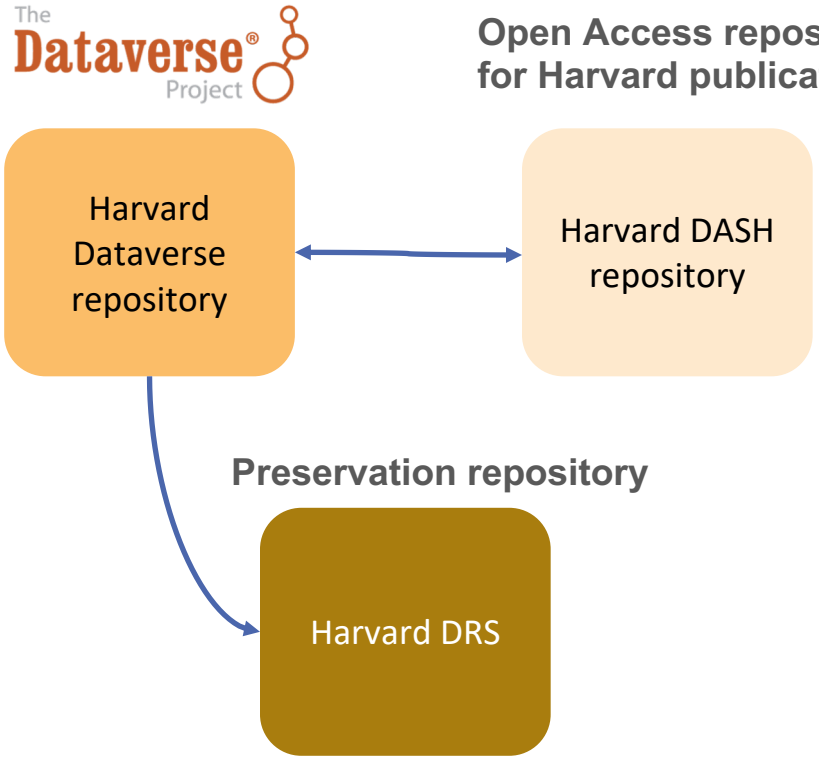
- Facilitate publication of data from research computing storage to Harvard Dataverse repository using **Globus endpoint**
- For **very large datasets** (>TBs), publish metadata to repository and keep data in research computing storage, connected via a persistent link
- Same for **sensitive data** with access controls and trusted secure storage

Support Research Workflows



- Support **advanced computational research workflows** and research objects
- Review **standards for packaging and metadata** (BagIt, RO-Crate)
- **Deposit workflows** to repository to facilitate reproducibility and repeatability

Connect Harvard Research Repositories



- Support **open science** and **long term access and reuse**
- **Connect datasets** with open access **publications**
- Copy curated collections to preservation repository, using **preservation standards**

Thank you!

Mercè Crosas, Ph.D., Harvard University
scholar.harvard.edu/mercecrosas @mercecrosas