The Dataverse Network™ Project

MAKING RESEARCH DATA MORE ACCESSIBLE
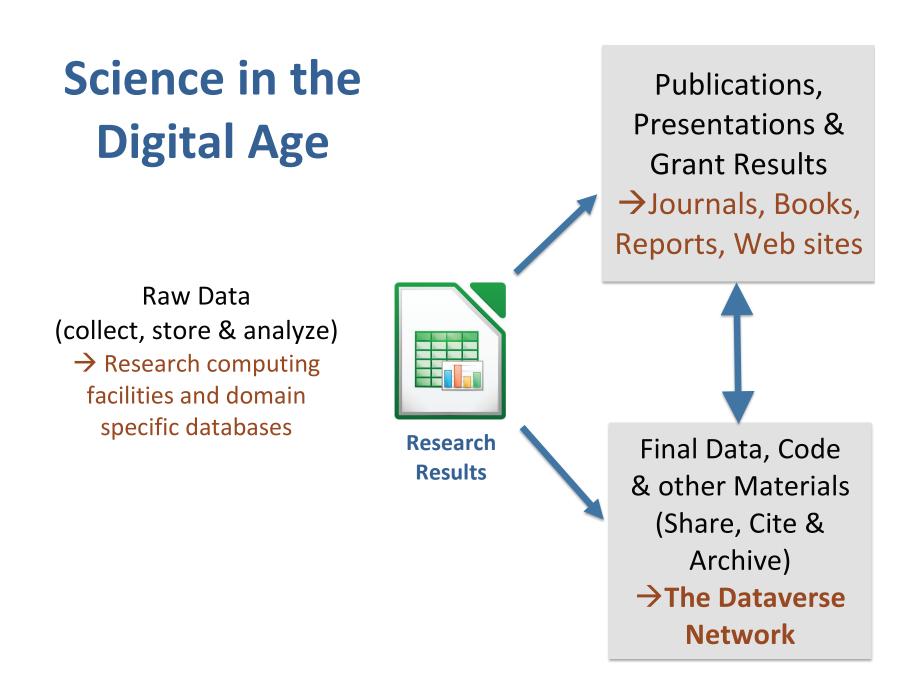
Eleni Castro (Research Coordinator) and Gustavo Durand (Project Manager, Product Development),
the Institute for Quantitative Social Science at Harvard University – http://thedata.org

CfA | **HARVARD LIBRARY** | **IQSS** The Institute for Quantitative Social Science

---

What is the Dataverse Network Project?

Science in the Digital Age

Raw Data
(collect, store & analyze)
→ Research computing facilities and domain specific databases

Research Results

Publications, Presentations & Grant Results
→Journals, Books, Reports, Web sites

Final Data, Code & other Materials (Share, Cite & Archive)
→The Dataverse Network

The Dataverse Network is a web application for:

Citing
Preserving
Analyzing
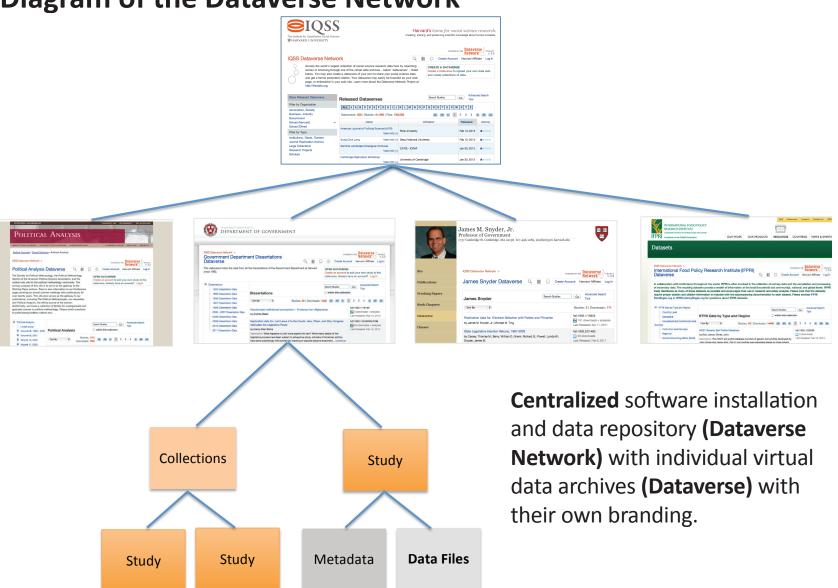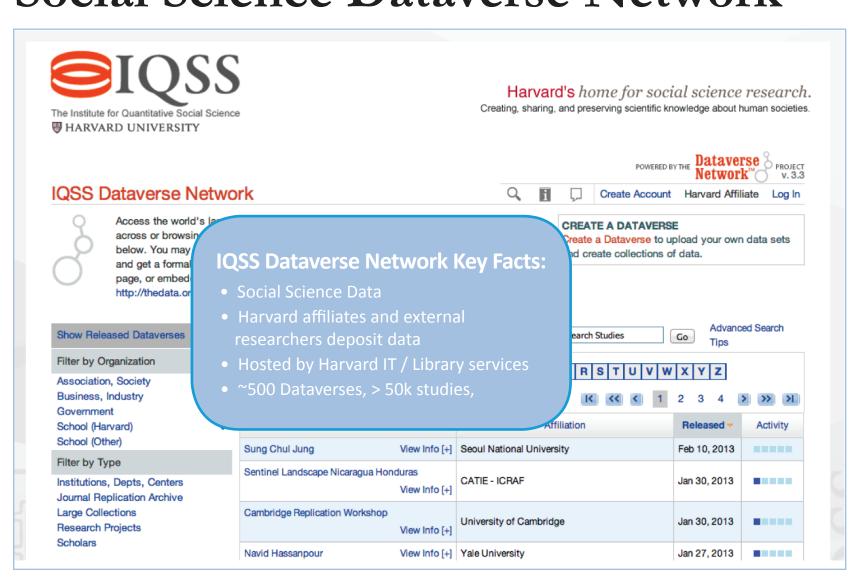**Sharing**

open source initiative

→Research Data

Key Benefits:
- Recognition, credit and control for data author
- Persistent citation between scholarly claims and data
- Central repository for archival and long-term access
- Validation of science through replication
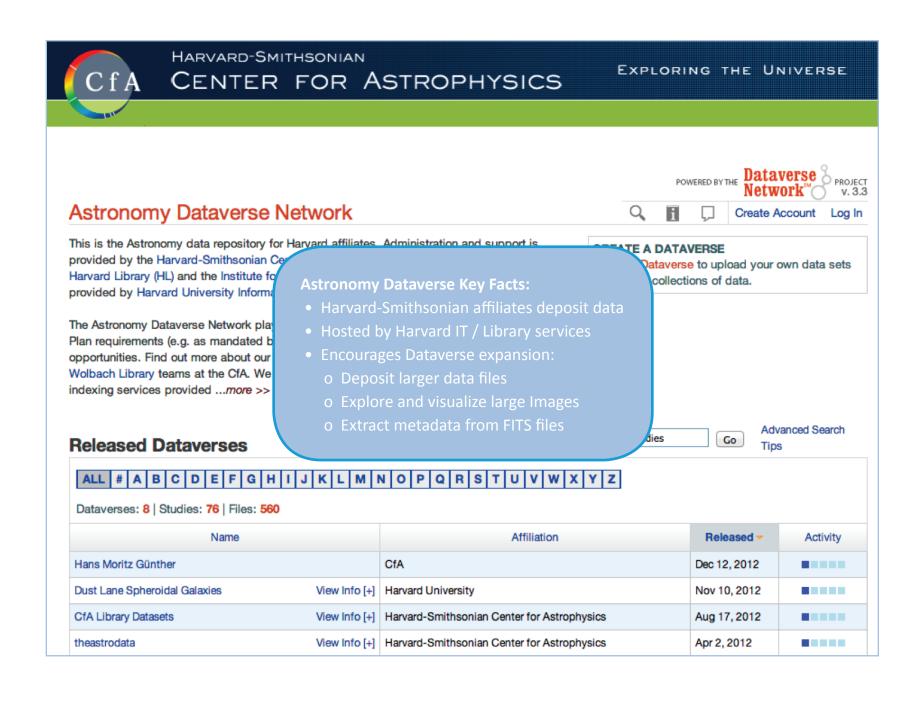
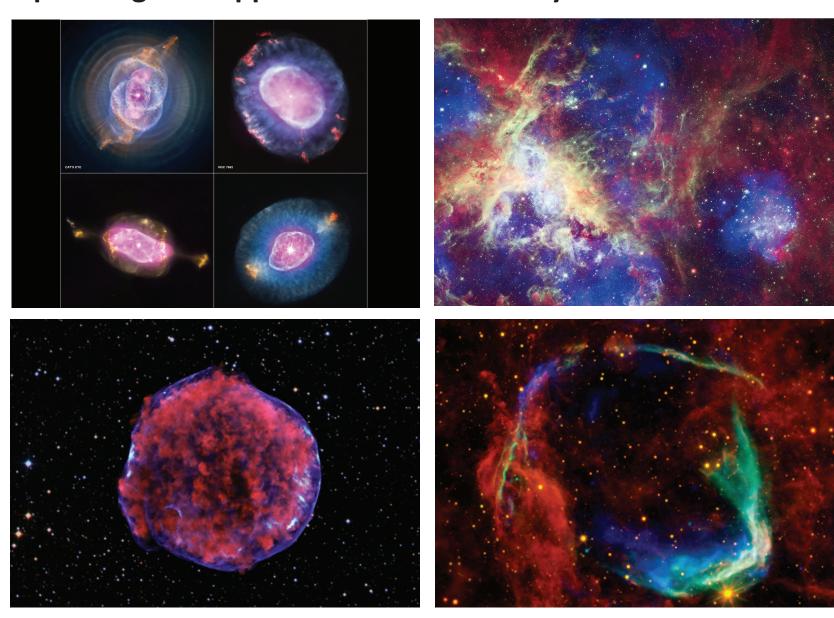### Diagram of the Dataverse Network

Centralized software installation and data repository (**Dataverse Network**) with individual virtual data archives (**Dataverse**) with their own branding.

Collections
Study | Study | Metadata | Data Files

Social Science Dataverse Network

IQSS
The Institute for Quantitative Social Science
HARVARD UNIVERSITY

Harvard's home for social science research.
Creating, sharing, and preserving scientific knowledge about human societies.

IQSS Dataverse Network Key Facts:
- Social Science Data
- Harvard affiliates and external researchers deposit data
- Hosted by Harvard IT / Library services
- ~500 Dataverses, > 50k studies,
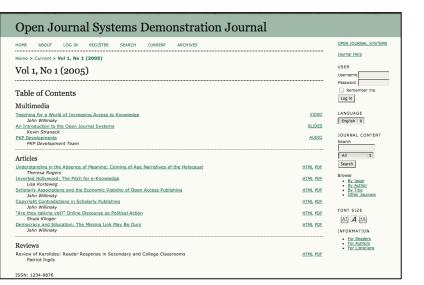
Upcoming DVN integration with Harvard Library systems:
- DASH (Harvard's IR): relate publications to research data
- Digital Repository Service (DRS): as a preservation backend
- DMP (when available)

Astronomy Dataverse Network

Harvard-Smithsonian Center for Astrophysics (CfA) Astronomy Dataverse Network

Astronomy Dataverse Key Facts:
- Harvard-Smithsonian affiliates deposit data
- Hosted by Harvard IT / Library services
- Encourages Dataverse expansion:
  - o Deposit larger data files
  - o Explore and visualize large images
  - o Extract metadata from FITS files

Upcoming full support for FITS astronomy data files

Harvard Library meets the Dataverse Networks

HARVARD LIBRARY

With Library Technology Services (LTS)

Hosted by Harvard Information Technology Services

Migration to centralized hosting of Dataverse Networks, in order to efficiently support multiple domains.

Collaboration Key Facts:
- The Library/DVN project is the main solution offered at Harvard for publishing, sharing and citing research data.
- The Digital Scholarship Program at the Harvard Library, together with IQSS and research computing at Harvard, is developing other strategic projects aligned with the program's goal to enable researchers with the necessary tools to manage research data. These projects will include:
  - Providing joint resources on the research data management cycle.
  - Continuing education of librarians and researchers in data management.
  - Joint advisory services to researchers about: storage, archiving, preservation and curation of research data.

OJS-DVN Project: Developing a Data Sharing System for Journals

Funded by an Alfred P. Sloan Foundation grant, the Dataverse Network Project in collaboration with the Public Knowledge Project (PKP) are on a two-year endeavor to make data sharing and preservation an intrinsic part of the publication process.
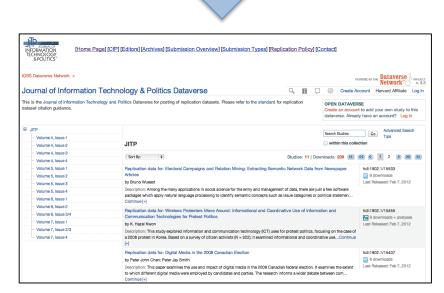
PKP PUBLIC KNOWLEDGE PROJECT

Author submits article + research data to a journal using PKP's Open Journal System (OJS).

Journal Management And Publishing System

Research data + metadata will automatically be deposited into a journal's Dataverse repository using API technology (SWORD). In conjunction, PKP is developing a plugin to allow for this seamless integration.

Journal's Research Data Repository

Projected impacts the OJS-DVN project

"The immediate impact of the project will be to increase the number of readily replicable articles published, and the number of social science journals that adopt best data management and citation practices. The broadest impacts of the project will be to increase the pace of discovery in the social sciences, and to broaden the research opportunities for younger scholars."
- Gary King, IQSS Director and PI of Project

"This project will help establish a formal linkage between scholarly publications and the underlying research data, thus increasing replicability and reusability of previous studies to verify and advance science."
- Dr. Mercè Crosas, IQSS Director of Product Development, and Co-PI of project.

Up next for The Dataverse Network

- Expansion of the IQSS DVN to Harvard's DVN to support other domains (medical, health, etc).
- DOI support, in addition to the current Handle support for persistent identifiers.
- Integration with Shibboleth for user authentication.
- SWORD 2 compliant API to deposit research data seamlessly from Journals.
- Full support for:
  - FITS astronomy data files
  - R files
- Integration with WorldMap to visualize geospatial data.
- Support for archiving and sharing sensitive data.
- New G UI for statistical analysis and subsetting of quantitative data sets, using Zelig statistical package.