Using Dataverse Virtual Archive Technology for Research Data Management

Jonathan Crabtree
Thu-Mai Christian
Amanda Gooch

UNC | THE ODUM INSTITUTE

H. W. Odum Institute Archive Services

- The Howard W. Odum Institute was founded in 1924.

- It is the oldest multidisciplinary social science university institute.

- Odum Archive Services is host to one of the largest catalogs of machine-readable data in the U.S.

- The Odum Dataverse Network (DVN) catalog includes polling, census, and other social science and health-related data.

The Problem

- Different needs for archives, data libraries, researchers, journals, funding agencies…
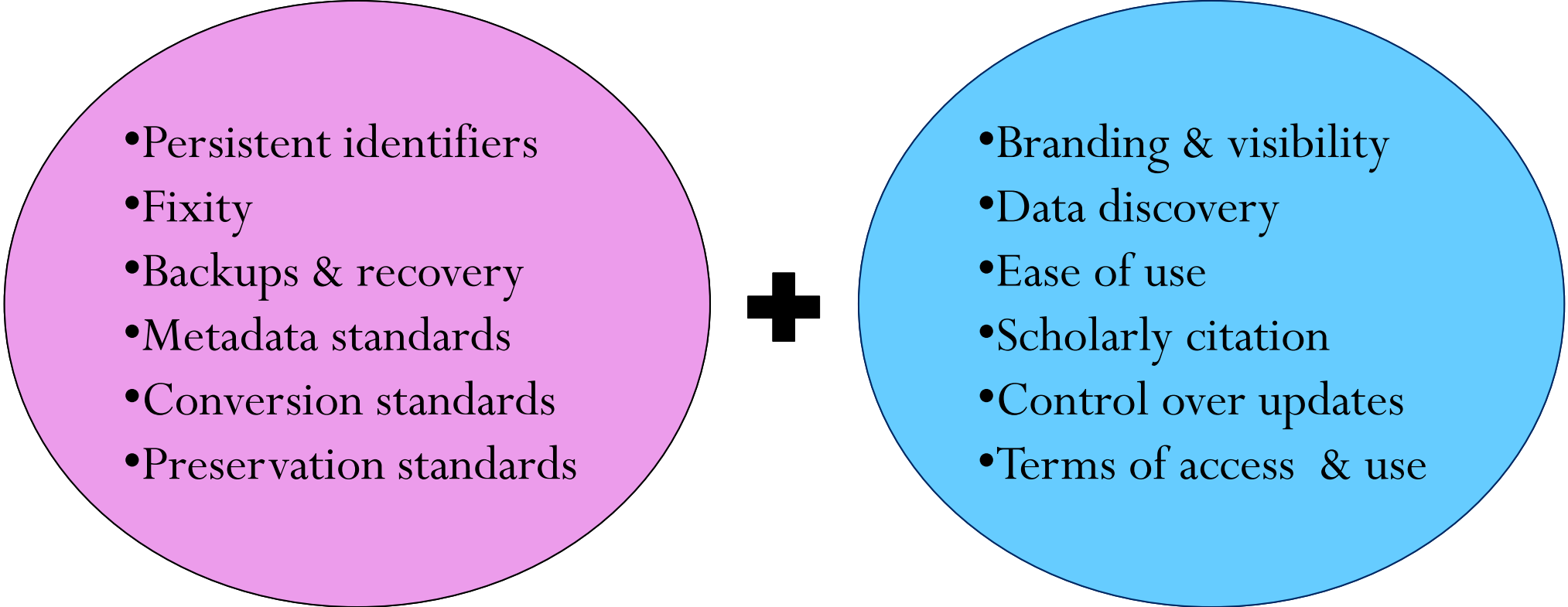

Cross, M. *Why the Dataverse Network?*
Available at: thedata.org

Odum Institute's Solution

- Dataverse Network: **centralized** professional archiving with **distributed** control and recognition

- Persistent identifiers
- Fixity
- Backups & recovery
- Metadata standards
- Conversion standards
- Preservation standards

+

- Branding & visibility
- Data discovery
- Ease of use
- Scholarly citation
- Control over updates
- Terms of access & use

Cross, M. *Why the Dataverse Network?* Available at: thedata.org

How It Works

Cross, M. *Why the Dataverse Network?* Available at: thedata.org
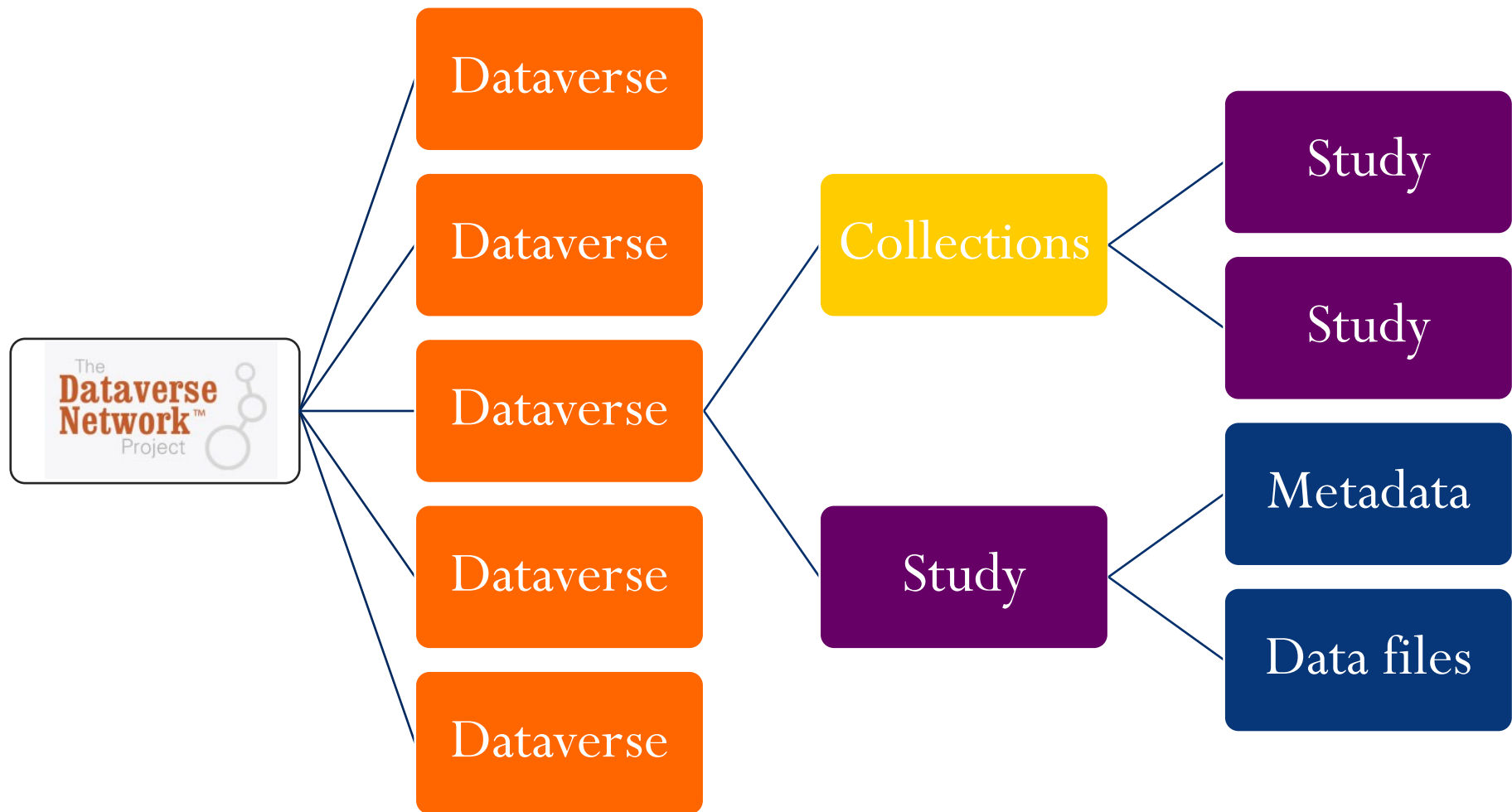
Supporting Data

Tabular Data:

STATA

SPSS

R

CSV + control card

Tab delimited + DDI

Social Network Data:

GraphML

Other data or relevant files:

All formats are accepted BUT only tabular files have value added data support

- Convert to a preservation format (data and metadata)
- Calculate Universal Numerical Fingerprint (UNF)
- Download in multiple formats
- Download a subset of the data
- Generate summary statistics
- Apply Zelig (R) statistical methods
- Visualize time series
- Define Terms of Use and Permission

Cross, M. *Why the Dataverse Network?* Available at: thedata.org
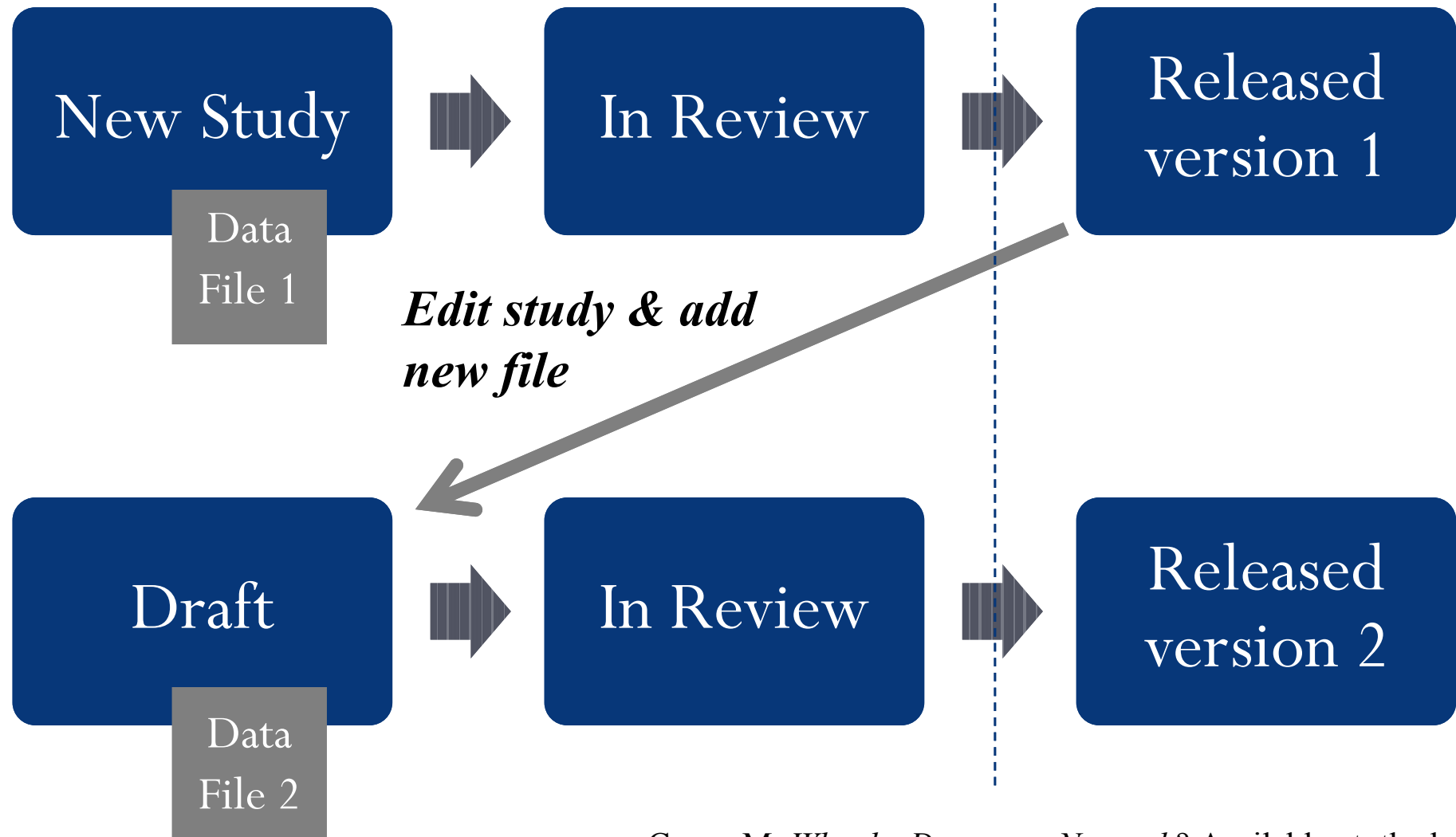
Creating Data Citations

- Author(s)

- Year

- Title

- Persistent URL and ID (DOI coming soon)

- UNF

- Distributor

- Version

- Other optional fields

Louis Harris and Associates, Inc., 1992, "Harris 1984 Female Veterans Survey, study no. 843002", http://hdl.handle.net/1902.29/H-843002 UNF:3:4VngKZgBorG/7T6aZSaq1g== Odum Institute;Odum Institute for Research in Social Science [Distributor] V1 [Version]

Cross, M. *Why the Dataverse Network?* Available at: thedata.org

Managing Data and Versions

Contributor, curator, admin view   **End user view**

New Study → In Review → Released version 1

Data File 1

Edit study & add new file

Draft → In Review → Released version 2

Data File 2

Cross, M. *Why the Dataverse Network?* Available at: thedata.org

Data Never Permanently Deleted

A study is never permanently deleted after it is released. Curators or admins can **deaccession** the study.

Released version N

De-accessioned

This study is deaccessioned. [Go to other study]

Edit study

Draft → In review → Released version N & I

Cross, M. *Why the Dataverse Network?* Available at: thedata.org
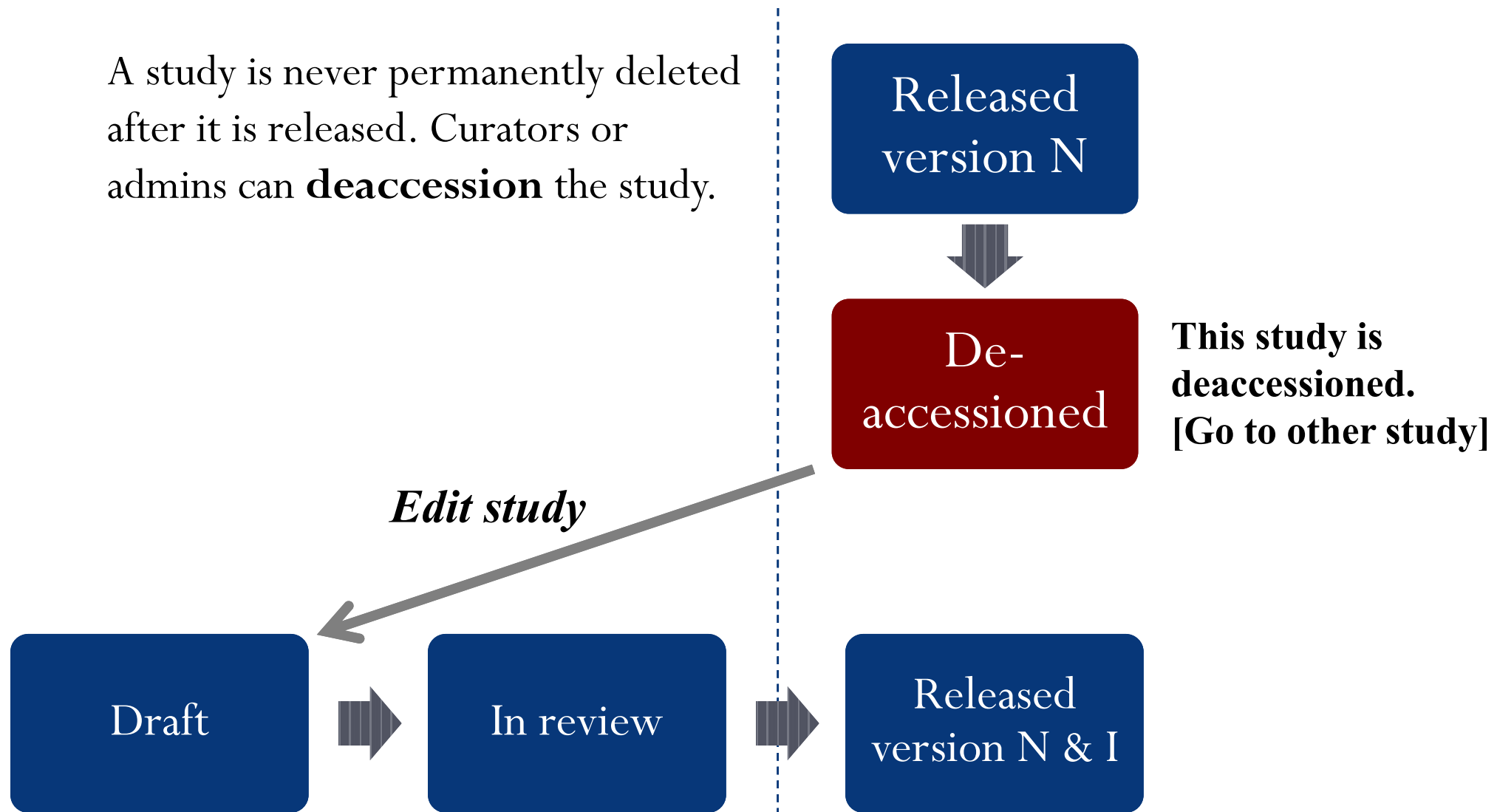
Supporting Standards

- Study and variable metadata are exported into **XML** (Dublin Core, Data Documentation Initiative – DDI, FGDC) and **MARC**

- **OAI-PMH** for harvesting metadata

- **LOCKSS** for data duplication in multiple locations

- **Z39.50** for distributed search

- **E-Z Proxy** to authenticate for data access

Moving Beyond Social Science

- Dataverse Network is cross-disciplinary
- We are expanding the study metadata and building communities of interested groups:
  - dataverse-community@googlegroups.com

Questions?

- Jonathan Crabtree, Asst. Director for Archives & IT
  - Phone: (919) 962-0517
  - Email: Jonathan_Crabtree@unc.edu

- Thu-Mai Christian, Archival Research Associate
  - Phone: (919) 962-6293
  - Email: tlchristian@unc.edu

- Email: odumarchive@unc.edu
- http://arc.irss.unc.edu/dvn/