

How does the ISPS
Data Archive
re-use data?

A key data curation task is appraisal and selection, with re-appraisal after initial selection. (DCC)

A well-curated archive ensures that, “data are accessible to designated users for first time use and reuse.” (DCC)

We argue that, in a replication archive, a key criterion for re-appraisal is whether the data and code reproduce the published results.

So, in addition to traditional curatorial tasks, dedicated data curation staff replicate analyses and validate published results for each study before publishing the files online.

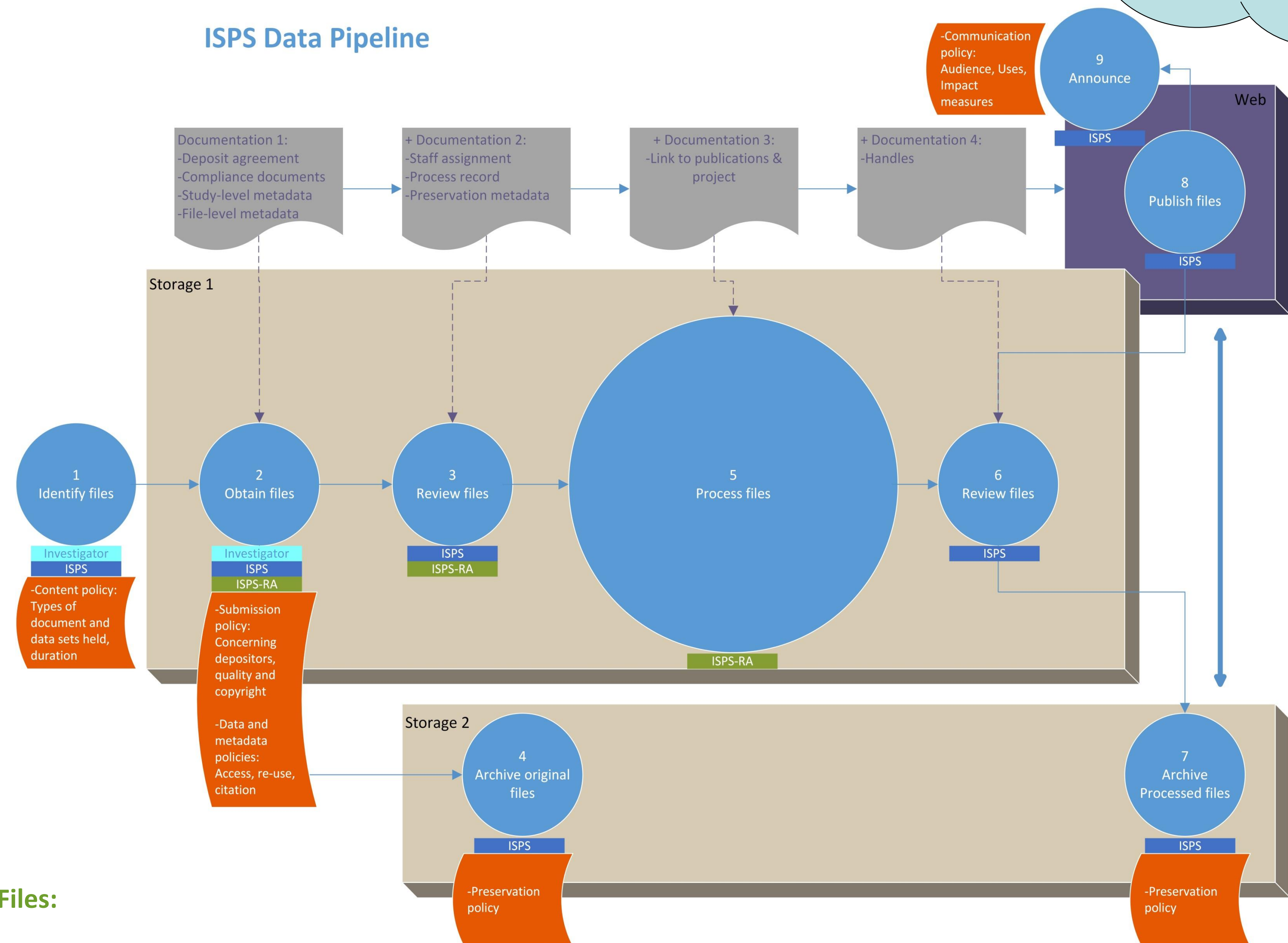
In practice, this has implications for: Resources, Expertise, and Relationships.

The Repository as Data (Re) User: Hand Curating for Replication

Limor Peer, PhD

Yale University. Institution for Social and Policy Studies

ISPS Data Pipeline



How does
replication drive
curation at the ISPS
Data Archive?

Process Files:

1. Assign staff to study and files
2. Move original files to Archive space
3. Make copies of processed files and move to collaborative space
4. Identify related publication and project
5. Rename all copied files for public dissemination according to ISPS Data Archive naming conventions
6. Check and complete variable-level metadata for each data file
7. Compare variable information, check for additional variables and recoded variables, check variable/value labels
8. Check all files for confidential and other sensitive information
9. Run the statistical code and check against published results
10. Re-write statistical code in R and check replication
11. Communicate with PI as needed
12. Create new DDI-XML file with variable-level information
13. Create additional files by converting to readable formats (e.g., ASCII, PDF)
14. Update study- and file-level metadata record
15. Update tracking documents: process record / general study database / status document