

# An Open Source, DDI-based Curation System for Social Science Data

Limor Peer, Ann Green, Niall Keleher, Jeremy Iverson, Dan Smith  
[limor.peer@yale.edu](mailto:limor.peer@yale.edu), [green.ann@gmail.com](mailto:green.ann@gmail.com), [nkeleher@poverty-action.org](mailto:nkeleher@poverty-action.org), [jeremy@colectica.com](mailto:jeremy@colectica.com), [dan@colectica.com](mailto:dan@colectica.com)

### About the project

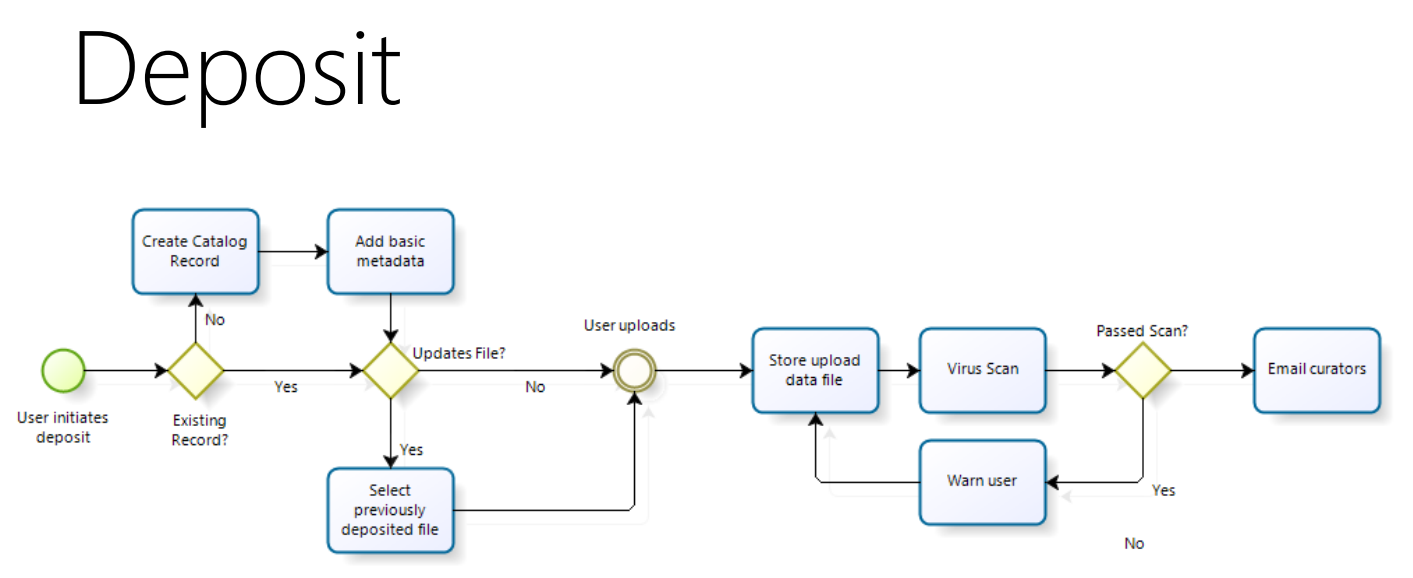
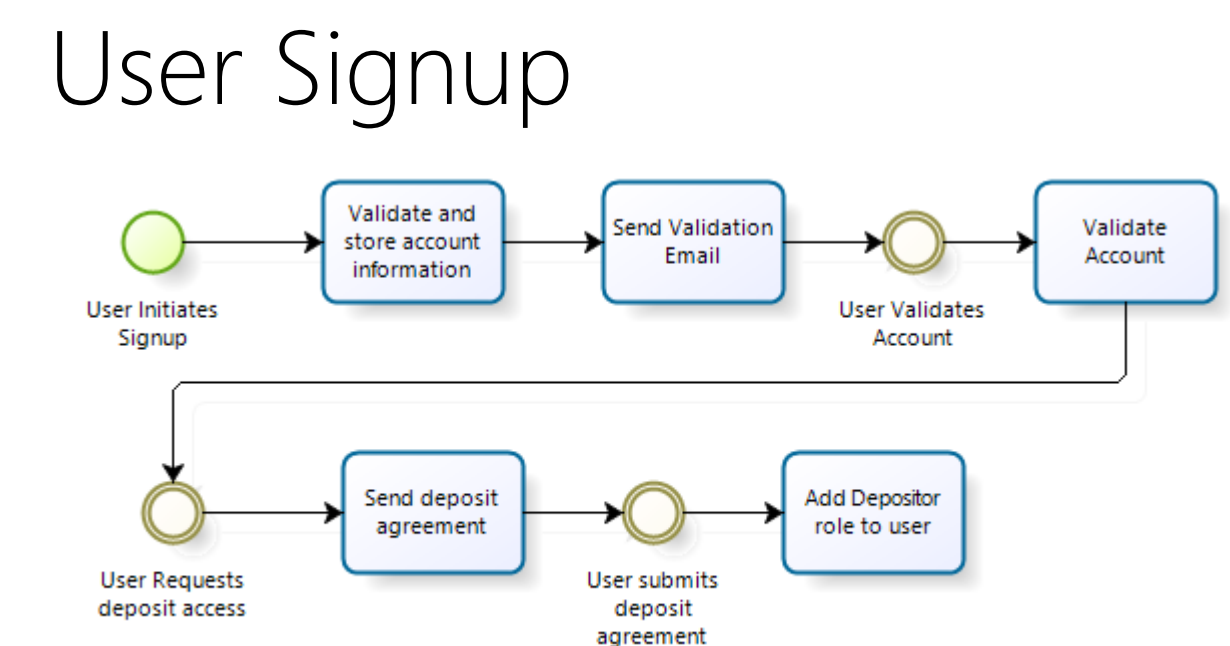
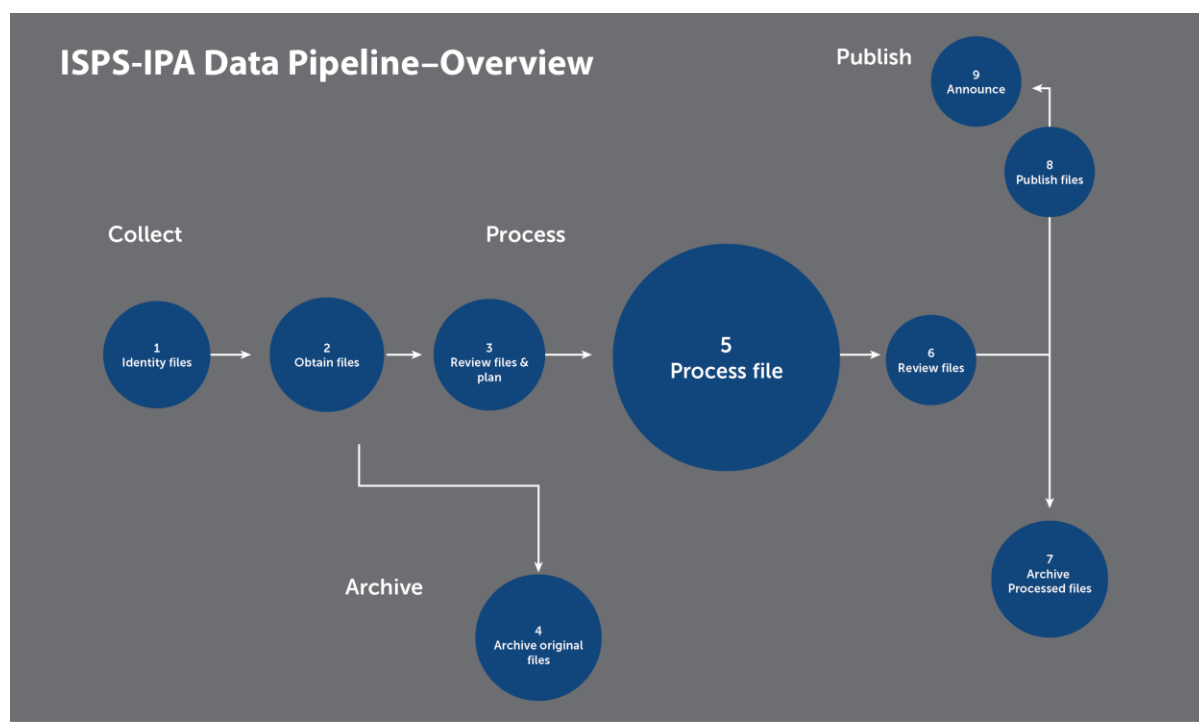
The Institution for Social and Policy Studies (ISPS) at Yale University and Innovations for Poverty Action (IPA) are partnering with Colectica to develop a software platform that structures the curation workflow, including checking data for confidentiality and completeness, creating preservation formats, and reviewing and verifying code.

Both organizations collect data from social science research that measures the impact of interventions—such as voter mobilization campaigns and microfinance programs—via randomized controlled trials in the real world.

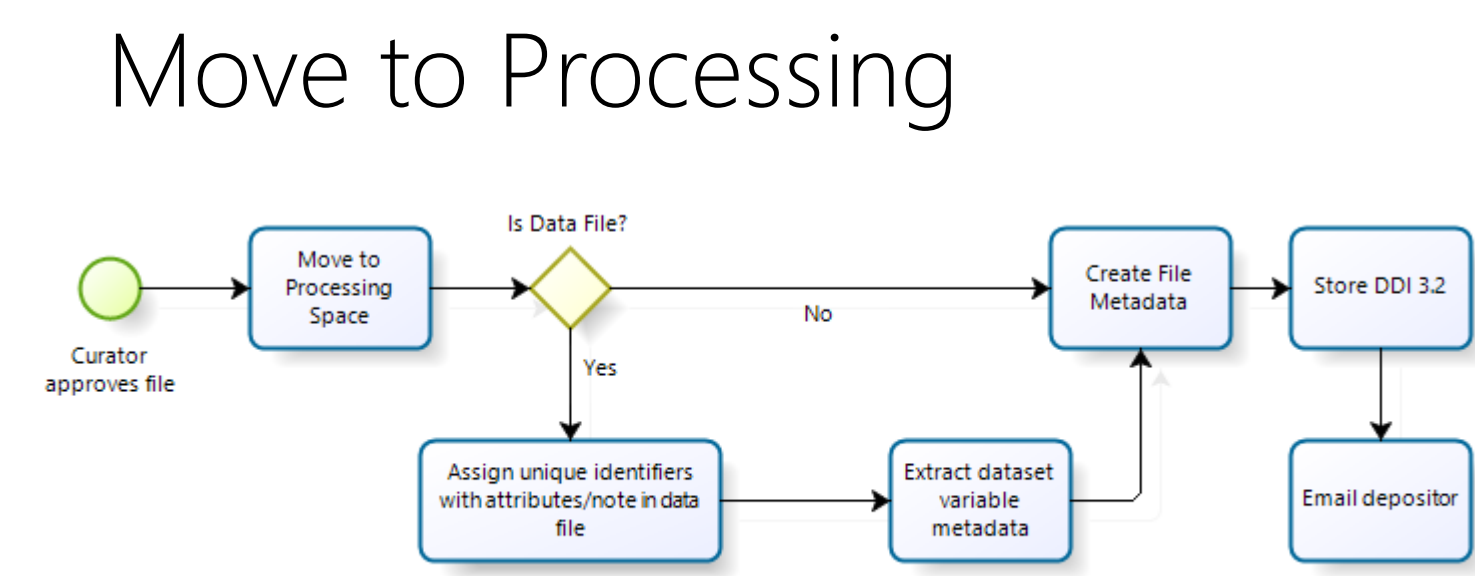
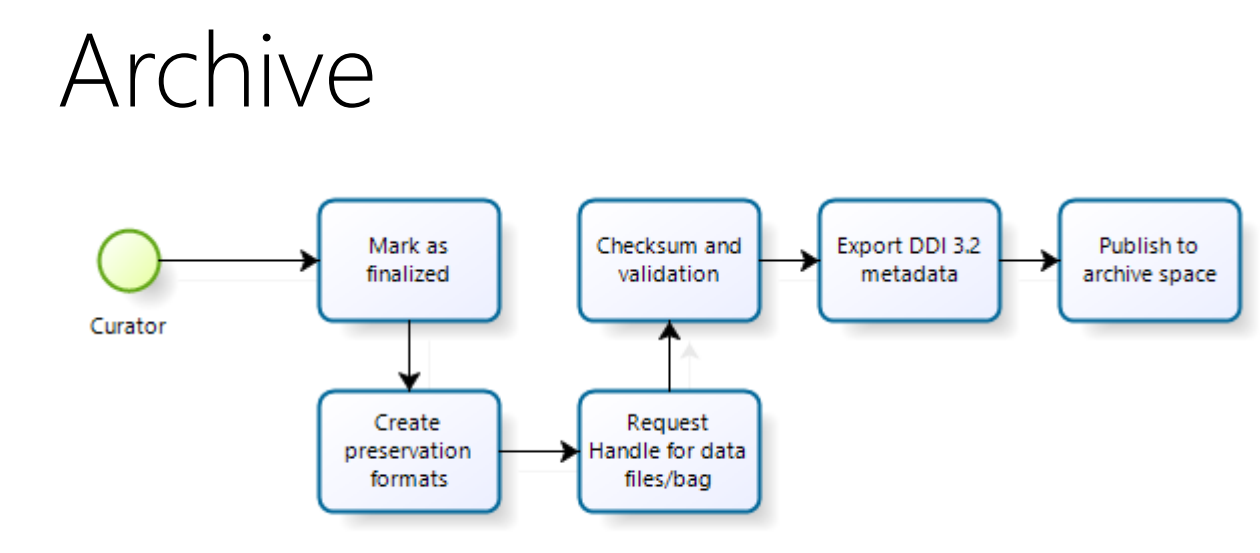
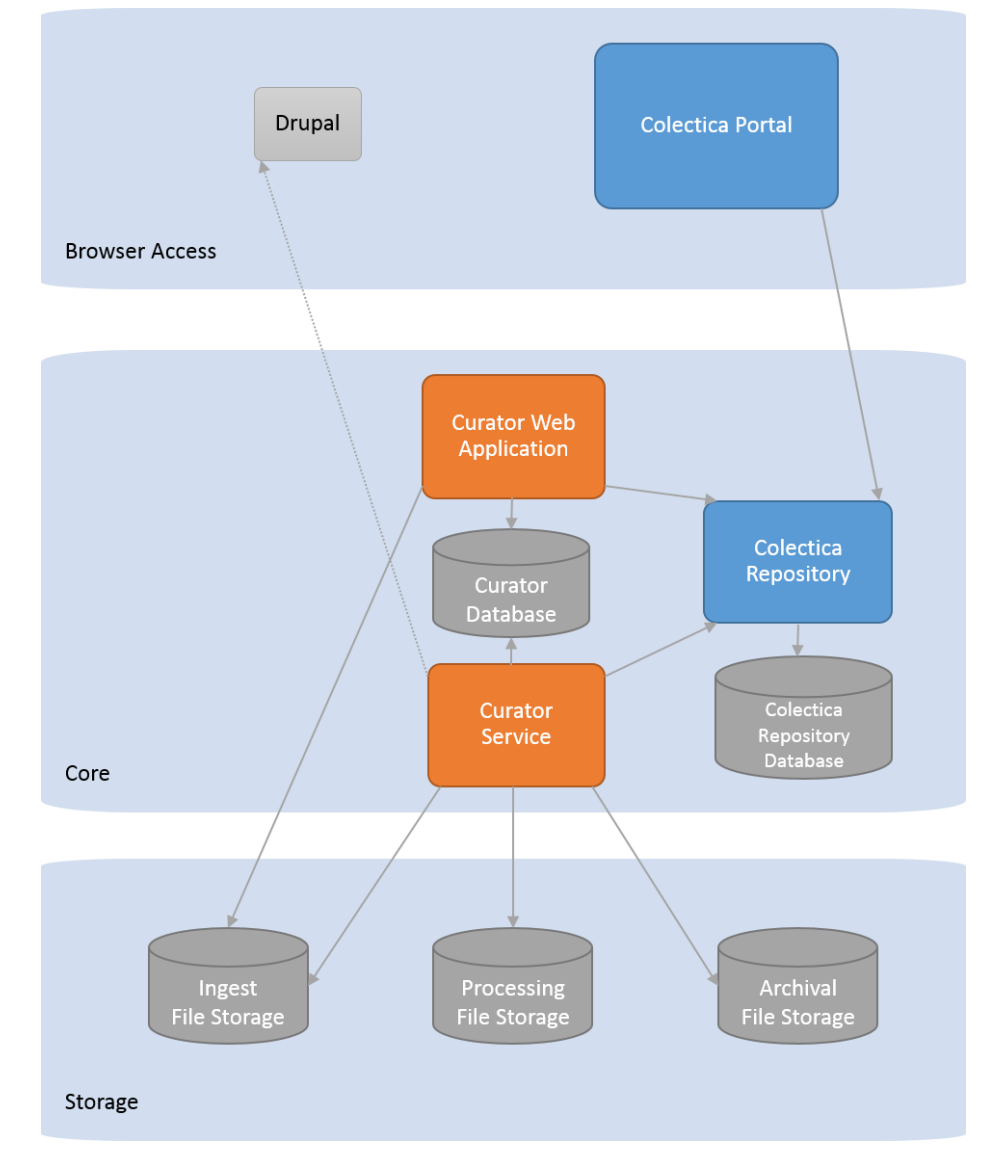
The software leverages DDI Lifecycle and combines several off-the-shelf components with a new, open source, Web application.

### ISPS and IPA requirements

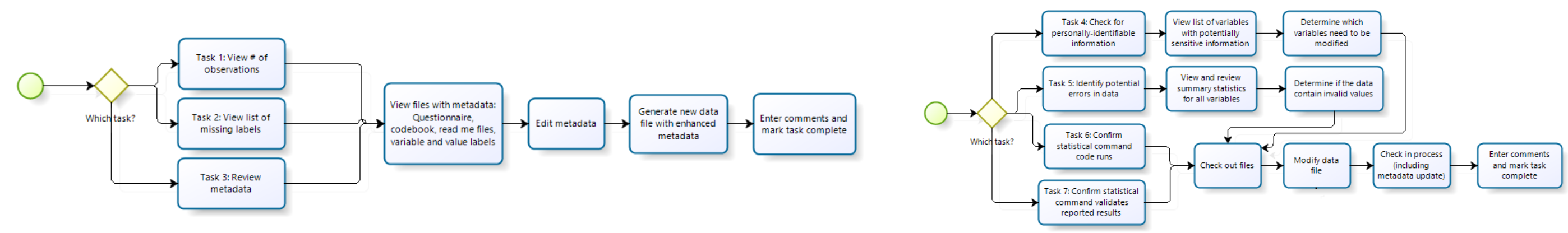
- Curation workflow management (dashboard)
- Track changes to files (provenance)
- Integrate metadata production with data and code review and cleaning
- Preservation metadata and formats
- Secure storage and access
- Smooth transition to public dissemination of content
- Preference for open source solutions



## System Architecture



## Processing, Review, and Metadata Enhancement Tasks



The Institution for Social and Policy Studies (ISPS) at Yale University advances interdisciplinary research in the social sciences that aims to shape public policy and inform democratic deliberation. The ISPS Data Archive has a replication archive (ISPS Data Archive), which pioneered the review of all data and code pre-publication.



Innovations for Poverty Action (IPA) is a 501(c)(3) US-based non-profit organization committed to performing high-quality randomized evaluations throughout the developing world. IPA's research involves widely distributed data gathering efforts.



Colectica is a software company specializing in providing integration and access to data using open data standards.

## Data Quality Review

Files

Documentation

Data

Code

Before

After