ICPSR

SHARING DATA TO ADVANCE SCIENCE

LinkageLibrary at ICPSR

Building a benchmark record linkage repository

Susan Hautaniemi Leonard
Abay Israel
Margaret Levenstein
Trent Alexander

In collaboration with Texas A&M University
Funded by the National Science Foundation

ICPSR INSTITUTE FOR SOCIAL RESEARCH
UNIVERSITY OF MICHIGAN

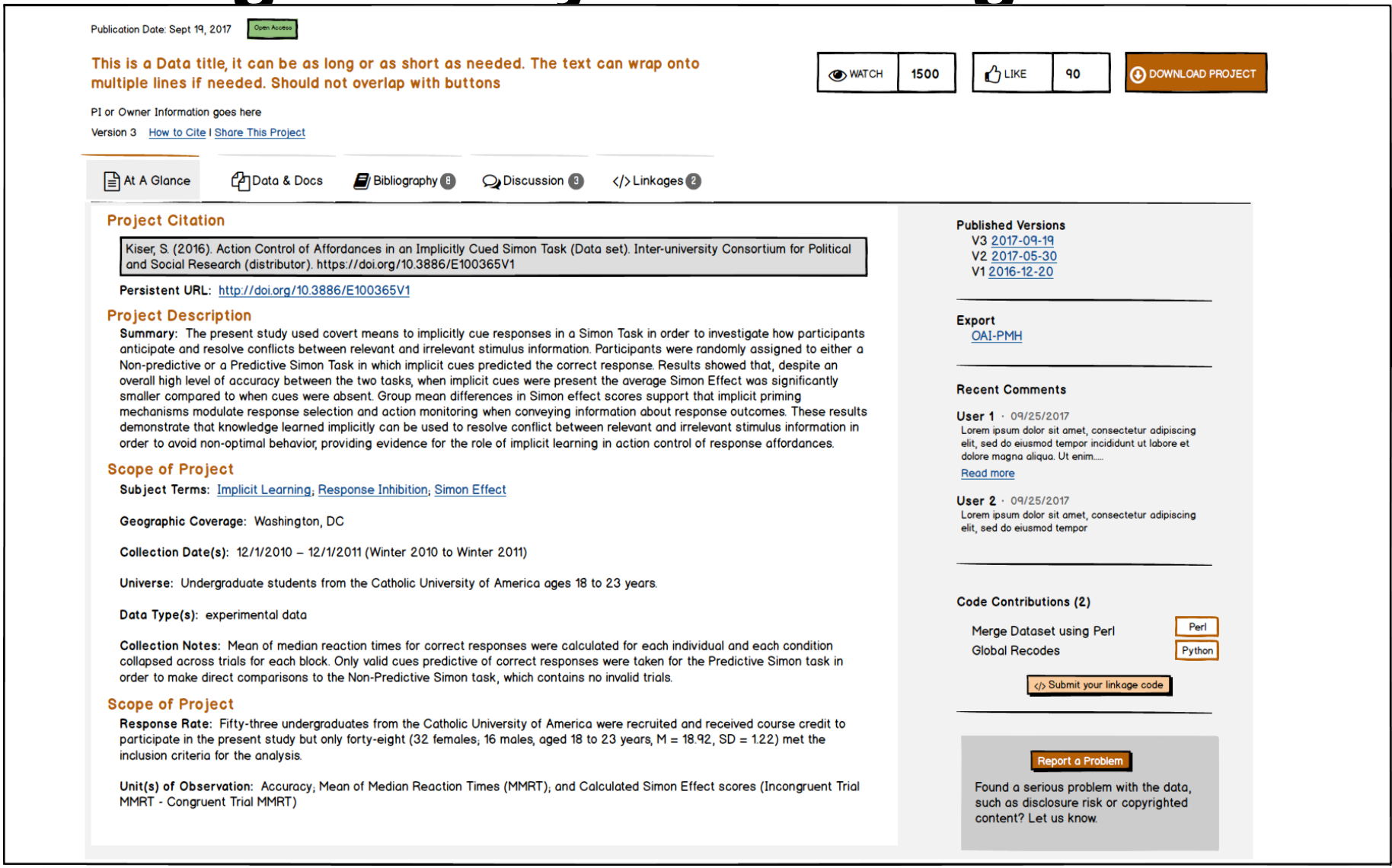
Our goals:

- Accelerate development of new record linkage and evaluation methods, and use on real data
- Improve reproducibility of analyses
- Develop critical collaborations between researchers, users, and data custodians
- Help close the gap between research and practice
- Train the next generation of multi-disciplinary data scientists who can lead the field.
- Build cross-disciplinary community around data linkage

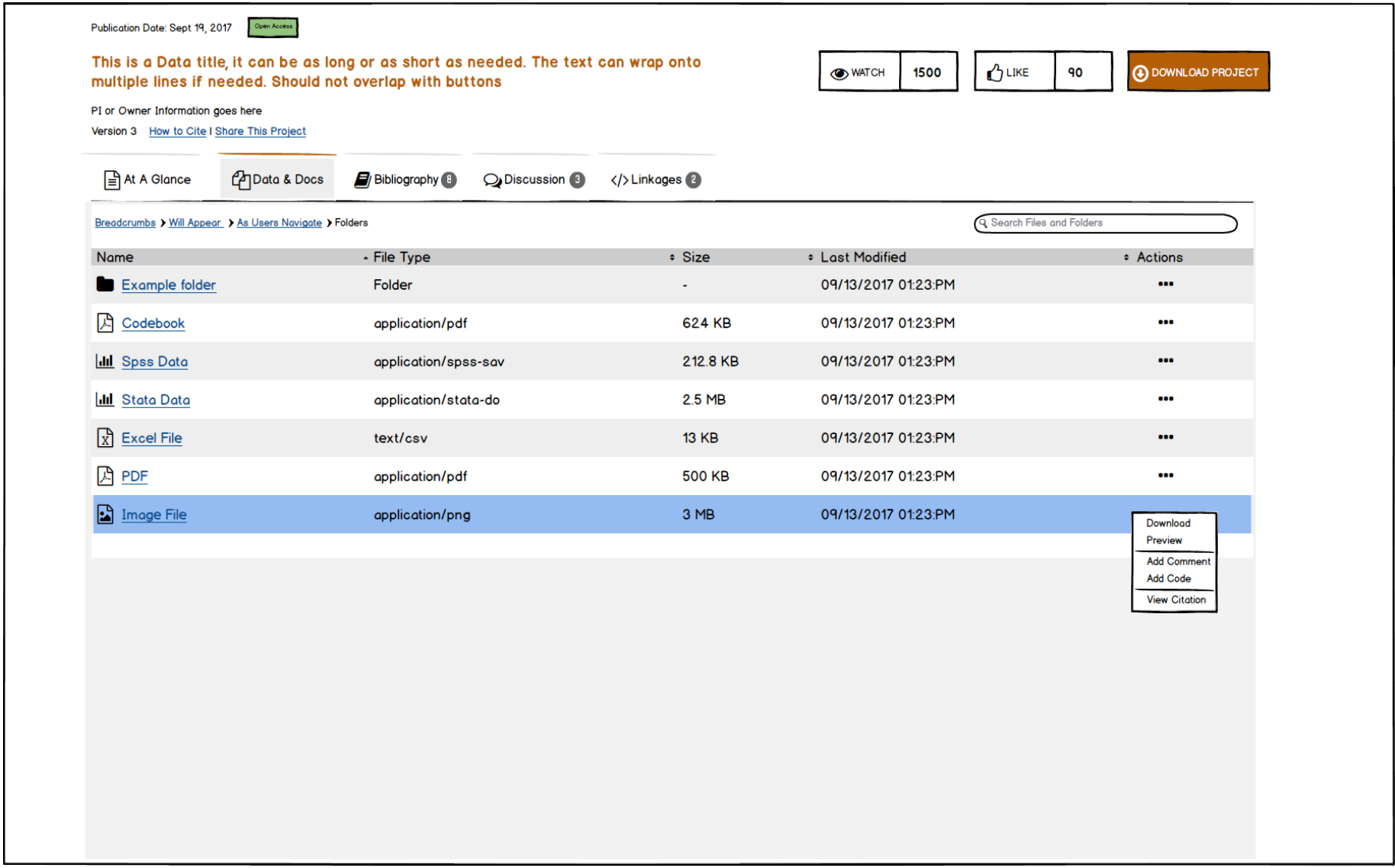
Our approach:

- Build a repository that is based on openICPSR
- Seed it with a variety of exciting projects
- Invite the research community to:
  - Visit the site
  - Download and use code and data
  - Create projects
  - Comment on existing projects
  - Contribute to existing projects

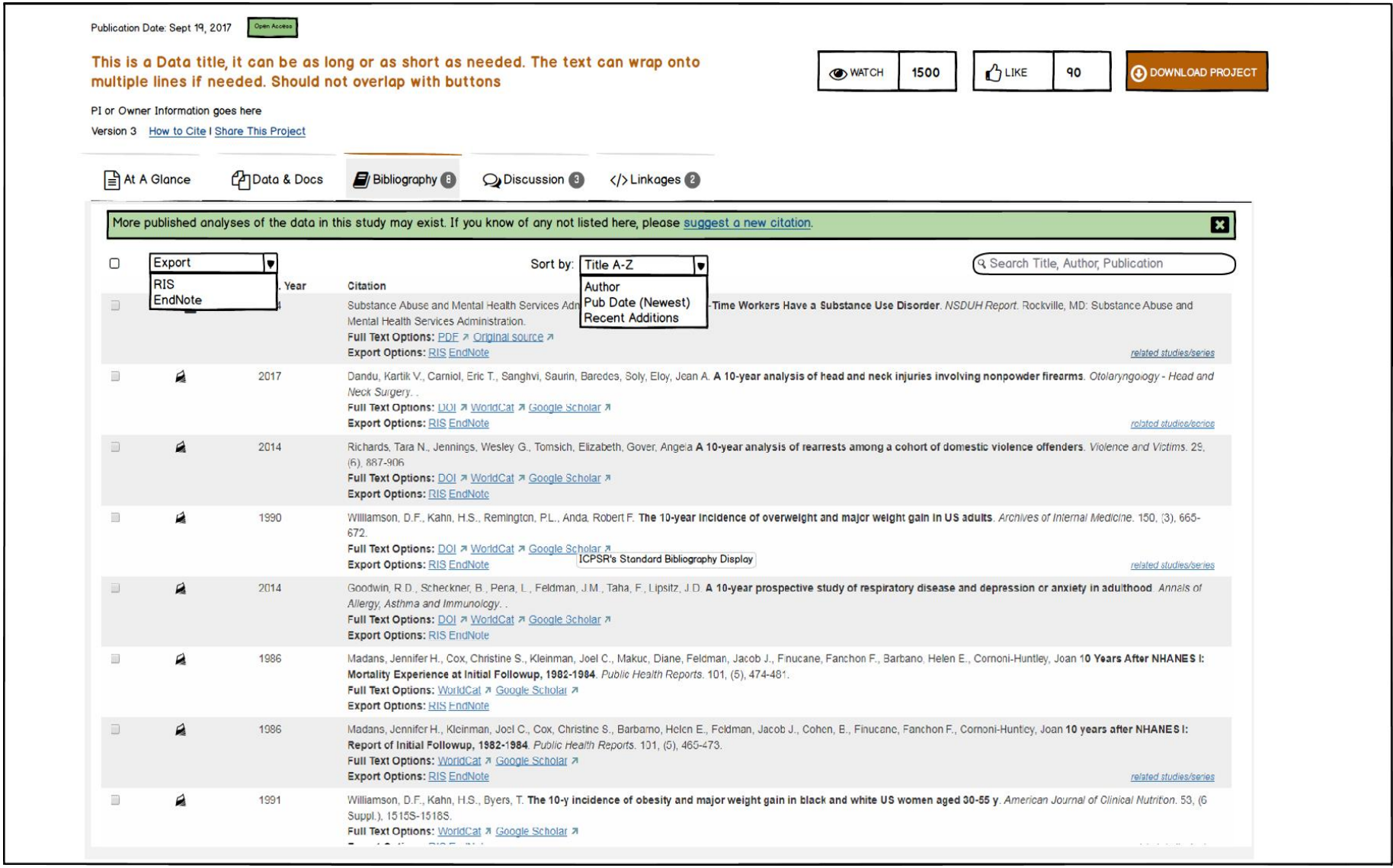
LinkageLibrary Home Page



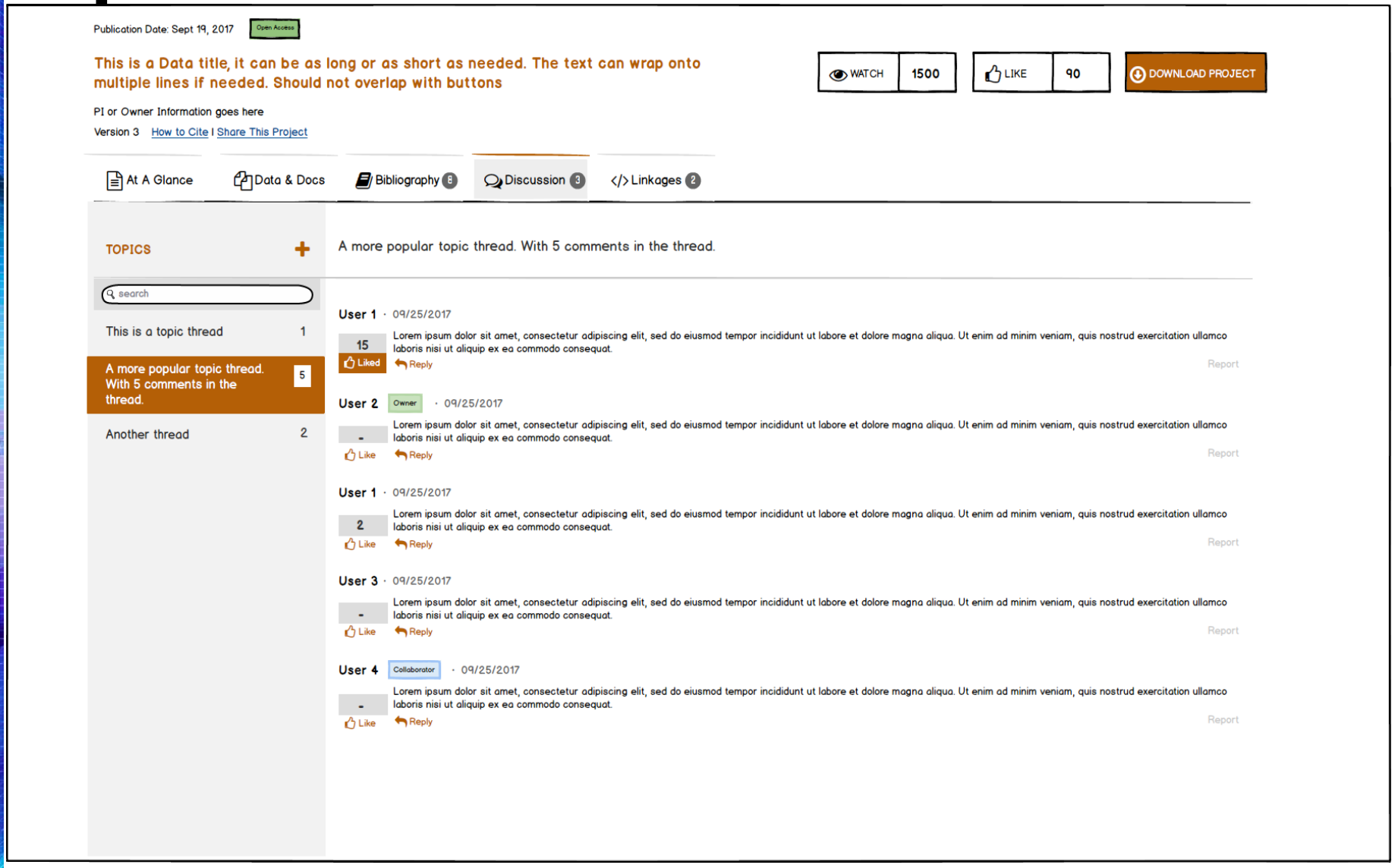
LinkageLibrary File Space



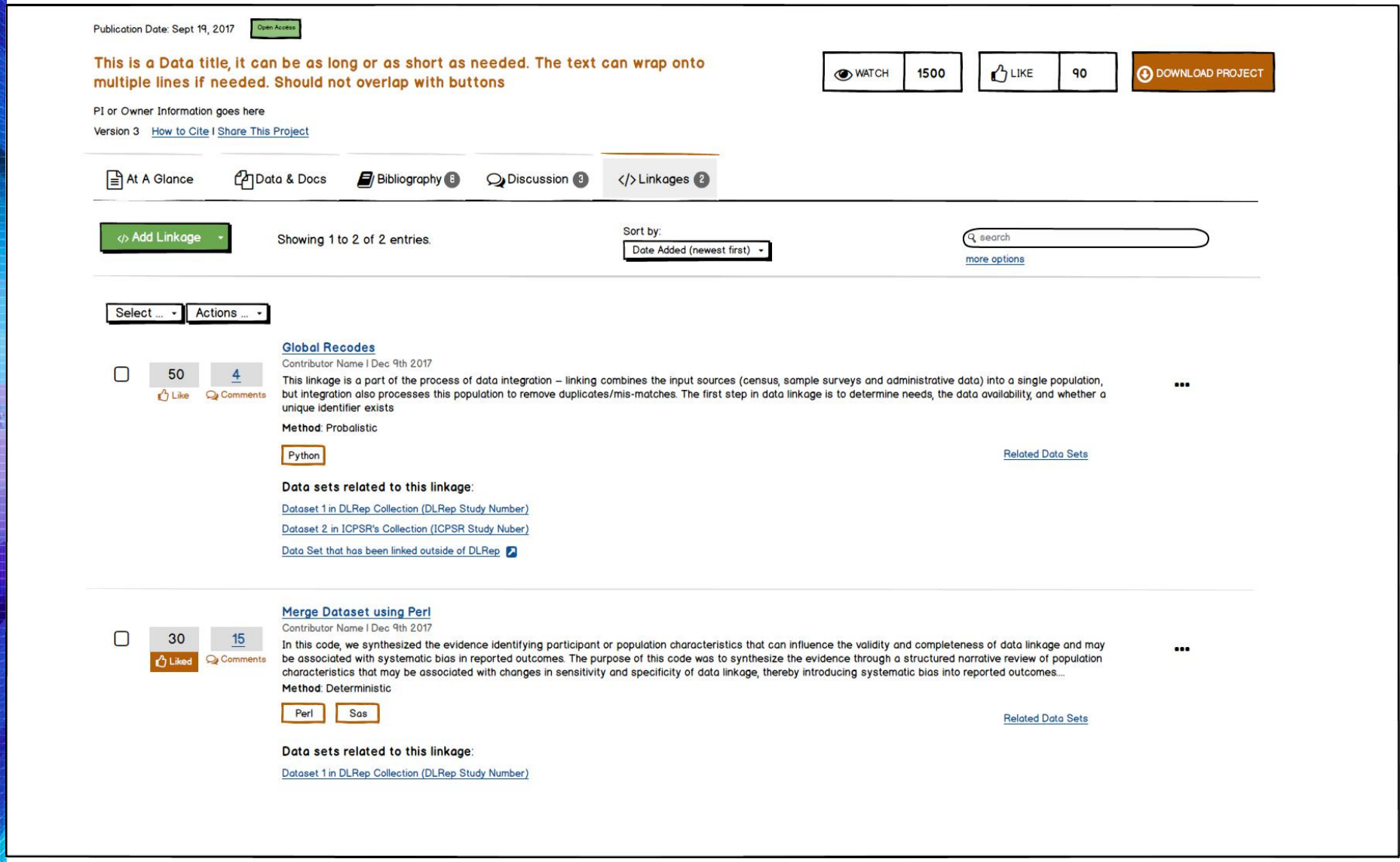
LinkageLibrary Bibliography Space



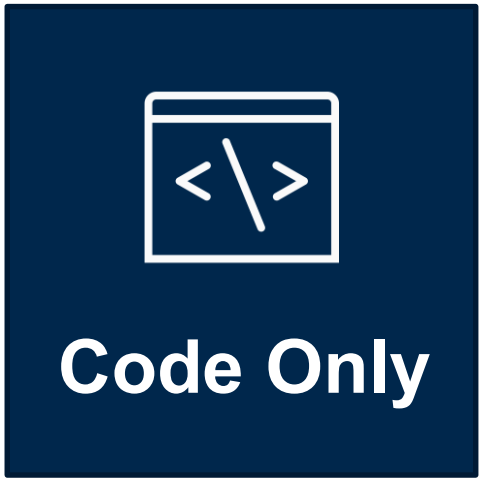
LinkageLibrary Discussion Space



LinkageLibrary Linkages Space



Three types of projects



Seeding the LinkageLibrary

- Bailey et al. **LIFE-M** project
- Levenstein et al. **CensHRS** project
- Jason Owen-Smith et al. **IRIS** project
- Antonie et al. **Sunshine List**
- Access to **CARRA** Census Bureau benchmark files

Join us!

- Lead and learn in a record linkage community
- Full metadata for linked data and methodology
- LinkageLibrary is looking for contributors and users
- Soft launch expected by year's end
- Contact Susan Leonard at ICPSR for more information (hautanie@umich.edu)