SEAD: Preserving Data for Environmental Sciences in Areas of Climate, Land-Use, and Environmental Management

SEAD
Sustainable Environment — Actionable Data
**http://sead-data.net**

Margaret Hedstrom (PI), Marietta Van Buhler (PM), George Alter (ICPSR), Bryan Beecher (ICPSR), Aditya Bhandari, Karen Woollams, Jude Yew - **University of Michigan**

Beth Plale (PI), Katy Börner, Robert H. McDonald, Kavitha Chandrasekar, Inna Kouper, Stacy Konkiel, Robert Light, Robert Ping - **Indiana University**

Praveen Kumar (PI), Md (Zaman) Aktaruzzaman, Rob Kooper (NCSA), Luigi Marini (NCSA), Terry McLaren (NCSA) - **University of Illinois-Champaign-Urbana**

James Myers (PI), Ram Krishnan, R. Lindsay Todd - **Rensselaer Polytechnic Institute**

Follow us on twitter @SEADdatanet or hashtag our poster #IDCC13 #SEADdatanet

1. Problem/Domain
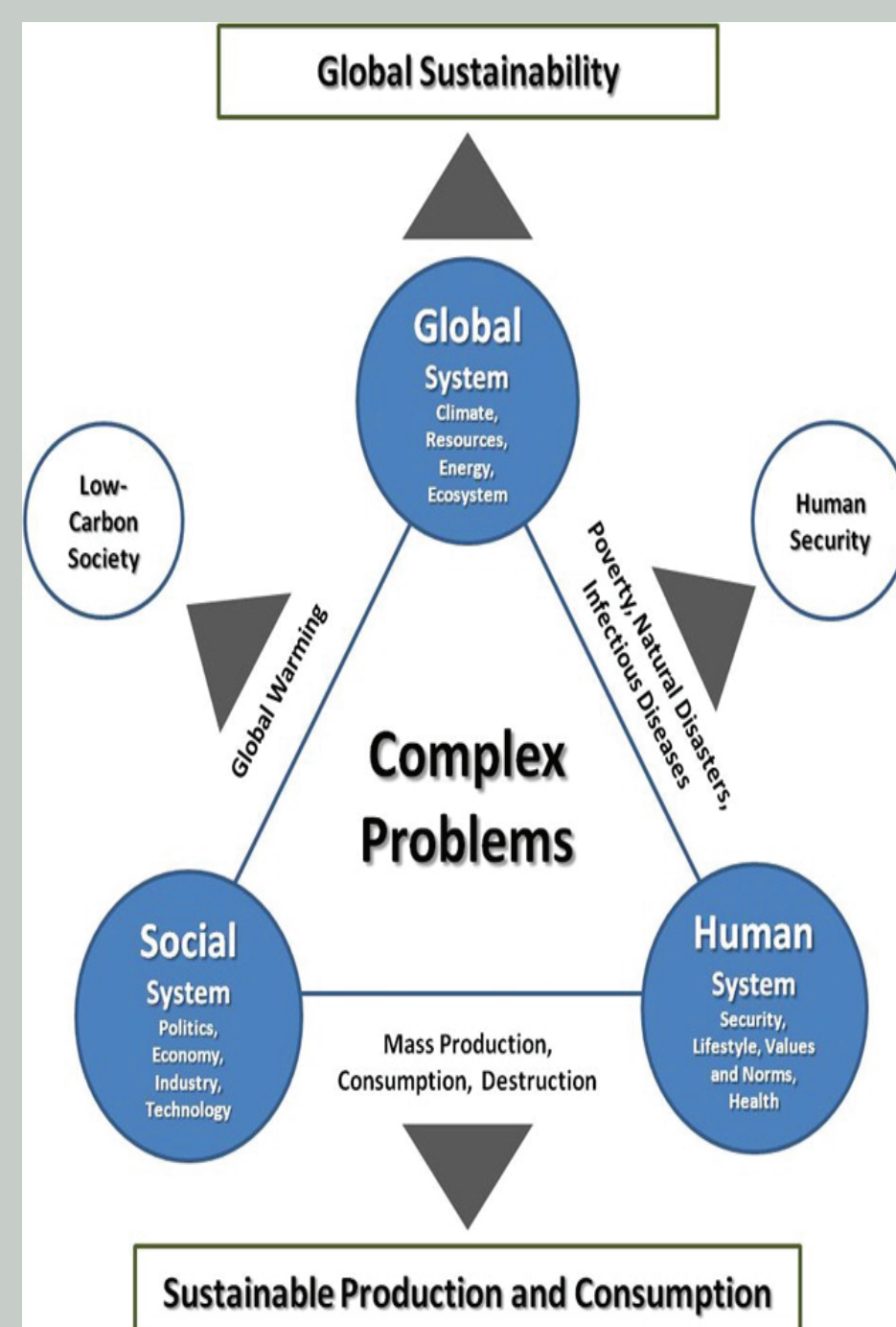
- Sustainability science is a data-intensive area that focuses on the complex interactions between nature and human activities.
- Sustainability research requires access to data from the physical and social sciences.
- But data are difficult to find, obtain and use because different disciplines collect, describe, and store their data in different ways.

2. Data Challenges in Sustainability Science

The long tail of scientific research data:

- Small and derived data sets
- Heterogeneous data
- Multiple sources of data
- Short-lived data with long-term value
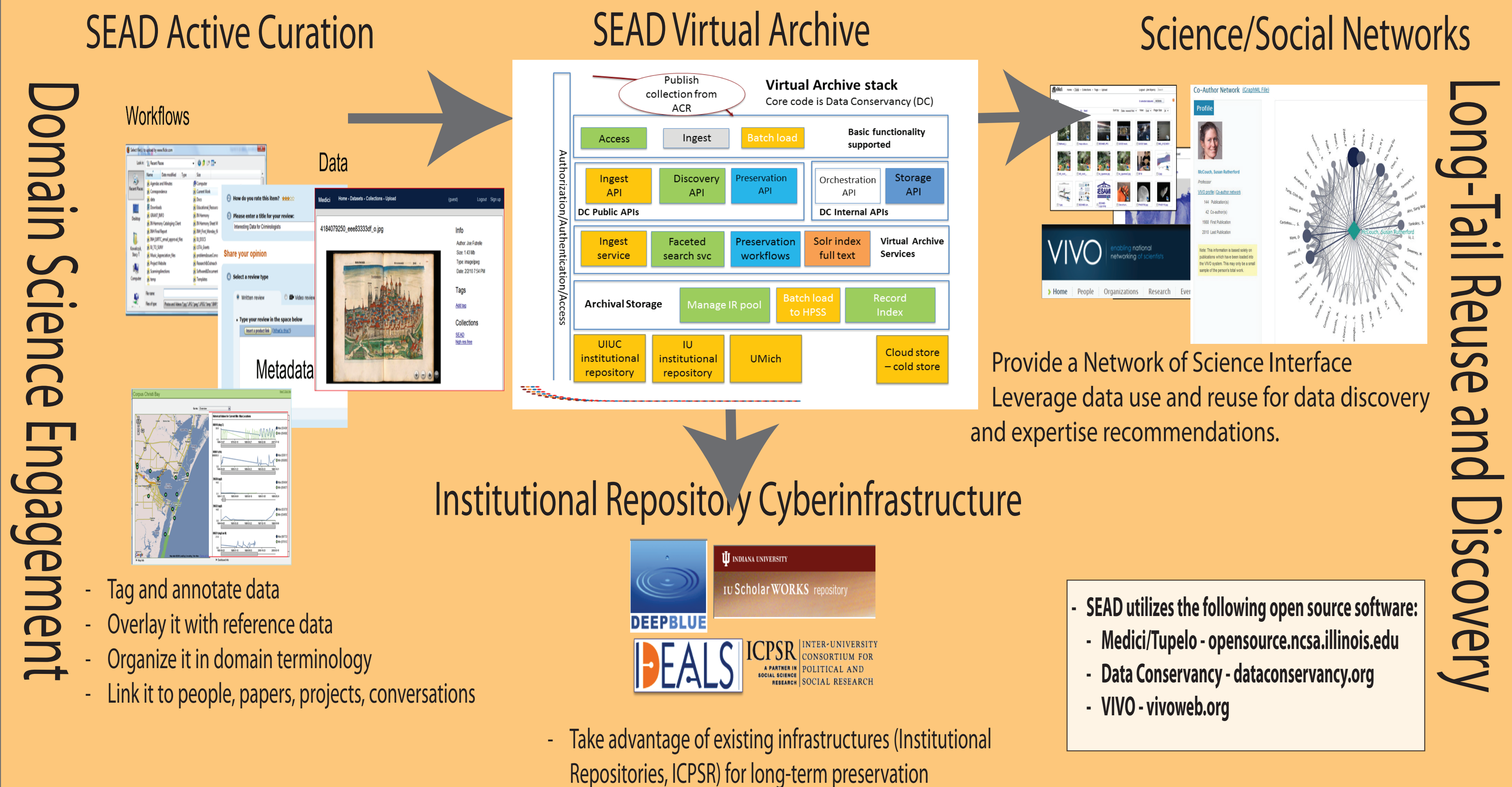- Value of data grows when combined & integrated

3. SEAD Goals

SEAD will address the needs of sustainability researchers to search for, aggregate, and maintain valuable data for the long term.

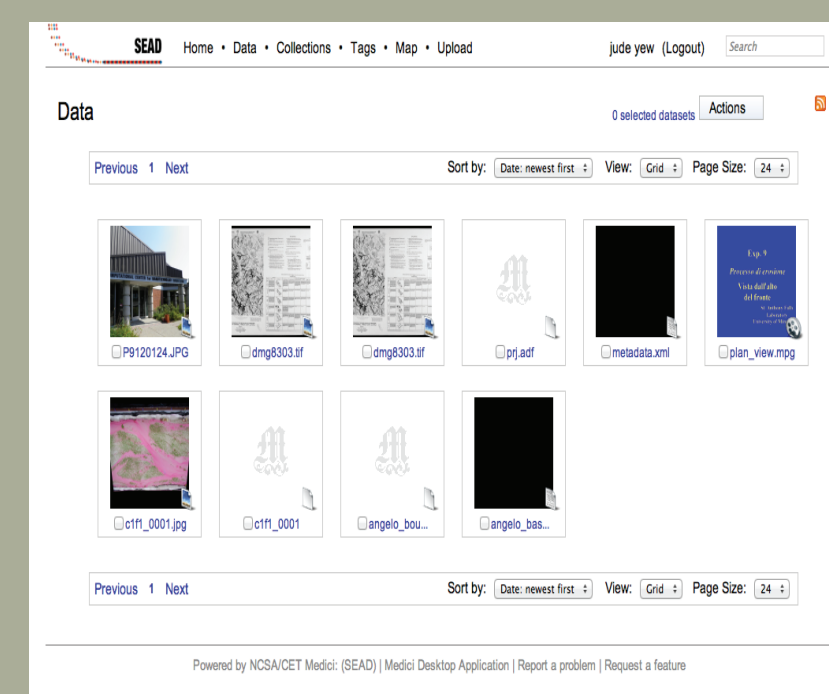To do this, the project seeks to build a prototype that:

- Applies existing tools and services to sustainability research
- Integrates these services into a generalizable "Active and Social Curation" infrastructure
- Enables researchers to collaborate and share data during active projects
- Packages and migrates data valued by the users to a federated repository infrastructure for long-term preservation

4. SEAD At a Glance

Domain Science Engagement

SEAD Active Curation

Workflows

Data

Metadata

- Tag and annotate data
- Overlay it with reference data
- Organize it in domain terminology
- Link it to people, papers, projects, conversations

SEAD Virtual Archive

Institutional Repository Cyberinfrastructure

- Take advantage of existing infrastructures (Institutional Repositories, ICPSR) for long-term preservation

SEAD utilizes the following open source software:
- Medici/Tupelo - opensource.ncsa.illinois.edu
- Data Conservancy - dataconservancy.org
- VIVO - vivoweb.org

Science/Social Networks

Long-Tail Reuse and Discovery

Provide a Network of Science Interface Leverage data use and reuse for data discovery and expertise recommendations.

5. SEAD Use Cases

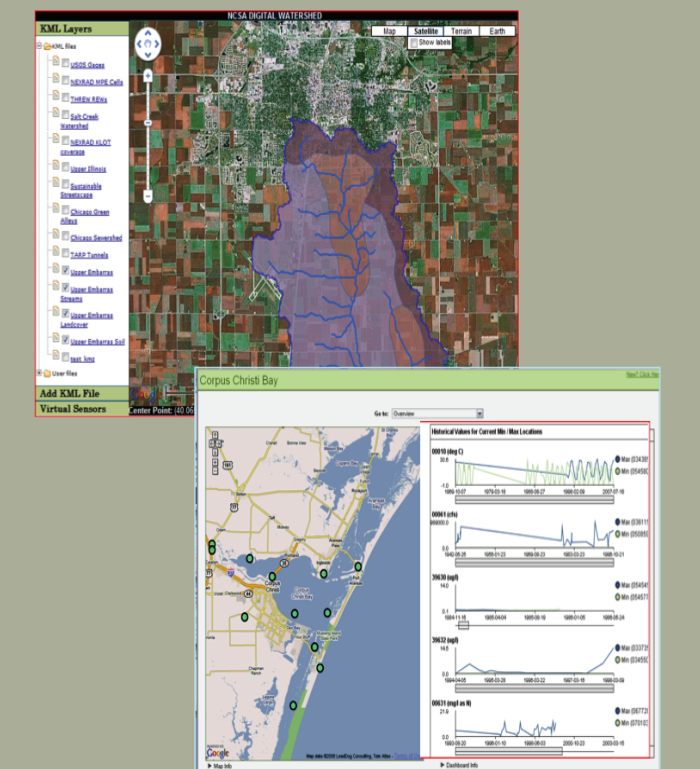### i) Able to ingest a variety of data types

- Users can store, manage, and share heterogeneous data types (e.g. images, geo-spatial images, sensor data etc.)
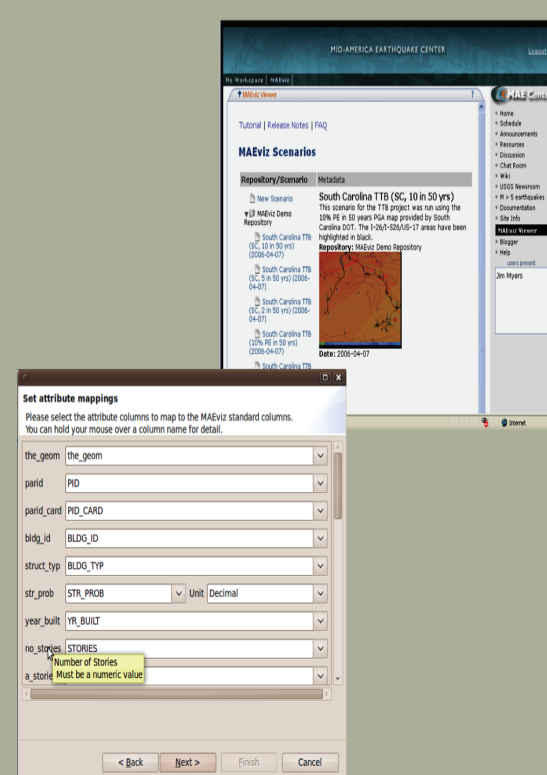
ii) Support data discovery
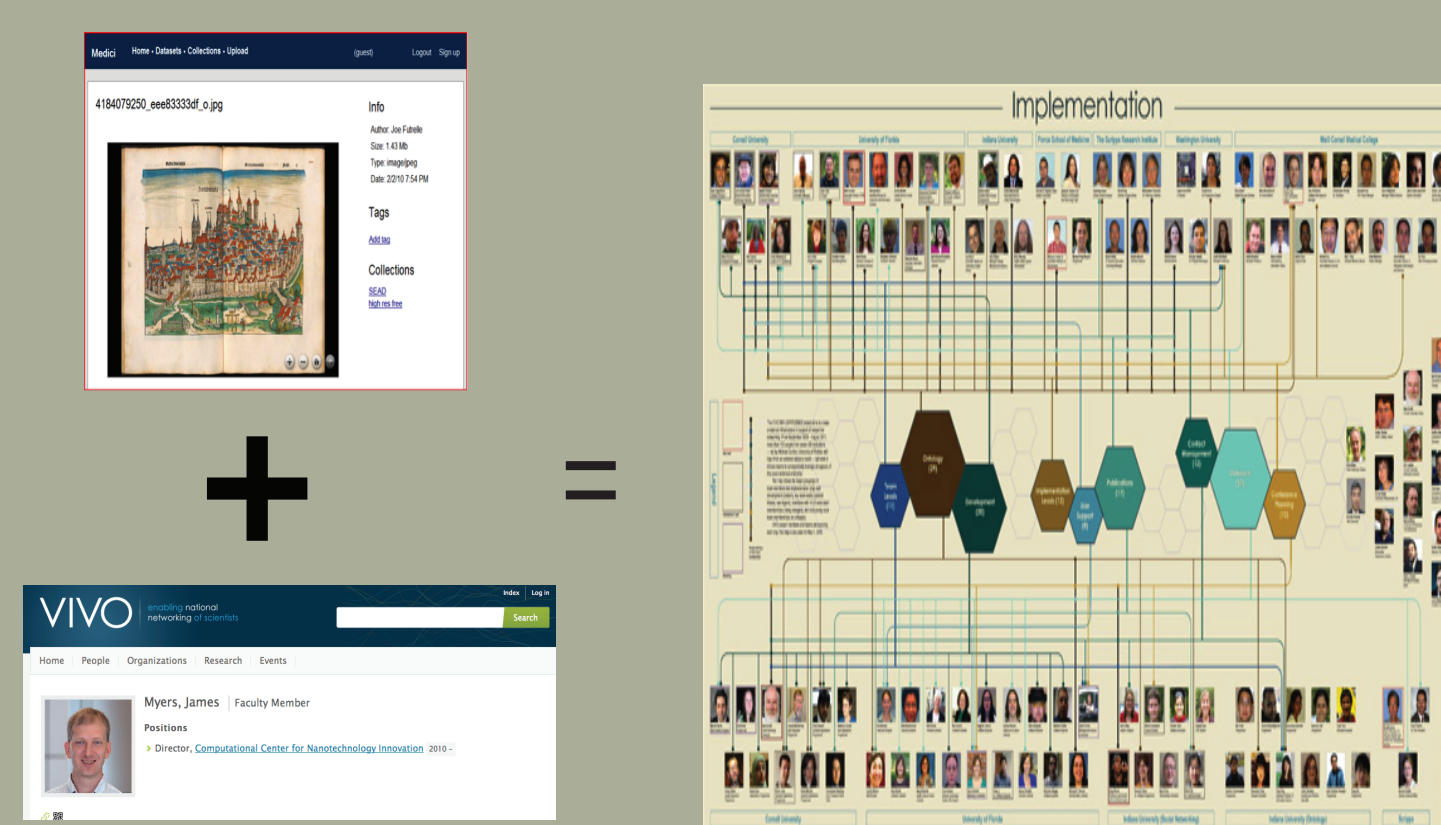
- Provide links between data, people and publications
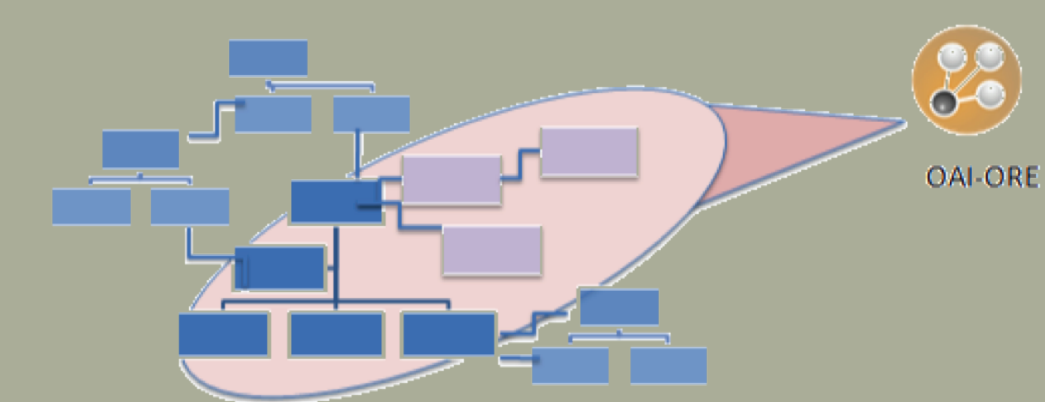
iii) Add value to existing data

- Users of data can provide additional metadata and annotations

iv) Create new data

- Combine data from multiple sources and contribute derived data back to SEAD

v) Community curation of data

- The community identifies and curates data of value
- These valued data will be moved to existing institutional repositories for long-term storage

- The SEAD team will work closely with the community of sustainability scientists to evolve these use cases over our first two-year timeline.
- SEAD will collaborate with scientists studying the Upper Great Lakes and Upper Mississippi River Basin and is a developing a test-bed partnership with the National Center for Earth-Surface Dynamics (NCED) to test NCED data in our 18-month prototype.
- Through this collaboration, SEAD will develop a prototype system using open-source software that engages sustainability researchers around data management, encourages sharing of data and metadata through new social and contextual interfaces, and leverages existing repository storage cyberinfrastructure.