The Community for Data Integration (CDI)
a community approach to managing science data

Leslie Hsu, U.S. Geological Survey
SciDataCon, 12 September 2016
How can a community of practice help with science data management?

- USGS is a **distributed organization** – practitioners may be isolated from those with similar interests.

- It is **difficult to keep up** with latest technology, tools, and policies.

- Some of the best innovative and practical ideas **bubble up from the workforce**.
The Community for Data Integration

Purpose

• **Advance understanding of Earth systems** through enhanced use of data and information including associated tools and techniques

• **Provide a forum for people doing work with data integration** to come together to share ideas as well as learn new skills and techniques

• **Advance overall USGS capabilities with data and information** by increasing visibility of the work of many people throughout the USGS
How does the Community for Data Integration achieve its goals?

- Annual proposal process
- In-person workshops and trainings
- Monthly virtual meetings
- Working groups
- External connections
- Frequent feedback from participants
Lucy, new USGS employee

Research CV

Federal Government Experience
A new world of policies

OFFICE OF SCIENCE QUALITY AND INTEGRITY
Scientific Data Management Foundation
Metadata for Scientific Data, Software, and Other Information Products
Review and Approval of Scientific Data for Release
Preservation Requirements for Digital Scientific Data

U.S. Department of the Interior
U.S. Geological Survey

Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data
Joining CDI

You should check out the Community for Data Integration

Search Products and Publications by:

- Data
- Tools
- Publications
- Best Practices

www.usgs.gov/cdi
Funded projects: A CDI-funded project in 2011-2012 turned into the USGS official data management website.
Special training events

Subject: Software Carpentry Workshop

Special trainings like the Software Carpentry workshop and Hackathon bring people together to meet and spend dedicated time on shared problems.
Special training events: Virtual training topics were chosen by the community. Topics related to the new data access policies were most popular.

2016 CDI Virtual Training Series

- Reviewing Metadata
- Version Control and USGS Software Policy
- Scientific Workflow and Reproducibility
Monthly virtual meetings

Scientist’s Challenge

Trying to solve a challenging problem and wondering if another group is working on the same issue?

We have Scientist’s Challenge “lightning talks” to kick off Monthly Meetings and post slides on the CDI community forum to stimulate discussion and generate community connections.
Working groups: Lucy found slide presentations from a recent data management webinar series on the CDI wiki

- Data Management Planning: overview and a USGS program experience
- Metadata: standards, tools and recommended techniques
- Best Practices for Preparing Data to Share and Preserve
**Member feedback:** Lucy contributed her idea for an Annual meeting session using the UserVoice platform, and voted on her peers’ ideas.

**What topics, events, and activities do you want to see at the CDI FY17 in-person meeting?**

- Metadata standards
- Machine readable formats
- Rescuing Legacy data
- Panel and Q&A on tools and policies for Data Release
The CDI Network:

Soon after joining the USGS, Lucy is part of a community of researchers, data managers, and IT professionals, spread across the nation and all with interest in promoting data integration and management.
Communication about useful tools and information for science data management is amplified by the member network.

You should check out the Community for Data Integration.
Community connections: CDI itself is just one node in a network of communities of practice supporting scientific data issues.
Community for Data Integration: a community approach to managing science data

• Connects the people, tools, and information needed for sound scientific data management.
• Brings together distributed, diverse staff.
• Spreads the word about new data management policies.
• Supports and disseminates new tools for science data management.
• Identifies shared needs across the USGS.

lhsu@usgs.gov    cdi@usgs.gov    www.usgs.gov/cdi
Main CDI activities

- Social media
- Working group activity
- Special workshops and trainings
- Request for Proposals

- Opportunities notifications
- Monthly meeting
- Annual meeting
A brief CDI history

- **2009**: Community for Data Integration
- **2010**: In-person meeting
- **2011**: USGS-wide Request for Proposals
- **2012**: In-person meeting
- **2013**: Community voting on Statements of Interest
- **2014**: Continued growth
- **2015**: Science Support Framework
CDI Working Groups

Dynamic groups that come together to solve common challenges

- High level working groups
- Focus groups

**2010**

- Tech Stack WG
  - joint with ESIP Tech Dive

**2011**

- Connected Devices WG

**2012**

- Data Management WG
  - Data Release Use Case, DOI Group, Data Management Policy, Science Data Lifecycle

**2013**

- Semantic Web WG
  - 3D/4D Viz

**2014**

- Citizen Science WG

**2015**

- Earth Science Themes WG

**2016**

- Communication WG
  - Science Center Strategy Dev, Process and Analyze Pages

*Dates are approximate as mentioned in CDI annual reports*