EarthCube Council of Data Facilities (CDF) Meeting Summary
July 22, 2016
ESIP Summer Meeting 2016
Summary based on notes from Inken Purvis, Sam Silva, and Shelley Stall

Co-chairs: Tim Ahern and Sara Graves (not present)

Actions/Decisions:

- ACTION: All users should register their repositories in RE3
- DECISION: Doug Fils was elected as the CDF liaison for the Technology and Architecture Committee (TAC)
- ACTION: CDF members should make additions to the current registry list by August 5th.
- ACTION: Subsequently, the EarthCube Science Support Office will create a survey. There will be a month to populate the survey.
- ACTION: Share with NSF the need to have technical staff, perhaps part of the ESSO, to implement needed (technical) activities.
- DECISION: Tim Ahern is selected as the Chair of CDF (2 year term)
- DECISION: Danie Kinkade is selected as Vice-Chair of CDF (1 year term)
- DECISION: Bob Arko is selected as Secretary of CDF (2 year term)
- DECISION: Sara Graves (1 year term), Jessica Hausman (2 year term), Chuck Meertens (1 year term), and Bernard Minster (2 year term) selected as at-large members
- DECISION: priorities have been suggested for development, they are:

    - Semantics
    - Brokering
    - Shared infrastructure
    - Making use of existing capabilities
    - Influencing future funding solicitations

Topic 1: Coalition on Publishing Data in Earth and Space Sciences (COPDESS)

- Slightly over 40 data publishers have joined
- Created directory of repository stack of authors/publishers to find appropriate data facility/repository
- Directory built with help of Centre of Open Science & AGU
- AGU primarily focused on reaching out
- Using Registry of Research Data Repositories (RE3) data
- Agenda item included navigating through the site to show how to get data ready for the directory. Specifically, this included:
    1) Walked attendees through registration process – created fake registration for demo
    2) Contact Shelley (SStall@agu.org) for further entries – don't enter name of repository

3) Login and access Directories of Repositories Management
4) URLs must be in URL format
5) Added lots of new fields for people to select from
6) Enter data type in your own words – DON'T HIT ENTER, hit + sign
7) Complete the rest – Shelley has a guide for any fields that may be confusing
8) Repositories are going to be endorsed by organizations that are already using them
9) Journal endorsements will be checked and verified – make sure you have one

ACTION: Request from Tim Ahern to encourage all users to register their repositories

Topic 2: EarthCube Architecture

- Report from Architecture Workshop 2016
- All Hands Meeting discussed that CDF must be involved for EarthCube to be a success
- CDF members discussed the key aspects of the architecture, remembering that it will not all be developed at once.
- Thoughts in response to the Architecture Workshop Final Report
    - Data facilities are the foundation
    - Thoughts on interacting? Is there a role for existing facilities? Is EC taking over technical development of key components of the architecture or not?
    - What things are being developed in the portfolio?
    - How to provide guidance back to the facilities?
    - Seed thoughts at All Hands Meeting discussion: too conceptual to say how people can interact with the system. It would benefit from moving from conceptual to a technical implementation plan in order to make progress.
    - What requirements are needed for people to start plugging into the system? Specifically, is a gap analysis needed?
    - From data facilities perspective: how many people came to repository because of EarthCube? How effective is EarthCube in bringing users to the repositories? Next step is getting to the point where we have architecture and can start considering implementation. This is an argument as to why both the facilities and the EC workbench need to tract metrics,
    - Leverage things that are already there? Why can't all repositories register data in a data form? Access data through common interface
    - NSF depositories make data available but not always usable – too esoteric
    - People don't give credit to the repositories they use – are the data facilities being taken for granted?
    - Some great ideas but a lot of repositories are already doing things being suggested – many infrastructures already exist outside the EarthCube environment – don't need to restart these things – many groups already using data management tools
        - Meeting will discuss Registry of Capabilities later

    - A real power data facilities have is an operational reality check – customers have to be served and implementing different technologies takes time effort & resources – you get what your developers know how to do – too much reinventing the wheel in the community – too many are doing the same thing – data facilities could be doing more: are there other things already out there – can we adopt someone else's stuff if it's already there? How much would that cost? EarthCube could get strategic about funding this sort of thing?
    - Centralized service for semantics would be great for linking up and exchanging information – developing new capabilities rather than inventing yet another vocabulary.
    - The data facilities exist – way to get them involved is not to make them change – develop brokering capability / find a broker who can do this.
- How are we going to move to actually do things – get facilities to talk to each other
    - Are we still too far behind to get real action? Can we get a proposal to get some real money?
    - Right now we are not moving along – these meetings are not helping (?)
    - Can we get action items and assign someone responsible for making them happen? Haven't done this in the past – we need champions, more structure, key working groups to report back on progress
    - 3 big federal agencies here with bigger budgets than EarthCube – we should figure out how to use them.
    - Priorities of Group: Semantics, brokering, shared infrastructure (making use of existing capabilities where possible), influencing future NSF/EarthCube-specific solicitations
    - We definitely need more dialogue so people are aware of what's going on in the various groups that are part of the repositories – better way to exchange information needed
    - Having more meetings probably not helpful – too many meetings already – can we have one thing to do this year that we can accomplish?

Topic 3: Workbench Components

- We're doing this capabilities inventory – in this context we should think about how this maps into the architecture
- Let's establish a goal for CDF for the year – something that really brings the facilities together – demonstrate the value of having CDF and make it easier for the next project
- A goal going forward could be conduct a survey, send out under CDF & get a consensus of everyone's thought on next steps
- What are the capabilities:

Goal to Take Away

- Establish specific action items going forward
- We need a CDF liaison to the EarthCube Technology and Architecture Committee: Doug Fils was unanimously selected.

TOPIC 4: CDF Facility Registry of Capabilities

More information can be found at https://docs.google.com/document/d/1-mA6HRdZRGTVq7STlo9p5rDH5NLrbxYU3xRObc4FkkU/edit?usp=sharing

- Concern noted that there were other registries – such as re3Data and World Data System. That had originally caused a pause to conducting the survey
- With the discussion of the EarthCube architecture the needs to have a full understanding of the CDF repository capabilities was important. The decision was made to move forward with the survey.
- ACTION: CDF members should make additions to the current registry list by August 5th.
- ACTION: Subsequently, the EarthCube Science Support Office (ESSO)will create a survey. There will be a month to populate the survey.
- Discussion about who is the audience for the information. The potential variation on answers is too varied for some of the topics.
    a. Re3data has many of these elements.
    b. Dataone has a way to harvest. We need a method for this information to be maintained going forward.
- Suggestion to register a URL where the information is located. Don't do another survey.
    c. Look at the 6 points that the DataOne member nodes share as a starting point.
- Guidance to need to stay in alignment with re3data.org
- This information is needed to help guide the implementation of the architecture.
- This task is driven by the group and needed by the group that will develop the architecture. Also the members of the CDF and NSF are aware of the current capabilities and services.
    d. Re3data does not have all records.
- An automated system is attractive, but who will develop that?
- Question raised about an implementation group in EarthCube and/or funds with a response noted that this is under development; however, we need a group working at the ESSO to begin the implementation.
- Currently, the ESSO award does not fund a technical person. That conversation is taking place including figuring out how to fund the position who will manage it going forward.
- CDF could make a recommendation that a facility should be co-located with the ESSO and have a technical support team EarthCube cannot move forward without a core technical group.
- There is a test bed funded project. EarthCube Integration and Test Environment (ECITE). It should be reviewed as a possible solution.
    - As a test bed it's not a persistent solution.
    - Advantages to placing the technical group with the office since it look like it is in production and not a short-term solution.
- ACTION – Share with NSF the need to have technical staff, perhaps part of the ESSO, to implement needed (technical) activities.

- Need to issue clear guidance on the level of granularity on the survey responses.
- Is there a conflict with DataONE? Can that be a possible full/partial solution.
    - This will work. Start with a survey to get the responses right and make adjustments.
    - Should put examples in the survey to help get the correct responses.
    - A steering committee should hone this and test with 3-4 diverse repositories.

TOPIC – Other Business - Tim

None was presented.

TOPIC – Election of Officers. Tim/Mohan

Tim – At the EarthCube All Hands meeting we asked for nominations for CDF leadership. There were 7 positions. And 7 names. All were contacted and willing to serve.

Tim deferred to Mohan to conduct the elections as he was included in the nominees.

Mohan – The CDF is structured such that leadership is decided upon outside the EC structure.

All 7 positions are open at this time. Terms have been adjusted to avoid a full turn over at the next election.

- Chair
- Vice-Chair
- Secretary
- At-Large (4)

Names that have been submitted for nomination along with the recommended position:

• Jessica Hausman – JPL PO.DAAC	Vice Chair (1-year)
• Bob Arko – Columbia R2R	Vice Chair (1-year), Secretary (2-years)
• Sara Graves – GHRC DAAC/IPSC NASA	At-Large #1 (1-year)
• Danie Kinkade BCO-DMO	Chair (2-years), At-Large #2 (2 years)
• Chuck Meertens – UNAVCO	At-Large #3 (1 year)
• Bernard Minster WDS/ Scripps	At-Large #4 (2 years)
• Tim Ahern – IRIS	Chair (2 years)

To be a CDF officer, the recommended person must be from a member organization of the CDF.

Discussion:

- Kerstin – Recommend that going forward we need a process that includes a few sentences from each nominee as to their desire and willingness to be a CDF officer.
- Jessica – Logistical concern - Her organization doesn't receive NSF funding, and therefore won't provide travel funds to participate in face-to-face meetings. This

will likely impact her ability to attend CDF meetings held at the EarthCube All Hands meetings. Meetings at ESIP are not an issue.

Selection of CDF Chair

Danie Kinkade – Rescinded her nomination for Chair.
Mohan – Those in favor of Tim Ahern as Chair of CDF, raise your hands?
All CDF members were in favor.
Tim Ahern is selected as the Chair of CDF

Selection of CDF Vice-Chair

Bob – Rescinded his nomination for Vice-Chair.
Jessica – Rescinded her nomination for Vice-Chair
Danie – Self-nominated for Vice-Chair. Doug Fils – Seconded.

Mohan – Those in favor of Danie Kinkade as the Vice-Chair of CDF, raise your hands?
All CDF members were in favor.
Danie Kinkade is selected as Vice-Chair of CDF

Secretary

Bob Arko has been nominated.
Bob stated he is happy to be Secretary for CDF.

Mohan – Those in favor of Bob Arko as the Secretary of CDF, raise your hands?
All CDF members were in favor.
Bob Arko is selected as Secretary of CDF

At-large member slate:

Sara Graves - At-large #1 – 1-year
Jessica Hausman At-large #2 – 2-year
Chuck Meertens - At-large #3 – 1-year
Bernard Minster - At-large #4 – 2-year

Mohan – All in favor for the slate as it is presented? [it was shown to the room on a projector]
All CDF members were in favor.
The At-large members have been selected for their designated terms.

Elections Complete.

CDF Registry – Suggestions as of July 22, 2016

ISO 16363

The ISO standard for trusted digital repositories (ISO16363) provides a useful construct for characterizing, discussing, and evolving the aspect of data facilities having to do with digital data repository services around the following areas:

● Governance and organizational viability – strategic planning, succession planning, contingency planning, etc.

● Organizational structure and staffing – workforce planning, professional development, etc.

● Procedural accountability and preservation policy framework – preservation policies, change management, transparency, information integrity, etc.

● Financial sustainability – business model, transparent accounting practices, risk management, etc.

● Contracts, licenses, and liabilities – deposit agreements, license management, intellectual property rights, etc.

Mohan's Facility Registry – list of potential entries:

- Name of the facility
- URL for the facility
- Mission statement
- Principal sponsoring agency, directorate, and division (e.g., NSF/GEO/EAR)
- Approximate annual budget
- Number of staff members
- Main domain/discipline served
- Provides data? If so, historical, real-time, etc.
- Data types, formats, standards, etc.
- Data access APIs, protocols, and services
- Restrictions on data access and use
- Collects and archives data from users/PIs? If so, links to curation/archival policy, procedures, back-up systems, disaster recovery procedures, etc.
- Maintain a database of users and other relevant information?
- Develops and provides software? If so, what types of software are provided? (e.g., data management, processing, analysis, visualization, synthesis, etc.
- List of software, fees for use, and licensing (e.g., open source, proprietary, etc.); If open source, what type of licensing?

- Software: programming languages, desktop vs. web-based, …
- Tools provided
- Collect and archive physical samples? If so, a brief description
- Provide user support on data, software and other topics?
- Provide training to users? If so, list training topics and frequency.
- Conduct user workshops?