

Workshop Report for the C4P Synthesis Workshop

May 13-15, 2015, LDEO, Palisades, NY

[Agenda](#)

Website: <http://earthcube.org/workspace/c4p/workshop-3-synthesis-workshop>

Workshop Convenors: Kerstin Lehnert, Anders Noren, and the C4P Steering Committee

Report prepared by: Leslie Hsu, C4P Coordinator

Table of Contents

[Summary](#)

[Workshop Goals](#)

[EarthCube Context](#)

[Recommended Practices](#)

[Taxonomy](#)

[Unique Identifiers and Metadata Requirements for Paleo Samples](#)

[Publications: a way to disseminate Recommended Practices](#)

[COPDESS: working with Publishers to implement Recommended Practices](#)

[Engaging the Community of Practice](#)

[The Future of C4P](#)

[Actions](#)

[Participants](#)

Summary

The third workshop of C4P, Collaboration and Cyberinfrastructure for Paleogeoscience, was held at Lamont-Doherty Earth Observatory in Palisades, NY. 28 participants attended, including both those that have been involved heavily in C4P and those that were relatively new to C4P's activities.

The goal of the workshop was to plan for the long-term impact of the outcomes of the C4P project. Discussion focused on documenting and disseminating recommended practices, engaging the community of practice, and future grants and workshops. Major outcomes were the planning of two articles about C4P activities, plans for the geology and biology communities to work closely to align metadata profiles for samples, the agreement about the importance of reaching early career researchers and disseminating C4P products such as the C4P resource catalog, and plans to reach out to editors about paleogeoscience data best practices via the COPDESS (copdess.org) initiative.

Workshop Goals

The primary goals of the workshop, as planned, were: 1) Alliances, Consortia, Councils, & Committees - How can the C4P community continue the RCN for long-term impact?, and 2) Advancing adoption & implementation of leading practices developed and recommended by C4P workshops.

An initial discussion in small groups at the start of the workshop identified the following main topics:

- Sustainability of C4P activities
- Collaborations, branding, visibility
- Standardization, evolution in future
- Data sharing
- Increasing participation beyond database insiders (everyone generates data)
- Ease of use of C4P tools and resources
- Promoting active data discovery
- Penetration of many related communities

EarthCube Context

Brief talks about different aspects of EarthCube familiarized all participants with the context of C4P as a Research Coordination Network within EarthCube. These were: EarthCube Building Blocks (Zaslavsky), Scope and Vision Workshop (Ferrini), Tech Hands Meeting (Fils), and Science-Tech workshop (Zaslavsky)

Ways that C4P could contribute to EarthCube are to:

- submit a formal Use Case,
- submit informal use cases or use scenarios to Building Block projects,
- test Building Block products,
- comment on the reports and white papers that come out of different standing committees (e.g., [Geoscience 2020](#) from the Science Committee).
- communicate our recommendations for best practices to the proper EarthCube committees for dissemination and integration.

It was mentioned that EarthCube projects may be transient, and that there are pros and cons of integrating our practices with projects that may not exist in the near future.

Recommended Practices

During the C4P project, several topics have been raised by the community as needing guidance on recommended best practices. These topics were introduced by steering committee members, followed by a discussion of recommended practices. The C4P Steering Committee plans to produce a paper on recommended practices for the paleogeoscience community from these discussions.

Taxonomy

(led by Mark D. Uhen)

The expression of taxonomy in the literature lacks community-supported best practices, leading to challenges in documenting, discovering, and using this information. The discussion focused on identifying recommendations that would improve collection and sharing of taxonomy data.

Recommendations for best practices include:

- a call for clear, concise statements regarding all nomenclature, and the avoidance of vague statements;
- a call for authors to include all taxonomic levels in systematic paleontology sections (do not skip over selected higher levels);
- a call to follow all taxonomic rules carefully, and;
- an encouragement to correct previous taxonomic problems that are highlighted in the publication.

Participants noted that funding agency or journal editor backing was necessary for appropriate motivation and implementation of these recommendations. Breakout groups discussed possible carrots and sticks (incentives and threats) that would help adoption of the practices. Carrots and sticks for the recommendations included:

Carrots:

- an error checking tool for taxonomies, and an associated electronic "verified taxonomy badge" for webpages, data systems, or papers, signifying that recommended best practices were followed
- including taxon authors in a bibliography, thereby giving proper credit
- providing editors with guidelines or software support for checking taxonomy
- involving professional societies to spread the culture of best practices
- generating a list of examples of best practice to help users and give credit to good examples

Sticks:

- working with funding agencies to set up requirements to follow best practices
- working with publishers and editors to set up requirements (see COPDESS section below)

- working with community data/object (sample) repositories to require best practices

Unique Identifiers and Metadata Requirements for Paleo Samples

(led by Kerstin Lehnert)

Unique identifiers for physical samples in the paleogeosciences are important for both interoperability of data systems and sample management by investigators. One identifier system that is quickly gaining adoption in the geology community is the IGSN, International GeoSample Number. A goal of C4P is to better integrate geo and bio sample identifiers.

The discussion led to many more topics for investigation, which could be part of a future proposed project:

- How can we integrate IGSN with Globally Unique Identifiers (GUIDs) in Bio?
- Does the IGSN system work for paleobio samples? (Is the current set of metadata ok?)
- If not, what additional fields are needed for a metadata profile for paleobio samples? (both at the global level and at allocating agent databases) That is, how to allow different communities to apply their own constraints in addition to a core set of metadata.
- What controlled vocabularies are needed? Vocabularies for time/age; taxonomy.
- How can we broaden adoption of these recommended paleogeoscience Unique IDs?

Recommendations included:

- the decision to align IGSN metadata to the DarwinCore standard

Publications: a way to disseminate Recommended Practices

(led by Jack Williams)

The C4P steering committee had been discussing a publication in EOS to communicate C4P lessons learned. At the workshop, the group decided to have two separate articles, a shorter one in EOS and a longer, perhaps more technical one in a different publication such as GSA Today. The second publication may describe a case study or use case (depending on how technical it is). A writing plan was formulated.

COPDESS: working with Publishers to implement Recommended Practices

(led by Kerstin Lehnert)

The Coalition for Publishing Data in the Earth and Space Sciences, www.copdess.org, intends to provide a framework for publishers and data facilities to implement and promote common policies and procedures for publication and citation of data. C4P discussed working with COPDESS and C4P-related data facilities to promote best practices for publishing and citing paleogeoscience data.

Engaging the Community of Practice

Presentations were given to describe existing or emerging projects that can help to continue the C4P-related community of practice after its funding period is over.

STEPPE (Dena Smith): STEPPE (steppe.org, Sedimentary Geology, Time, Environment, Paleontology, Paleoclimatology, Energy) is an NSF-supported consortium that is a voice for research on the sedimentary crust, deep-time and earth-life system. STEPPE provides a database of white papers and reports, a funding database, and announcements of ongoing events and new opportunities. Proposals for funding of up to \$15,000 for workshops is available. STEPPE may provide resources that C4P could include in a renewal proposal.

iDigBio (Gil Nelson): iDigBio (<https://www.idigbio.org/>) is an NSF-funded digitization effort for every specimen in North America. The tools and software are available through APIs. iDigBio holds frequent workshops for training investigators on the digitization of samples. C4P hopes to work with iDigBio into the future, and should look into what components may be transferable or reusable in C4P.

Deep Time Data Infrastructure (Mike Meyer): Deep Time Data Infrastructure develops, curates, and integrates diverse data resources from mineralogy, petrology, paleobiology, paleontology, paleotectonics, paleomagnetism, geochemistry, geochronology, genomics, and proteomics. Components include a Deep Carbon Observatory dataset browser and statistical and visualization tools. C4P and DTDI may find opportunities for collaboration and leveraging of each others resources.

Paleobiological Data Consortium (Jack Williams): The Paleobiological Data Consortium is a proposed project to maintain communication between geodata providers, biodata providers, the open source community, the users, and other related communities. The Consortium could act as a vehicle to promote and adopt best practices from C4P.

Open Core Data (Doug Fils): Open Core Data (OCD) is an initiative to address the need for a common mechanism to collect and distribute core-related data. OCD is working with the EarthCube Building Block project GeoLink to create ontology design patterns and Linked Open Data. C4P-related data systems and community could benefit from these activities.

IEDA (Interdisciplinary Earth Data Alliance) (Kerstin Lehnert): IEDA proposes a new multi-institutional alliance to enhance and support services and sustainability of long-tail data systems in order to enhance data discovery, access, and integration across disciplines. The proposal funds a few testbed communities to take part in the alliance, take advantage of EarthCube-created tools, and use existing IEDA infrastructure for shared data services. C4P-related communities may be Alliance members in the future.

C4P CINERGI Viewer (Ilya Zaslavsky): The CINERGI EarthCube Building Block project is focused on building a catalog of interoperable and well-described resources to the user community. The C4P RCN contributed a list of hundreds of dataset, software, and physical sample repositories, compiled by student interns in the summer of 2014. CINERGI created a Community Resource Viewer that allows easy browsing, discovery, and community curation of these C4P-related resources. Workshop members decided that a working group should check and improve the metadata for these resources.

The Future of C4P

Small groups discussed future directions, opportunities, and questions from the C4P community. Highlights included:

- The need to contact the geochronology and paleoclimate communities for future RCN activities. A recent geochronology report, "It's About Time: Opportunities & Challenges for U.S. Geochronology" (Harrison, T.M., Baldwin, S.L., Caffee, M., Gehrels, G.E., Schoene, B., Shuster, D.L., Singer, B.S.) was provided.
- The importance of establishing closer Geo-Bio collaborations, but the need to be strategic about selecting the correct activities to target first, since funding agencies are wary of the overlap. A draft document for best practices that could be integrated into the publication and data submission process was proposed.
- The continuation of specific topics via working groups, including Early Career Outreach, Generating Best Practices.
- A hackathon to work on implementing new ideas with both computer scientists and geoscientists.

For the last session of the workshop, When asked what aspect of the synthesis workshop topics they would like to remain engaged in, several more ideas emerged.

- Linking C4P and the EC3 EarthCube RCN to have a dialog about linking laboratory and field data. Possible topic for workshop/funding: Tools and Reproducibility in the Field Sciences.
- Participating in the sessions that the EarthCube engagement team has set up at GSA and AGU.
- Continuing work on visual representations of geologic time.
- Choosing a well-determined time (e.g., the Mesozoic) and incorporating different data to determine the Standardization and Best Practices needed.
- Continuing to promote the resources that C4P has created, including the CINERGI Resource Viewer and the webinars.
- Increasing relevance of the C4P Use Case by figuring out C4P community grand challenges such as time-3D coordinate integration/implementation/viz.; Catalogs: how do people do data discovery; taxonomies.
- Have a hackathon where DarwinCore is presented to C4P community, the participants determine what is required to build an extension, and build it.

- Educating early career scientists about standards and integrating with the iSamples EarthCube RCN.
- Training domain-specific information literacy at the undergrad level.
- Continue outreach to Libraries

Actions

The high priority action items were:

1. A Best Practices Document (MUhen volunteers in making an outline/Google Doc)
 - a. taxonomy (RGuralnick, JohnWz, MUhen)
 - b. age data (NoahM, (DColeman and BSchoene))
 - i. what does someone with a fossil need to include in a publication to make the age very clear (to others not necessarily in that field)
 - c. samples (KLehnert, ANoren, AMolineaux)
 - d. deadline ~ GSA / October
 - e. Names listed are leads but everyone should be able to see/comment/contribute (Open Science Project platform? MUhen will explore)
2. Darwin Core alignment (RGuralnick, JWieczorek, KLehnert, (JWilliams), DFils)
 - a. What deadline: Sept/Oct, TDWG
3. Hackathon
 - a. Revisit this
4. Workshops
 - a. Lunch workshop at GSA specifically targeting Early Career, and should present Best Practices, get feedback on where they see challenges for adoption
 - b. There will be a samples breakout session at ESIP in July
5. Cleaning up the CINERGI catalog
 - a. MUhen, MMeyer, IZaslavsky, LHsu, DFils
6. Need to continue to do outreach to Early Career Researchers.
7. Continue reaching out to editors through COPDESS.

Participants

Carter, Megan	Lamont-Doherty Earth Observatory
Coleman, Drew	University of North Carolina
Dong, LingLing	Lamont-Doherty Earth Observatory
Ferrini, Vicki	Lamont-Doherty Earth Observatory
Fils, Douglas	Consortium for Ocean Leadership
Gross, Wendy	NOAA Paleoclimatology Program, National Climatic Data Center
Guralnick, Rob	University of Florida
Hsu, Leslie	Lamont-Doherty Earth Observatory
Ito, Emi	University of Minnesota
Lehnert, Kerstin	Lamont-Doherty Earth Observatory
McLean, Noah	University of Kansas
Meyer, Mike	Carnegie Institution for Science
Molineux, Ann	University of Texas at Austin
Mookerjee, Matty	Sonoma State University
Nelson, Gil	iDigBio/Florida State University
Nichols, Jon	Lamont-Doherty Earth Observatory
Noren, Anders	University of Minnesota
O'Connor, Patrick	Ohio University
Olszewski, Thomas	Texas A&M University
Park-Boush, Lisa	University of Connecticut
Schoene, Blair	Princeton University
Smith, Dena	University of Colorado
Song, Lulin	Lamont-Doherty Earth Observatory
Uhen, Mark D.	George Mason University
Wahl, Eugene	NOAA Paleoclimatology Program, National Climatic Data Center
Wieczorek, John	UC Berkeley
Williams, Jack (John)	University of Wisconsin
Zaslavsky, Ilya	San Diego Supercomputing Center