

LC Strategic Priorities 2018 - approved April 2018

For distribution to NSF PMs, ESSO, EAG and EarthCube Governance Committees and Teams.

The EarthCube Leadership Council (LC) has established immediate programmatic priorities based on feedback it received from the NSF (in response to the LC's October 2017 draft roadmap - https://docs.google.com/document/d/1zFQGs-vxS2y1GQn0Xdg5TnBp6cxTdZPY3P_Rir24yuM/edit?usp=sharing). The NSF identified three priority areas for the LC to focus on and made suggestions on possible activities to achieve them, which the LC has discussed during the preparation of this document. Progress on these Priorities will be accomplished by the LC and other governance committees, ESSO, and the EC community.

In this document, the LC describes the strategies for activities over 2018 and beyond, the order it will pursue these activities, the resources needed, and other related activities to help the EarthCube community make progress on the priorities. Some activities are already happening while others will start soon.

“Short Term EarthCube Priorities”¹

The three priorities identified by NSF involve development and assessment of an EarthCube registry, science community engagement, and the feasibility of developing an EarthCube workbench. The LC also plans to

- Monitor and assess progress and impact of priority activities, defining metrics and milestones;
- Engage and interact with EarthCube committees, EarthCube Teams, and the ESSO;
- Evaluate the resources needed (those that could go into the ESSO's APP for FY2019 were discussed with ESSO already, and those that can go into the next EarthCube solicitation are under discussion currently). Additional unmet needs will also be discussed.
- Develop more rigorous timelines, action plans, responsible parties and monitoring of action plan progress.
- Keep the EAG apprised of all plans.
- Develop a plan for shared assessment of EC activities and outcomes, particularly with respect to end user utilization and experience

Each of the three priority areas are discussed below in light of their role within the EarthCube program. However the order of the topics in this document does not imply higher priority.

Priority A: Science Advancement

Specific Activity: Science Community Engagement. Engagement in this context means developing strategies to get active scientists to learn about and use existing EarthCube components, to identify data, software and database needs, to promote FAIR principles with respect to NSF-supported GEO data and other entities via EarthCube efforts, and to develop a staged priority structure for this potentially far-reaching effort, so that high-impact and low-effort activities can be addressed first, and with some emphasis on science funded by the NSF-GEO directorate.

¹ These refined priorities and actions were developed at a A F2F meeting of the LC on Feb 7-8, 2018

The Science Committee will be charged with assessing and providing answers to the following questions (paraphrased) from NSF, with assistance from the Engagement Team. Active SC discussion of the general topic is already underway:

- What is the scope of this effort? Design a strategy to assess the scale and resources needed.
- What is feasible and on what timeframe?
- How can the Science Committee work to scope this effort in collaboration with the Engagement Team / TAC and, supported by ESSO staff?
- How can data facilities be better leveraged by the research community to connect users to capabilities? Determine the correct DSEs or EarthCube projects that have data and services for determined use cases
- How will user feedback be gathered and acted upon in a consistent way?

Additional priority actions:

- Explore the use of a single point-of-contact, EarthCube branded resource to raise awareness and usage (e.g., a workbench or portal), with SC, TAC and CDF input.
- Provide input to NSF solicitation for projects to be lead by earth science leads and with technologist co-PIs. these projects will implement EarthCube tools into science workflows, with science applications driving the activities.
- Monitor impact of EarthCube funded engagement projects like ASSET, CRISP, and ECITE
- Work with geoscience community, especially through the new NSF-GEO-CI data workshop effort (2018-2019) to provide input on the types of end user capabilities that are preferred: visualization tools, output data format types, ease of use, selected analysis tools, etc.
- Work with DSEs to promote new services on their website and link to the demo workflows (get visibility at the locations scientists access data)

Priority B: Registries: Resource Integration and Reuse

To allow for new technical and scientific capabilities to be built from existing resources, and to maximise reuse of resources, LC wishes to integrate NSF's cyberinfrastructure investments, including but not restricted to those funded by the EarthCube program. These will be done initially via ESSO-funded pilot projects: **Data Repository Registry** (specifically, a registry of data repositories, related P418, P419, and metadata/GUI efforts) and **Resource Registry**. Overarching goals include:

- Promote within EC the idea that resources to be incorporated not as winners vs losers of the EarthCube ecosystem, but instead as what is more and less central to current science workflows.
- Focus should move away from products per se, to providing means and mechanisms for scientists via tools and data access. This is a way to refocus EC to be more relevant.
- Keep registries open to all data, all resources, "EC or not".
- Develop these registries so that anyone can put stuff in, so long as they follow guidelines

Specific Activity #1: Data Repository Registry. Evaluate the impact of the P418 pilot data center registry effort and establish a plan to both assess P418 outcomes and build on these efforts, along with expansion to other future registry components. Two proposed follows-ons are P419 (production software) and a GUI +Metadata expansion effort, to be conducted as pilots in FY 2019.

In response to NSF suggested activities and questions, the following steps have been implemented thus far or are planned:

- Develop working demonstrations of the P418 pilot using Use Cases and put in for a demo at the 2018 All Hands Meeting (AHM). To this end, the Science Committee (SC) was tasked by the LC with soliciting and vetting Use Cases, and then passing them to the P418 leads so they could work directly with Use Case proponents on the demos. As of Mar. 2018 the SC work is completed and the Use Case teams are working with the P418 PIs on the demos.
- Ask the EarthCube Technical Officer to establish milestones and basic assessment metrics for P418, with input from TAC and CDF as needed.
- Have the SC propose metrics for the impact of P418 on the science community. Identify who else should assess & provide feedback? TAC, CDF? NSF-GEO
- Discuss possible follow-on activities, which as planned currently include openly recompeting a second year of P418 ("P419") to aid in data center efforts to populate the registry and provide pilot funds to establish a pathway towards a parallel and fully interoperable (with P418/419) registry effort on EarthCube non-data components.
- Promote interaction through the ESSO and EarthCube governance groups (e.g., the CDF and other GEO DSEs = Data Service Entities) for trial adoption of the P418 methods and technologies. The specifications for P418 should be written up to promote consistent adoption methodology across DSE's & CDF, have the Technical Officer gather high level information about obstacles to adoption, and how we can promote roll-out to other (even non NSF-funded) data facilities.
- EarthCube governance and ESSO establish assessments of current and future registry efforts in multiple key areas. Attempt to use a combination of governance groups and ESSO. If this doesn't work, stand up an assessment working group from TAC and SC. Use nascent ESSO Technical Officer and the older TAC catalog of capabilities to facilitate discussions. For instance, items on the to-do list include:
 - evaluate progress toward P418 deliverables
 - § responsibility is with ESSO, but LC will monitor the project timeline
 - § Require a report to ESSO and LC from the P418 advisory team
 - adoption by data facilities
 - § ESSO and the Technical Officer to monitor how many metadata/data providers will be in the registry by end of year 1 and provide updates to LC
 - sustainability
 - § ESSO will identify who will prepare a plan for sustainability of registry. Who will maintain it, who will run it?

Specific Activity #2: Resource Registry. Create a registry of EarthCube technical resource components, including standards and interfaces.

Once fully realized, such a registry should enable end users to discover scientific and technical resources (software, tools, vocabularies, etc) that are relevant to their research. It will allow the developers to determine which technical components exist and may be reused, and provide NSF program managers and reviewers with a tool to evaluate the novelty of proposed work. It will also

provide a vehicle to notify the community about new products, increase visibility, and gain recognition. Seed money for this effort has been included in the ESSO proposal to the NSF. The EarthCube TAC Resource Registry WG has defined a controlled vocabulary for the EarthCube resource component registry, and is working with the ESSO on an initial registry. Formally, the project has not yet started, but is an LC priority in 2018.

Priority C: Workbench: Scientific Workflow and Data Platform

The idea of a workbench has been around since the earliest days of EarthCube, but the term has meant different things to different groups of people. The lack of both a list of desired capabilities and an assessment of existing technologies developed by other groups has made this effort more difficult to define. The EarthCube LC has established a Tiger Team to address these overarching and related questions in Spring 2018, with an aim to provide direct answers to questions received from NSF on our 2017 roadmap workbench activity, and to make a high level evaluation of some existing Workbench capabilities (e.g., National Data Service, CyVerse, etc..).

While existing “workbench software” takes a variety of forms (e.g. interactive websites, apps, Jupyter notebooks), they tend to share some common features such as providing (1) a single point of access to a large collection of tools and resources, (2) the ability to choose resources and then “compose” them into some type of workflow, (3) the ability to record or “journal” a user’s workflow and then save it for reuse and modification by others and (4) access to remote computation and storage. The third feature is considered important for the reproducibility of scientific results.

Specific Activity: Workbench Exploration. Consider the usability of existing workbench tools for EarthCube. Definition from the EarthCube TAC Resource Registry WG: A *platform* is a composite software entity that enables execution of a variety of tools. A *workbench* is a kind of platform that is user extensible and can support workflow execution.

Paraphrased NSF questions and suggestions that the Tiger Team, and then the LC, will evaluate:

- What is the EarthCube WorkBench, what capabilities will it contain, are there different versions for different user-groups, and what is the rationale for the type chosen?
- What is the value to a geoscientist and how will that be determined?
- Can a workbench be closely tied to cutting-edge scientific research projects?
- How do we establish guidelines for a user-facing workbench, and what is feasible and sustainable?
- What level of resources should be invested?
- What is managed by ESSO vs through projects? Is there a need for pilot projects, and if so are those solicited?
- Stand up Workbench Working Group (WWG) and create a charter for the group working with TAC / SC (*currently underway in March 2018 as a Tiger team led by Sara Graves*)
 - Examine and collect information on possible workbenches and workflow engines
 - Understand the role and technical feasibility of different workflow engines (Pegasus, Kepler, Fireworks, VisTrails)
- Write a Workbench white paper if deemed necessary.