A CRISP Approach for EarthCube:
Collaborative Resource Incubators (CRI) for a Sci-Tech Portfolio

The Collaborative Resource Incubators Working Group:

David K. Arctur, University of Texas Austin

Yolanda Gil, University of Southern California

Anna Kelbert, Oregon State University

Jay Pearlman, University of Colorado Boulder

Ken Rubin, University of Hawai'i

Elisha M. Wood-Charlson, University of Hawai'i

December 11, 2015

This document recommends a new overarching process and fundable organizational units that would move EarthCube forward through the formation of multi-investigator sci-tech efforts called "Collaborative Resource Incubators" (CRIs). These incubators are driven by science-focused cyberinfrastructure needs expressed by the community following the first few years of EarthCube activities which are not well-served by current EarthCube funding mechanisms. CRIs are jointly conceived by scientists and technologists drawing from existing ideas and resources, resulting in a portfolio of possible sci-tech projects that could be undertaken by the EarthCube program. This CRI Sci-tech Portfolio (CRISP) enables a process where projects are prioritized by the community, and leads to proposals submitted to NSF for peer review and potential funding. The overall process described herein allows projects to be long-lived collaborations open to volunteer contributions, and to progress over time from innovative concepts to mature infrastructure solutions. CRIs build on existing EarthCube funded projects and also integrate the extended resources available from the broader geosciences and cyberinfrastructure communities. The proposed CRISP process and its associated CRI funding mechanism represent a novel approach to pursuing science-driven community infrastructure, charting a practical path to address major goals of the EarthCube initiative.

1. Background and Motivation

This document recommends a new process and fundable organizational unit that would help the EarthCube community develop a self-sustaining, long-lived science/technology (sci-tech) collaborative infrastructure that grows organically to respond to the needs of the science community with latest generation technology solutions.

A **Community Proposal.** This document was synthesized by members of the EarthCube Science Committee (SC) and Technology and Architecture Committee (TAC) selected as representatives to the SC's Science Vision Workshop[1] and the TAC Tech Hands Workshop[2] (Spring 2015), and to the subsequent joint SC/TAC Sci-Tech Feasibility workshop[3] (April 2015). As a result, it derives from experiences of the science and cyberinfrastructure community with EarthCube. Feedback on an earlier version of this proposal was collected (Summer 2015) through discussions with the SC and TAC, the EarthCube Leadership Council, the NSF, as well as with the broader community at the 2015 EarthCube All Hands Meeting[4]. This document represents what could be an important component of the mid-term strategy that the EarthCube Leadership Council proposes to the NSF.

Problems Addressed. This recommendation was designed to address fundamental shortcomings unveiled by EarthCube activities to date. Specifically, it was conceived to: 1) accelerate the current rate of progress on exposing the Science community's specific needs for an EarthCube infrastructure, still limited by the ability of most EarthCube scientists and technologists to communicate together on core needs and technologies, by more deeply embedding scientists in the visioning and planning processes; 2) ensure that scientists can experiment with technology as it is being developed so they can affect and have increased ownership of project outcomes rather than being given a final product; 3) enable and deepen sci-tech collaborations for the extended periods of time needed to develop complex infrastructure, from the identification of requirements to prototyping to subsequent integration and deployment of technology; 4) break the currently limited ability of non-funded individuals to participate in EarthCube funded projects by facilitating a more open mechanism for community input on project priorities and for individuals to contribute to projects of interest; 5) provide flexibility to pursue opportunistic synergies and collaborations as they arise through continuous and interacting project activities.

Approach. Our approach is to create a framework that embraces the fact that different geosciences disciplines have vastly different situations with respect to technology and also engagement with EarthCube. Scientific groups with common interests progress from early community formation, to agreements to data repositories and metadata standards, to sharing of software and computing resources, to integration of resources across disciplinary boundaries. While some communities have adopted advanced data management and high-end computing technologies, others are in early stages of organizing themselves and articulating information and technology needs and requirements. While some communities have adopted data sharing practices, others lack community

[1] http://earthcube.org/event/earthcube-scope-vision-workshop
[2] http://earthcube.org/group/tech-hands-meeting
[3] http://earthcube.org/event/sci-tech-workshop
[4] http://earthcube.org/info/about/2015-all-hands-meeting

agreements on repositories or metadata standards. While some are well organized and have defined requirements, others are still discussing what their needs are and learning what technology is available. Our approach is to provide a mechanism to empower geoscience communities at whatever stage they are in this spectrum, give them a context and launching point, learn about technology capabilities available in other areas, and forge the kind of collaborative projects needed to move their community (and others) forward.

Summary of Recommendations. Our core recommendation is a new kind of fundable organizational unit called Collaborative Resource Incubators (CRIs). CRIs would bring together small (4-6 individuals) groups of technology and science PIs (plus student, researchers, early career people, or developers as appropriate) from the science and technology communities within EarthCube to work closely together for a year or two to tackle complex long-term challenges in an agile evolutionary manner. Depending on the focus of the proposed CRI project, this group would be mostly comprised of geoscientists (e.g., to enable community visioning projects), or of computer scientists (e.g., to realize science vision in a specific operational product). As part of the envisioned process, multiple CRIs will be active at any one time and collectively will evolve from germinal science visions to exploratory infrastructure implementations to mature deployments, with each incarnation having a different set of active partners as the focus shifts from what is learned at each stage. We anticipate an accelerated rate of innovation will arise from these close and sustained sci-tech collaborations which have not arisen from existing EarthCube funding mechanisms. We also envision that CRI developments and outcomes will be publicly available and coordinated with the overarching architecture and tool developments. The recommendation also envisions a novel and open process for community-nourished cyberinfrastructure development. This process is designed to facilitate the germination and evolution of CRI projects through a CRI Sci-tech Portfolio (CRISP) that will be explicitly managed by the EarthCube governance. This CRISP process exposes crucial aspects of EarthCube activities and products, offering a structure that would enable the EarthCube community to oversee, enrich, prioritize, and evaluate funded projects and to build on their products.

Relationship with Existing EarthCube Projects. CRIs would be a new type of EarthCube funded project. As of Fall 2015, the EarthCube solicitation[5] has four main types of funded projects for community building and cyberinfrastructure development: Research Coordination Networks (RCN), Building Blocks (BB), Conceptual Designs (CD), and Integrative Activities (IA). Each has different levels of expectations, deliverables, and corresponding funding levels, as described in the EarthCube solicitation. CRIs would be different in encouraging long-term planning and development activities across geosciences disciplines and technology areas, requiring clearly defined mechanisms for community participation, and being expected to enable subsequent CRIs or other EarthCube projects to continue from their work.

Document Overview. The rest of this document gives an overview of the recommended CRISP process, describes the nature of the CRIs, and discusses how they relate to and build on earlier EarthCube efforts.

---

[5] NSF 13-529 solicitation, http://nsf.gov/funding/pgm_summ.jsp?pims_id=504780&org=NSF&from_org=NSF

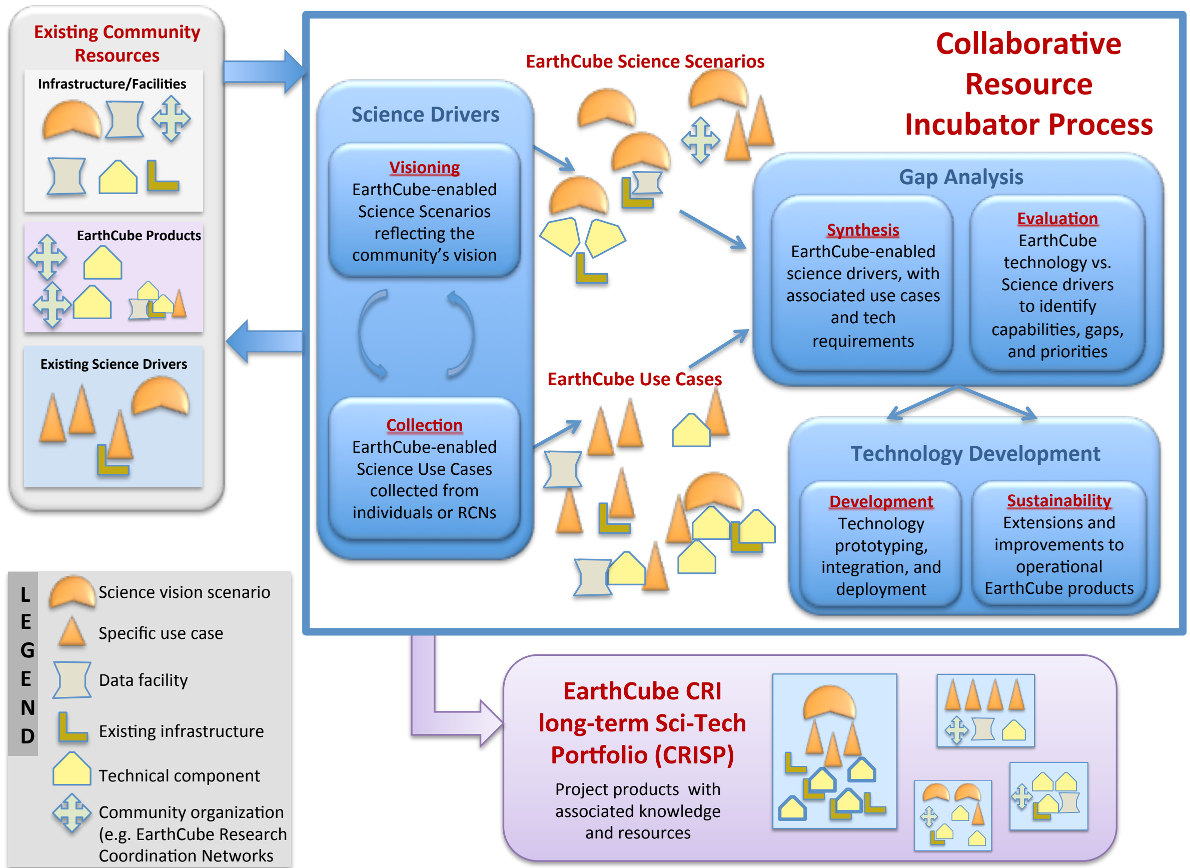

Figure 1. **Collaborative Resource Incubators (CRIs)** will be new funded organizational units with multiple PIs having cross-cutting science and/or technology expertise who will draw on community resources to formulate driving science vision scenarios and use cases, synthesize requirements for prototyping technology components, and integrate and evaluate new capabilities. Multiple CRIs can be active at the same time, each focusing on a different set of cross-domain science problems and technology solutions. All outcomes from these projects will enter the **EarthCube CRI long-term Sci-tech Portfolio (CRISP)**, which will accumulate publically available EarthCube products at all stages of maturity. CRISP products are envisioned to drive and inform further CRI proposals as well as proposals to other EarthCube funding mechanisms.

2. Recommended Process: Collaborative Resource Incubators for a Sci-tech Portfolio (CRISP)

Figure 1 gives a diagrammatic overview of the recommended process and its key activities, which are described in detail in the rest of this section. This process would move science goals forward over a long period of time through a series of short-lived activities formulated as CRI projects, whose products will generate an evolving EarthCube CRI long-term Sci-tech Portfolio (CRISP), or EarthCube Portfolio for short, that articulates over time the science goals, current state, available resources, identified gaps, and prospects for technology development for a science community.

The figure depicts a proposed organic, self-sustaining process for EarthCube community development. Key components of this process are **Collaborative Resource Incubators (CRIs)**. CRIs are envisioned as new funded organizational units with multiple PIs with cross-cutting science and/or technology expertise. A major component of the CRI process is the ability to fund groups of scientists and technologist to conduct science-driven planning and implementation (a first for EarthCube), for instance to formulate technology-savvy science scenarios and synthesize them with technology requirements. CRIs will draw on community resources to formulate driving science vision scenarios and use cases, synthesize requirements for prototyping technology components, and integrate and evaluate new capabilities. These resources include products of other EarthCube projects and of other NSF programs, data and software from existing facilities and repositories. CRI project outcomes will enter the publically available **EarthCube CRI long-term Sci-tech Portfolio (CRISP)**, which will accumulate EarthCube products at all stages of maturity that are envisioned to drive and inform further CRI proposal submissions and/or submissions to other EarthCube solicitations and NSF-CISE opportunities. The process would be jumpstarted and continuously guided by the outcomes of Science Driver CRI projects; outcomes from these projects would enter the CRISP in the shape of formal technical requirements. All proposal submissions to the CRI program would be prioritized through the standard NSF peer-review and budget allocation procedures. This primary feedback loop is depicted with orange arrows in the Figure.

An additional input to the CRISP would come from the pool of existing resources, which includes all data and other infrastructure facilities, existing science drivers, along with all other EarthCube products from BB and IA activities. These products will enter the long-term EarthCube portfolio after an EarthCube leadership review and prioritization, in a secondary feedback loop depicted with blue arrows. These resources will also continue to inform future CRI project developments. The ultimate goal of the proposed strategy is that all CRI efforts will contribute to, reuse, develop and maintain previous projects' outcomes found in the long-term EarthCube product portfolio.

The CRI process and CRISP will be most effective if it is informed by, and receives input from, the EarthCube community, EarthCube governance, and NSF program managers. Rather than establish a formal mechanism for oversight, we propose that the CRI projects be aligned with EarthCube strategies and community needs by the PIs and that the visioning and prioritization for these projects be evaluated as part of the normal NSF peer review process. In this regard, CRI project proposals will be based on community-driven priorities. EarthCube governance would provide feedback into CRI activities in much the same way as it does to other EarthCube activities (e.g., via discussions and feedback on presentations from CRI researchers to the relevant governance committees).

Collaborative Resource Incubators (CRIs)

The CRISP process would enable the engagement of scientists with a shared interest, ideally from different geoscience communities. Initially, *visioning* efforts would be carried out and result in general science scenarios, leading to identification of existing resources as well as technology gaps and needs. Use case *collection* projects could come from communities or from individual investigators by articulating specific scientific research scenarios that would spell out the roles of scientists and technology components. Activities could also focus on mapping specific use cases to science vision scenarios, annexing existing community resources, and identifying general technology gaps and requirements. *Synthesis* projects would then focus on the collaborative (sci-tech) development of requirements and formulation of new technology components as appropriate, with corresponding *evaluation* projects focused on testing and assessment of existing capabilities and gaps. Finally, *development* projects would focus on the development of technology components, while other projects would focus on *sustainability* of deployed technology components. Examples of science goals and a possible CRISP process to advance them are given below.

CRI projects will draw from community resources including products of prior EarthCube projects, infrastructure from other NSF programs (e.g., DIBBs[6]), data facilities, and broader resources from other funding sources (e.g., government agencies, private-sector solutions and international resources).

CRIs would involve at any given time a small set of groups of lead PIs (4-6 individuals) in both science and technology. The size of the projects will depend on the community and what the activities require. As CRIs evolve through the process, they may involve a different set of active partners as the focus shifts at each stage. This will provide flexibility and agility to address the challenges at each stage. CRI projects will have specific requirements for involving the community, particularly early career researchers, as described below. We envision some "keystone" science participants who will remain involved throughout the process to see the technologies developed actually transition to science practice.

The results of these multiple CRI activities and outcomes will form the EarthCube Portfolio (CRISP), which would embrace both germinal and more mature projects. The portfolio would consist of an evolving set of project outcomes at various stages of refinement and with associated community resources, which would be visible to the community and promote direct engagement by domain scientists and technologists. Additional inputs to the portfolio could come from the pool of existing EarthCube and other resources, through the process of review and prioritization by the EarthCube community and leadership.

Science Drivers

Science scenarios and use cases capture specific contexts where technology can be used in science research, and lead to requirements that drive the development of technology capabilities. Therefore, they play a crucial role in facilitating communication and fruitful interactions between computer scientists and geoscientists. Yet to date they have been difficult to develop for much of the geoscience community that will be served by the eventual EarthCube technology, largely because scientists tend to not think of their day-to-

[6] NSF DIBBs solicitation: http://www.nsf.gov/pubs/2014/nsf14530/nsf14530.htm

day work and their future infrastructure needs in this manner, making it difficult to develop rich and specific scientist-informed use cases. Results of several recent EarthCube workshops where technologists and scientists discussed this situation indicate that both groups feel a guided framework for facilitating such interactions, informed jointly by the science goals and by a thorough understanding of technological capabilities, is needed to direct EarthCube in its efforts to support original geoscience research. It is the long-term nitty-gritty back and forth interaction between scientists and technologists that CRIs will foster which will result in the most useful, innovative and well-used future technologies from the EarthCube project.

Recognizing this, EarthCube should adopt a two-pronged strategy for stimulating and driving its cyberinfrastructure development:

(1) A domain-driven visioning effort that would assemble science vision scenarios that have emerged as a community consensus from the End User Workshops, Research Coordination Networks (RCNs), and other geoscience community activities. These science scenarios or 'generalized' use cases would elaborate the vision of the EarthCube Strategic Science Plan Geoscience 2020[7] (§5). These scenarios would articulate science goals, highlight technology requirements, and point to existing resources (e.g., an existing data facility or an organized community) relevant to achieving those goals.
(2) A user-driven/computer-scientist-aided collection effort that would focus on documenting use cases contributed by individuals or a coherent group of scientists (e.g., RCNs) across a wide spectrum of geoscience domains, whose research agendas would be advanced through specific integrative science and (existing or to yet be developed) technology and resources.

This dual approach would result in broad science scenarios and specific science use cases that could then be mapped, through a synthesis activity, to identify relevant technology capabilities, organize communities around synergistic topics, elaborate commonalities across scenarios and cases, and annex or develop unique resources as appropriate.

The challenges of communicating technology needs across the computer scientist-geoscientist divide, the person-hour intensive nature of this effort when scaled up to a large number of use cases, and the difficulties of doing this within the current EarthCube funding vehicles suggest that a new type of organizational unit and funding mechanism for these activities is necessary. In fact without the CRISP process, the community is at risk of not having a sufficiently broad and representative group of multi- and cross-domain use cases and workflows to realize the full range of expectations for the eventual EarthCube infrastructure. A CRI would also effectively connect with a broader range of users to better address community-wide opportunities and challenges. Much of the CRI support would be in the form of salary for participants, with a smaller amount for travel to support group work and/or small products-focused workshops.

[7] http://earthcube.org/document/2015/earthcube-strategic-science-plan

Furthermore, the involvement of scientists should not stop once science scenarios and use cases have been formulated. It is crucial that the technology development activities be open and supported to encourage participation and experimentation by interested communities and individuals at all stages. In addition, there must be room for different evolution paths for the projects initially conceived, in some cases going through cycles of redesign and exploration of other solutions, in other cases progressing to integration and maturing into eventual deployment. This evolution takes time and therefore project participants must see a prospect for long-term collaboration and funding.

Evolution of CRIs and Accomplishment of Science Goals

To illustrate how science communities can progress through the CRISP process, we show three examples in the side boxes: omics, field science, and advanced cross-disciplinary data-model integration for predictive science. We discuss how communities around these themes have evolved so far through EarthCube activities, and how their arc can be continued through the proposed process.

Example: Omics

A nascent omics community was formed in EarthCube through an Ocean Omics End User Workshop held on August 2013. This initial meeting focused on the community's Visioning of science scenarios. This resulted in a workshop report and a publication to disseminate the results to the community (Gilbert et al 2014). These publications describe several critical science scenarios, and they also highlight many of the community's cyberinfrastructure constraints and suggestions next steps.

One of those recommended steps was to form an EarthCube Research Coordination Network (RCN). The "EarthCube Oceanography and Geobiology Environmental Omics" (ECOGEO) RCN – funded in August 2014 – is active in EarthCube governance and recently hosted their first workshop aimed at streamlining the community's science scenarios and improving their collection of use cases. The ECOGEO workshop report (available on the ECOGEO EarthCube website) summarizes outcomes from these aims, but also contains an extensive discussion on evaluation of technology versus the science drivers. These sections of the report include an in-depth discussion of existing capabilities (many of which are external to EarthCube), highlights several key resource/cyberinfrastructure gaps, and identifies the immediate, science-driven priorities. This RCN is essentially carrying out the Use Case Collection and Gap Analysis activities outlined in the CRI process.

During the final year of the RCN, ECOGEO will focus on improving the 12 use cases from the first workshop. ECOGEO will work with EarthCube's TAC WG on Use Cases to work on a Synthesis activity to generate technology requirements. The second workshop planned will focus on community engagement and training. At that point, the ocean omics community will have progressed to the next stage of maturity, and will be ready for the next phase of technology development to address identified gaps and to consider overlaps and differences with use cases, work flows, and technology solutions available to other ecological omics groups, such as Critical Zone observatories and in other terrestrial ecosystems. For instance, the CRISP process would enable crucial activities to grow technology solutions beyond ocean omics to merge with those of aligned research communities in geochemistry, microbiology and paleobiology.

[Gilbert et al 2014] "Meeting report: Ocean 'omics science, technology and cyberinfrastructure: current challenges and future requirements (August 20-23, 2013)." Jack A Gilbert, Gregory J. Dick, Bethany Jenkins, John Heidelberg, Eric Allen, Katherine R. M. Mackey, and Edward F. DeLong. Standards in Genomic Sciences. 9(3): 1252–1258, 15 June 2014. doi: 10.4056/sigs.5749944.

Example: Field Science

The EarthCube RCN "Earth-Centered Communication for Cyberinfrastructure — Challenges of Field Data Collection, Management, and Integration" was founded to engage geoscientists who collect field data in a variety of settings and for a variety of uses. The RCN has held two field trips with diverse participation including geoscience, computer science, information science, linguistics, and social sciences. On the geosciences side, they included sedimentologists, petrologists, structural geologists, and paleontologists, leading to a variety of use cases and requirements. These field trips resulted in a Visioning activity that led to a better understanding of the requirements for collecting data in the field, and for sharing data not only about samples and field observations but also about the laboratory and computational analysis products that are derived from the samples and field measurements once back in the lab.

An article about this Visioning activity reported on the requirements for field data collection tools, and an initial compendium of metadata properties for the samples collected. The CRISP process would enable this community to expose the requirements identified to technologists and engage them for Collection of specific use cases, and for Synthesis and Evaluation activities to identify relevant existing technologies and gap analysis. These activities would lead to coordination with the iSamples RCN, which is concerned with sample repositories and sample metadata and is producing complementary requirements that should be consolidated.

[Mookerjee et al 2015] "Field Data Management: Integrating Cyberscience and Geoscience." Matty Mookerjee, Daniel Vieira, Marjorie A. Chan, Yolanda Gil, Terry L. Pavlis, Frank S. Spear, and Basil Tikoff. EOS Letters, 96, 2015. Published on 13 October 2015. doi:10.1029/2015EO036703.

Example: Cross-Disciplinary Data-Model Integration for Predictive Science

One of the major challenges of integrating data and model predictions is to fit the resolution of data sampling to the resolution of the models available. Available data sets are collected at specific spatial and/or temporal scales, and they are often different from the scales that model developers work with, including spatial/temporal undersampling or variable sampling density. These issues become significantly more complex when attempting to merge data and models across different science domains to develop next-generation Earth systems understanding and cross domain predictive capability. Examples include seismology, ecology and petrology. All of these domains have access to some well-developed cyberinfrastructure, and yet share general scale mismatch issues. This is a very common theme in the End-User Doman workshop reports, which could be considered initial Visioning activities.

A CRI could be formed to do Use Case Collection and Synthesis on seamless model-data integration across wide spatial scales, sampling density, and science domains. Working on the same research topics, this group could make significant progress on outlining the problems and suggesting solutions for computing environments to provide model and data set interoperability for complex system description and predictions. For example, the seafloor volcano community is concerned with links between mantle structure/composition, magma generation, eruption size and frequency, mass and energy flux between the deep earth and the deep ocean, the impacts of these fluxes on microbial and macrofaunal ecology, and long term impacts on ocean chemistry ("mantle to microbe") [Rubin and Fornari 2011]. The limitations of sampling rate and sampling density for model predictions in one of the domains in this community (mantle geochemistry as deduced from erupted lava compositions) are described in [Rubin et al 2009].

The CRISP process would enable the formulation of projects involving experimental scientists and model developers across-domains together with technologists at a deep level, enabling them to formulate a new community based on shared science goals, and moving from Science Drivers to Synthesis and Gap Analysis to Technology Development activities via the CRISP process.

[Rubin and D.J. Fornari, 2011] "Multidisciplinary Collaborations in Mid-Ocean Ridge Research", EOS, Trans. Amer. Geophys. U., 92:16, 141-142, DOI: 10.1029/2011E0170002."

[Rubin et al., 2009] "Magmatic filtering of mantle compositions at mid-ocean ridge volcanoes", Nature Geoscience, 2, 321-328.

CRI Project Requirements

The community should have visibility into the CRI projects and weigh in the EarthCube Portfolio. This requires that projects have mechanisms to include the community. Specific mechanisms for community involvement should be required for CRIs, such as:

1. A plan for public and periodic checkpoints of progress, each checkpoint consisting of a report of how community input was incorporated in the work, and clear requests for further community inputs and creative forms for community participation in the project
2. A mechanism for scientists to become early adopters to test new capabilities and provide early feedback. Ideally, the CRI project would support such participation through funding (e.g., 1 month salary for faculty), and give priority to early career researchers when appropriate.

CRI projects should also be required to publish all their products in ways that are useful to the community and can be incorporated in the EarthCube Portfolio. In addition, CRI projects should be required to write specific articles targeting dissemination of results to the community (along the lines of the articles cited in the boxes above), to organize town halls, and to coordinate community involvement through the Engagement Team and the Liaison Team.

Relationship with EarthCube Governance

It is crucial that the CRI projects have a synergistic connection to EarthCube governance as a practical means to ensure a strong connection with the community's goals, priorities, and resources. CRIs would be asked to contribute to governance specifically by participating in committee activities, and by reporting their activities to committees for review and feedback. EarthCube governance committees would suggest synergies and collaborations among projects, and ensure that all the results of interest of the CRI projects are appropriately incorporated into the EarthCube Portfolio. EarthCube governance would also have the authority to add to the EarthCube Portfolio from the pool of existing non-CRI resources, to ensure maximum reuse of products and infrastructure that are close to maturity and have community buy in, and to allow EarthCube to build upon these through CRISP-driven proposal submission and the NSF peer-review process.

These close interactions with the EarthCube governance structures will ensure that: 1) the CRIs are guided by and benefit from community input, 2) the results of CRI projects are captured as community assets, 3) there is necessary coordination among projects, and 4) there is interlinking to other NSF programs and project results.

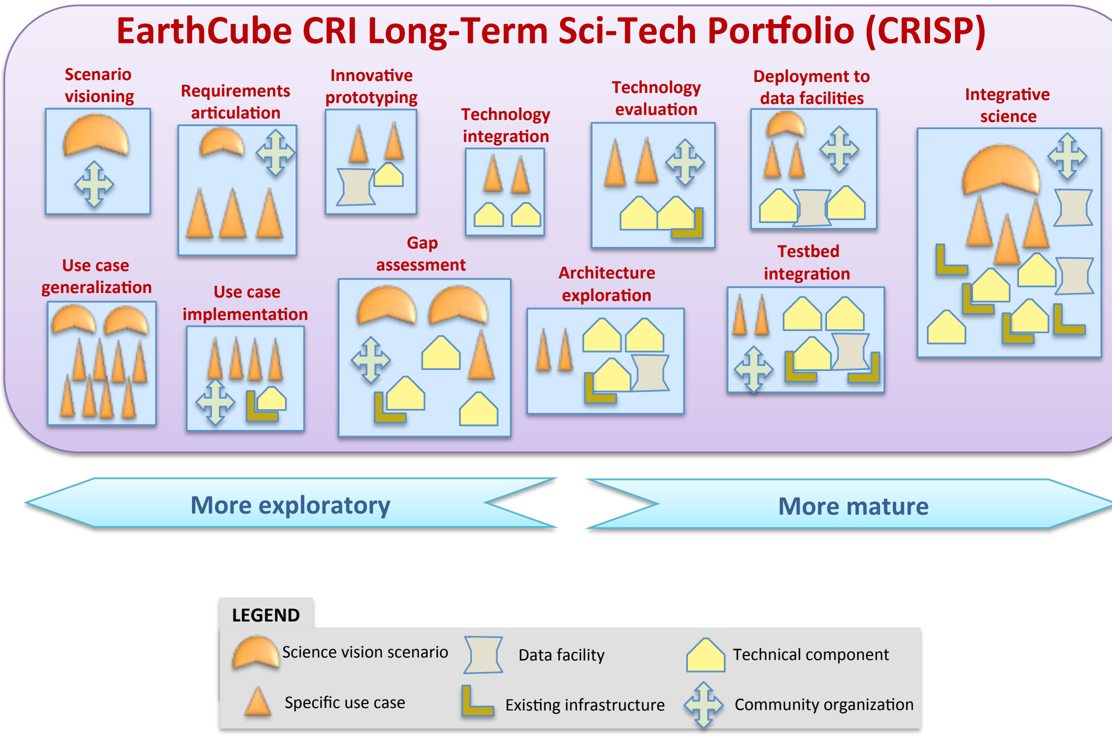

Figure 2. The **EarthCube Portfolio** will include all CRI products, as well as selected non-CRI resources of interest to the community. It will be a compilation of drivers, ideas and solutions to pressing geoscience research questions. The Portfolio will evolve and grow as CRI projects continue to be formulated to advance their driving science and research goals. The EarthCube Portfolio is envisioned to contain resources of varying degree of maturity and complexity. It should be organized and managed as a library that is accessible to the community for use and comment.

3. The EarthCube Portfolio

The EarthCube Portfolio would consist of an evolving set of project outputs and associated resources that would be made available to the community, promoting direct engagement by domain scientists and technologists throughout the process. From the perspective of CRIs, the Portfolio would include products from the activities of individual CRIs (contributed as community resources) and other approved resources, which would all be accessible to the community. Contribution to the EarthCube Portfolio would be a required deliverable of all CRI projects. Where reasonable, new CRIs would be required to build upon the resources already in the Portfolio.

Types of CRI Products in the EarthCube Portfolio

A wide range of project types is envisioned. For example, germinal activities can result in modest resources (e.g., a framework document meant to jumpstart a discussion) that can help catalyze a community of interest. On the other end of the maturity spectrum, the functional portfolio resources that are already associated with science users communities could be adopted for maintenance by the NSF or other funding sources, to ensure sustainability. Therefore, the EarthCube Portfolio would include a heterogeneous range of science-driven project ideas and functional technology with different composition and degree of maturity, implementation, and interdisciplinary work.

Figure 2 illustrates a diversity of potential CRI products in the EarthCube Portfolio:

1. Scenario visioning: These could include a science vision scenario with several specific use cases that can be mapped to that driving scenario. Initially, they may or may not include technology components, but could include a specification of these components.
2. Use case challenge problems: These would specify the roles of scientists and technology components in solving a science problem, and could include datasets and challenge science questions to be addressed when tackling the use case.
3. Technology gap and requirement identification: Joint projects with scientists and technology developers would lead to identifying technology gaps and requirements based on science use cases or vision scenarios.
4. Implemented science use cases: Implemented technology components resulting from direct interaction between computer and domain scientists, initially embedding full technological specifications driven by use cases and eventually including general reusable implementations of those specifications.
5. Tech explorations and transitions to data facilities: Of particular interest would be new technologies driven by and eventually deployed in existing data facilities, with synergistic science requirements and a clear path to sustainability.
6. Technology integration opportunities: These could be specifications of integrated feasibility demonstrations, testbed experiments, or integrations of mature systems.
7. Exploration of architecture concepts: Integrations of selected technology capabilities and relevant standards to experiment with specific architecture concepts and test them in a science use context.
8. Innovative prototyping: An environment to explore new ideas, facilitating rapid prototyping and the collection of early feedback from scientists.
9. Science-driven technology integration: Integration of existing technology components to address science vision scenarios and use cases.

   Commercial partnerships: Specifications or developed technology components that involve industry collaborations.

The EarthCube Portfolio would evolve over time in focus areas and maturity levels, reflecting the community's interests as they shift due to opportunities, science priorities, and technology evolution.

Managing the EarthCube Portfolio

Managing the CRISP process and the resulting EarthCube Portfolio would require infrastructure components. This new infrastructure would allow the EarthCube governance units to involve the community in managing the EarthCube Portfolio. Specific capabilities that will be needed include:

- A long-term persistent repository for the collected CRISP products
- An organized metadata-based repository of past and ongoing CRI project characteristics, including participants, status, and products
- Community comment mechanism to provide early feedback
- Individual subscriptions to specific projects, ranging from high-level periodic updates to receiving daily details
- An IP framework that supports collaborative project work and competitive funding proposals, and that clarifies what kinds of inputs each project is willing to accept and what recognition (monetary, credit, or other) will be given to contributors (like the IP frameworks that are used in OGC, W3C, and other bodies)

4. Building on Prior and Ongoing EarthCube Activities

CRIs would build on the products of past and ongoing EarthCube funded projects, but would have fundamental differences.

Differences with Current EarthCube Funding Mechanisms

CRIs are envisioned to differ from other EarthCube funded entities in the following ways, using NSF vocabulary paraphrased from the various NSF13529 Program Solicitation versions[5]:

- **Differences with EarthCube Research Coordination Networks (RCNs).** RCNs are intended to "advance geosciences cyberinfrastructure through interaction, discussion and planning between geoscientists and cyberinfrastructure experts." RCNs are therefore primarily scientist planning networks. CRIs differ primarily by expectation of the depth of interaction between technologists and scientists and the products that articulate next steps in the CRISP process.
- **Differences with EarthCube Integrated Activities (IAs).** The goal of the EarthCube Integrative Activities is to "enable geoscientists to participate in EarthCube. Projects are intended to improve access to the products of geosciences research so that a broader array of geosciences communities may help shape future EarthCube activities and outcomes." IA projects are used primarily to build up new capability from existing infrastructure, whereas CRIs are for planning new capabilities that address specific categories of technology solutions based on emerging science community needs.
- **Differences with EarthCube Building Blocks (BBs).** The objective of Building Blocks is to "contribute to the EarthCube enterprise by developing, integrating and utilizing current cyberinfrastructure in coordinated efforts to serve a broad segment of the academic geosciences community by contributing increased capabilities to through the integration of existing technology components; creation or modification of cyberinfrastructure to overcome barriers or inefficiencies as identified by the geoscientists, and development of essential components of an EarthCube

framework." BBs are focused efforts to develop a specific technology to solve one or more fairly narrowly defined science needs, largely driven by technologists, with scientist participation. CRIs are envisioned to plan for development of more complex technology solutions to a broader range of immediate scientist needs using more direct and continual scientist-technologist interactions, and to promote early career scientist participation.

- **Differences with EarthCube Conceptual Designs (CDs).** The objective of Conceptual Designs is to develop "the initial enterprise architecture design of EarthCube, including the set of structures needed to serve the data and information needs of the geosciences community". While the goal of CD projects is to produce documents with EarthCube architectural design criteria, CRIs are meant for projects that explore particular aspects of architecture or architecture concepts that address some clearly identified requirements from previous Visioning, Collection, Synthesis, or Evaluation CRIs.

CRIs would build upon and integrate the products of current and future BB, IA, CD and RCN projects to provide integrated technology solutions. A CRI might be thought of as generally falling between IA and BB in terms of expected deliverables and funding level, with a similar to perhaps shorter project duration, but the CRI mechanism should be flexible enough to the needs of specific aims of the CRIs, the domains they serve, and the technologies they are planning for.

Fundamentally, CRIs would be considered part of an overall process that encourages community participation.

Drawing from EarthCube End User Workshops and RCNs

From mid-2012 through 2013, EarthCube sponsored 24 End User Workshops targeting a broad spectrum of Earth, atmosphere, ocean, and other geoscience-related research domains. These workshops spanned a wide range of geoscience domains with various levels of pre-existing coordination, community cohesion, and cyber-resources[8]. The broad goals and aspirations of these communities were distilled by the EarthCube Science Committee into the EarthCube Science Strategic Plan in Spring of 2015 ("Geoscience2020"). Recently (e.g., at the 2015 All Hands Meeting, as well as several TAC and/or Science Committee workshops in the months before it) it became clear to a significant number of highly engaged EarthCube participants that this Plan, as well as the workshop reports themselves, lack the specificity required to develop actionable use cases for technology development, which in part drives our perceived need for CRIs.

One additional outcome of those workshops was the development of several domain specific Research Coordination Networks (RCNs) designed to engage a broader research community and explore their needs for technology and cyberinfrastructure, specifically highlighting gaps that hinder current research efforts. RCNs are funded to host two

[8] The combined executive summaries of EarthCube End-User Workshops can be found at http://earthcube.org/sites/default/files/doc-repository/CombinedSummaries_12Dec2014.pdf

community workshops to crowdsource ideas and develop plans to address the identified technology and cyberinfrastructure challenges. Some of these challenges may already be addressed through existing EarthCube cyberinfrastructure. However, several RCNs, such as ECOGEO (environmental 'omics), represent scientists, technologists, or research topics that do not fit into traditional geoscience frameworks. Therefore these RCNs have unique domains needs, specific use-cases, and science scenarios that require an additional CI resource or bridge be constructed before EarthCube can begin to enable their science.

A CRI would be an environment and vehicle to enable the end-user workshop communities, other science domains not represented by RCNs, and the aforementioned RCNs that are not traditionally "geoscience" research domains to build resources or designs that allow a community to overcome identified gaps in their science research and better leverage current/planned EarthCube resources.

Furthermore, CRIs can provide a working environment for new opportunities in crosscutting research themes. The EarthCube End User Workshops were generally focused on specific disciplines/domains, yet many of the advances needed to address grand science challenges (see the 2015 EarthCube Strategic Science Plan), such as climate change, are highly cross-disciplinary. Looking to the future, facilitating more complex use cases across disciplines and having an environment for test and validation is a natural strength of CRIs for geoscience research.

Drawing from EarthCube Technology and Architecture Foundations

As EarthCube's initial funded projects progress, they will result in technology components (e.g., the EarthCube Building Blocks), architecture concepts (e.g., the EarthCube Conceptual Designs), initial integrations of technical capabilities, and other technological advances. They will provide components that can be incorporated into the EarthCube Portfolio.

CRIs will help facilitate the mission of the EarthCube Technology and Architecture Committee by providing a funding framework that will enable: 1) coordination between technology components and scientific requirements, 2) identification of technology gaps, 3) development of testbeds for evaluating technology components, 4) assessment of standards and other architecture concepts.

CRIs would also provide a framework where the geosciences and the cyberinfrastructure experts can work closely together on long-term joint research initiatives, which would significantly extend the value of EarthCube to the broader community.

5. Benefits of Maintaining an EarthCube Portfolio

The benefits of maintaining an evolving EarthCube Portfolio include:

- Science communities with different degrees of technology adoption can engage EarthCube at different stages of the proposed process

- Scientists can relate to the scenarios and use cases and seek technology partners to formulate projects
- Individual scientists can map their needs to use cases and scenarios, vote on their prioritization and be involved in proposing solutions
- A flexible and inclusive environment for exploring new research directions
- A sandbox for scientists and technologists to understand, brainstorm, and explore ideas
- An open, collaborative setting that would encourage scientists in smaller facilities to reach out and get involved in manageable and meaningful ways
- Technology developers can analyze the scenarios and use cases and propose collaborations with scientists that developed them
- NSF can use the EarthCube Portfolio as an indication of priorities and consensus approaches representative of what the community wants to pursue and what it needs to build on
- Other agencies, foundations, and institutions can volunteer funding or resources to pursue particular projects in the EarthCube Portfolio
- Identification of well-developed and used technology components could be adopted for maintenance, promoting sustainability
- Outreach to other scientists who would be able to understand what EarthCube can do for them by looking at the developed and demonstrated science scenarios and use cases
- Inform government observers of what is happening in EarthCube in terms of advanced cyberinfrastructure capabilities and their science impact
- Other government agencies could browse through the EarthCube Portfolio and decide to fund projects of interest
- Leverage other funding as individual researchers can align their funded projects from other sources with EarthCube and explore synergies, pursue integration efforts, and make contributions to EarthCube
- Commercial partners can be engaged to address specific science needs that are synergistic with industry interests

The CRI approach would have two novel research contributions of potentially transformative impact on cyberinfrastructure:

- A novel process and underlying framework for community-driven cyberinfrastructure development with mechanisms for project evolution, prioritization, and sustainability
- A novel life-long learning environment where EarthCube Portfolio resources would be used for hands-on demonstration of cutting-edge synergistic technical and scientific advances that could be made accessible to researchers, students, decision makers, and the public

Acknowledgments

This document is an outgrowth of discussions that began at the EarthCube Sci-Tech Feasibility Workshop held on April 23-24, 2015 and has been enhanced by the input and

comments of the EarthCube community. Workshop participants were Tim Ahern, David Arctur, Ruth Duerr, Dave Fulker, Yolanda Gil, Basil Gomez, Tom Guilderson, Anna Kelbert, Danie Kinkade, Kim Miller, Jay Pearlman, Steve Richard, Kathleen Ritterbush, Ken Rubin, Durelle Scott, Sarah Stamps, Elisha Wood-Charlson, and Ilya Zaslavsky.

Appendix: Document History

The first version of this document[9] was released to the community for comments on May 13, 2015, and was presented and discussed at the EarthCube 2015 All Hands Meeting. The current version addresses major points of feedback received, including:

1. The specific community needs for this new type of project and the distinctions with existing EarthCube funding mechanisms have been clarified and expanded.
2. The terminology has been changed to "collaborative resource incubators". The original document used the term "collaboratories", which already has a specific meaning (e.g., the iPlant collaboratory).
3. The document includes a summary of its history, how it was generated by the community, and where it is intended to go.
4. The role of the Data Facilities has been expanded, so they are clearly identified as partners in the development of initial use cases and concepts all the way to transition mechanisms.
5. The document clarifies how the concepts proposed build on earlier EarthCube End User Workshops.
6. The kinds of projects that would involve technology components have been fleshed out in more detail.
7. The relationship of the proposed approach and the governance of EarthCube has been clarified, in particular what would be the oversight by the current governance units.
8. The size of the envisioned CRI projects is now explicitly mentioned.
9. The role of EarthCube governance has been clarified.
10. Examples of CRIs based on communities currently engaged in EarthCube have been included.
11. Requirements for CRI projects have been outlined.

[9] http://earthcube.org/document/2015/earthcube-collaboratories