EarthCube Advisory Committee Report
March 11, 2016

EarthCube (EC) began in 2011 as a decade-long program to develop cyberinfrastructure for the academic geosciences community and to help address the scientific challenges for studying the Earth System. The EarthCube Advisory Committee (AC) has been asked to guide the development of EarthCube by providing external scientific advice to the EarthCube Leadership Council (LC) as well as NSF Geosciences management. The formal mechanism for providing this advice is through this committee report to NSF program managers, which we view as necessary for their strategic planning, in addition to providing advice to be shared directly with the LC and other appropriate parties. The goal of this report is to provide actionable recommendations to the LC and the NSF that will enable the EarthCube program to move forward and begin to thrive over the next 5 years.

The AC held its inaugural meeting at the NSF on February 4, 2016 at a reverse site visit (RSV) attended by members of the EarthCube Leadership Council and Support Office together with relevant NSF personnel from GEO directorate and CISE/ACI. In person attendees at the meeting are listed in an attachment to this report, along with the detailed charge to the AC. Additional members of the LC and other EarthCube Committees participated by teleconference at various stages of our discussions.

The AC was presented with 4 broad questions involving review of the Strategic Vision Document, reviewing the Roadmap Document, evaluation of Community Engagement, and recommending evaluation of next steps towards Program Sustainability. Presentations and discussions with the LC at the Feb 4 meeting helped inform a series of detailed findings and recommendations connected to the charge, presented in Sections 2 through 6 below. These findings collectively point to fundamental concerns with the trajectory of this program, summarized in Section 1.

1. GENERAL CONCERNS

The AC notes that the LC has demonstrated an exceptional level of volunteer commitment to the success of NSF's EarthCube program. We recognize the cultural, social, and scientific challenges faced by the LC in attempting to promote a grass roots movement enabling the geoscience community to move towards a new data-enabled cross-disciplinary research paradigm. Over the past several years, steps have been taken to establish a governance model, populate the EC committees, and begin the work for EarthCube. The large number of community workshops and efforts to build consensus on community goals has clearly strained available intellectual and time resources. In weighing these struggles against accomplishments from the first 5 years, **it is evident that significant structural changes are needed to create a thriving sustainable program within the next 5 years.**

The AC identified three interrelated problems impeding progress in the EarthCube initiative. The first is the persistent lack of clear definition of "EarthCube." Given the time and financial resources invested, there should be no ambiguity about program definition or objectives at this time. The second impediment is the reliance on a largely voluntary LC and a fully voluntary governance structure. Given the lack of clear program definition, self-selected stakeholders are

free to tailor their understanding of EarthCube as they see fit, precluding the development of a common vision.   The third impediment is the lack of interconnection between LC activities, proposal solicitations, and assessment of outcomes from funded projects. In essence, the LC and the EarthCube PIs appear to be operating "open loop," except where interactions are mediated by the NSF.   This, again, is a severe impediment to progress toward a common vision.

Based on these concerns, the AC recommends that a full EarthCube program review be conducted, which may result in a revamping of program objectives, governance structures, and implementation plan.   More information on these concerns can be found in the detailed findings and recommendations presented in Sections 2 through 6.

2. STRATEGIC VISION

Charge: How well does the strategic vision represent the scientific and technological goals of the geosciences community?   Does it present an appropriate framework to guide future EarthCube community activities?

There was general agreement on the panel that the strategic vision document articulated suitable high-level scientific and technological goals. However, it falls short in establishing a framework for guiding future EC activities to accomplish these goals. Our perception is that the vision suffers from a conflict between the idea that one part of EarthCube is a voyage for the GEO community into a new scientific culture, vision, and workflow models while another part is perceived as a results oriented structured project that is in need of succinct implementation plans. Our overall discussion led to the following series of Findings and Recommendations in this section.

2.1 Finding: The definition of EarthCube remains unclear to the GEO community.
The question "what is Earthcube?" remains open, despite the 5 years elapsed since the first charrette where this was discussed. The AC felt that this lack of definition has two causes.   The first is the extremely broad scope of EarthCube, with many disconnected parties pulling the program in different domain-specific directions. The second cause is a pervasive community view that EarthCube is an infrastructure to be implemented. In reality, we see it as a 'program.' with two interrelated elements: innovation and production. The program should continue to foster innovative exploration around the explosive field of data science, while simultaneously considering ways to leverage the most applicable results for advancing geoscience research. Innovation is a continuing and evolving process with no specific destination, but continued justification of the program rests on the ability to implement a set of tools that meaningful enhance the productivity and scope of research.

Recommendation 1: Earthcube solicitations should be written to attract a broad base of innovative thinkers that cross the disciplines of data science and geoscience. EarthCube is a program within CIF21, and yet the data science component of the program is underemphasized. Solicitations should also be written to target the challenge of interoperability across traditionally disparate data collection methods and repositories. The first solicitation category should invite "blue sky" thinking, the second requires decisions that may exclude some players.

Recommendation 2: Articulate a concise and clear definition of EarthCube that embraces the innovation aspects of the program as well as the concrete objective of creating usable infrastructure.   The program definition needs to embrace discovery in both data science and

geoscience. At the same time a vision and suitable metrics are needed for evaluating research products and leveraging the best results to develop enabling cyberinfrastructure.

2.2 Finding: Transition to an LC-empowered governance model
The initial 5-year phase of EarthCube (which we will refer to as "Phase I") has involved development of a community-based self-selected governance model. Its evolution from "grass roots" has attracted a large base of impassioned participants who have laid a valuable foundation through end-user workshops, reports, and sponsored research projects (in particular, the Building Blocks projects). However, this federated self-selective governance model is unlikely to be able to achieve consensus or be sufficiently representative to push the program forward.

Recommendation: We recommend an expedited shift in governance strategy to include a substantial top-down component. It is appropriate to consider this new model as part of "Phase II" of the program and it should be a near term goal. We recommend that the LC be empowered to make critical decisions about architecture implementation and become directly involved in defining proposal solicitations from NSF.

Recommendation: In the long term, an entirely different governance model may be needed that involves a selected set of senior stakeholders in each discipline area, who would be empowered to lead capacity building efforts within their discipline while also facilitating coordination with other discipline leaders. The UNOLS program was cited as a potential organizational template for EarthCube, and there may be other models worth considering. This issue is revisited in Section 4.

2.3 Finding: "System of systems" is a critical concept in the EarthCube program, but it is not clearly defined.

Recommendation: If "System of Systems" represents the guiding paradigm for realization of the kind of data-enable science advanced by EarthCube, then the phrase must be defined in a way that lends itself to an implementation strategy. The AC understands the phrase to refer to the aggregation of multiple heterogeneous data repositories and tool sets under a single EarthCube system. Interoperability is the cornerstone objective in this approach. The LC is free to correct this understanding if it is incomplete or in error.

2.4 Finding: The competing "innovation" and "production" elements of EarthCube need better articulation to alleviate the ambiguity and lack of realistic program scope sensed by the Geoscience community.
The strategic vision does describe the parallel tracks of discovery in data science and Geoscience embraced by EarthCube, and leaves open the implication that some form of production infrastructure will emerge, but the dual objectives should be described more directly (pursuant also to Finding 1.1).

Recommendation: Articulate the parallel "production" and "innovation" tracks within the program more directly. Innovation is embodied in the "Building Blocks" solicitation and in the data-enabled geoscience advanced by leveraging the most effective data science products in a functioning architecture.

2.5 Finding: Insufficient emphasis on enabling cross-disciplinary science through EarthCube.

Recommendation: Most of the defined Use Cases concern ways in which data science enables discovery within a single sub-discipline of geoscience.   More emphasis should be placed on enabling truly cross-disciplinary geoscience research.

2.6 Finding: Strategic "Imperatives" are ambiguous

Recommendation: The imperatives in the strategic plans should convey the most important things needed for the program *not* to fail. These should be specific, and extraneous imperatives removed.

3. ROADMAP

Charge: What is the quality of the roadmap? Are the short, medium, and long term plans and actions appropriate for meeting the strategic vision? Are the plans achievable? **How well has the roadmap, particularly the development of architecture, been informed by and incorporated existing resources, EarthCube community activities and funded projects?**

The panel felt that the Roadmap document contains many elements of a more fully fleshed out strategic vision document, but lacks an actionable management and implementation plan. The development paths from organic community driven grass roots may not be conducive to consensus in planning at the necessary level of detail. The lack of specificity leads to the following recommendations.

3.1 Finding: The program lacks a concrete 5-year implementation plan.
The current roadmap document contains 103 action items.   While many are laudable generic goals, they lack necessary specificity for implementation planning and are not attached to concrete metrics.   The AC believes the lack of specificity is strongly related to the governance issues mentioned previously. The roadmap needs to provide a firm understanding of what can be expected from this investment over the coming 5 years.

Recommendation 1: The Roadmap documents should be supplemented, or converted to, an "implementation plan" or "implementation roadmap".   This document should include specific prioritized milestones, and clearly defined metrics against which program progress is to be measured.

Recommendation 2: The documents should also make the period of consideration clear.   In this case, the implementation roadmap is understood to be a 5-year plan. It has been said that the "path to infrastructure takes decades," but we need a more concrete implementation plan to justify continuing the program beyond the next 5 years.   Short-term explicitly defined solutions are needed to avoid volunteer fatigue and external criticism.

Recommendation 3: While tackling the above imperatives, the EC needs to retain some of the grass roots aspects necessary to allow participation of communities that are just beginning to try and understand how cyber-infrastructure can assist in the development of databases, new knowledge, or new capabilities to extract knowledge.

3.2 Finding: The development of functional infrastructure is impeded by unrealistically broad scope, lack of standards, and lack of metrics.

Recommendation 1: There is a need to define and/or adopt specific standards that ensure projects are integrable and data repositories are interoperable. These standards should be established by a small group of technologists and geoscientists, informed through discussions with the broader community. All grant awardees must be required to follow these standards.

Recommendation 2: Based on these standards, there will be a need for solicitations to promote implementation of the "system of systems" approach. The LC should look to "low-hanging fruit" in this effort– for example, areas where quality data are not utilized, or obvious synergies are not exploited.

4. COMMUNITY ENGAGEMENT

Charge: How well integrated and coordinated is the scientific community in the roadmap and other EarthCube activities? Are the plans to engage various communities appropriate?

The committee noted that considerable time has been spent in developing a governance structure. The numerous end-user workshops have been valuable for some individuals, and several have produced useful reports. Persistent engagement occurs within funded RCNs. However, the panel felt the need for broader synthesis of workshop outcomes so that knowledge gained can be transmitted and best practices established across the EC community. Sustaining EC will require exploitation of existing resources and planning the next steps after each workshops including evaluation of emergent cross-cutting themes. Further effort is need to identify any domains that are missing or poorly represented.

4.1 Finding: Insufficient emphasis on data science objectives in the roadmap
There should be encouragement for research in the area of data science, rather than just focusing on data curation. The Roadmap does not address technologies needed to advance the data science agenda. The planning to connect technology development and science goals is inadequate. These connections represent the most exciting facet of the EarthCube initiative.

4.2 Finding: The gap between the Building Blocks projects and a functional "EarthCube architecture" is enormous.

Recommendation: This gap is likely not bridgeable through a community governance model, as discussed previously. A hand-picked group of experts is likely needed to make hard but informed choices based on results from funded projects. The two-year duration of the awards was also deemed unlikely to lead to successful implementation in an architecture. The AC commented that the development of an implementation roadmap may be premature. The LC may wish to develop this document after the products from the Building Block projects can be evaluated later in 2016. This exercise will determine which products are most likely to move the EarthCube architecture forward.

4.3 Finding: Proposal solicitation structure may not be conducive to broad interdisciplinary (GEO-CISE) involvement. Although the "EarthCube" community touches several thousand people, the engagement must become both larger and stronger for the program to be sustained and successful.

Recommendation 1: A different approach to the solicitations involving feedback and oversight from the LC could lead to substantially increased community involvement at the same funding level.

Recommendation 2: Funding a smaller dedicated group within the community may lead to the concrete infrastructure products that will ultimately make the case for a much broader group of scientists to engage in EarthCube.

4.4 General Recommendations:

Recommendation for LC: Consider whether there is an opportunity or sufficient knowledge to be gained by examining some of the end-user workshops that reported out well – to extract best value or best practices out of the best end-user-reports.

Recommendation for NSF: Future solicitations should contain input from the leadership Council about the standards and formats that must be considered when producing or integrating data sets that will foster community knowledge of the constructs needed

5. NEXT STEPS

Charge: What next steps should be taken by the EarthCube community? What are gaps or missing elements of the strategic vision and roadmap? Are there other communities or initiatives the EarthCube community should consider?

Overall the committee recommends that the next steps be towards tangible results that produce a recognizable step forward in the EarthCube Program.

Recommendation for LC: Write down the guiding principles for what is clearly meant by the system of systems and what standards it would need to address so that the community expectations can be channeled. Then get down into prioritized items that address both sides of the problem, the innovation gathering of science communities into exploiting the data assets (analysis world) as well as influencing the NSF awards solicitations where it is appropriate.

Recommendation for the LC: It appeared to the panel that the LC did not up to this point have a viable structure or mandate to produce an implementation plan. This needs to be corrected. The panel challenges the Leadership Committee to solve this dilemma, perhaps through a more authoritative management construct in the Project Support Office under the guidance/direction of the LC.

Recommendation for LC: The panel concluded that the leadership council should be prepared to meet the panel again in about 9 months to report on progress.

Recommendation for LC and the NSF: In the end, the AC was pessimistic about the ability of EarthCube program to evolve into a thriving enterprise in its second five years under the current organizational model. What is likely needed in "Phase II" is a new structure, deliberately designed to realize the requested implementation plan. This structure should involve a select set of experienced community leaders from each of the GEO discipline areas. These leaders would be empowered to build data science capacity within their disciplines, while also coordinating with other leaders to work toward realizing the (still undefined) EarthCube vision. The detailed implementation of this new governance model should await a broader review of the EarthCube Program, as discussed in Section 1.

Observation: The panel also noted that the available documents were silent on security. Activities that are intended to integrate across boundaries or federate across institutions and geographies need to be aware of and conscious of the ever-increasing need to align with the Federal Information Systems Management Act (FISMA).   FISMA is best postured on the front end of a development with an ultimate objective of withstanding third party audits so that EarthCube complies with "best practices" security guidelines.

6. PROGRAM SUSTAINABILITY

6.1 Finding: It is difficult to assess at the present juncture whether EarthCube will be sustainable.

Recommendation: The recommendations outlined above involve a transition to Phase II. They include the need for succinct definition of EarthCube, changes in the LC's governance approach to structured rather than consensus driven decision making, restructuring the process for more focused definition of program solicitations, and producing a bona fide implementation roadmap. These are seen as prerequisites to adopt best practices, communicate standards, and evolve to a production track. Only then can a sustainability plan be defined.

The AC feels that it is essential for the LC to move rapidly into Phase II as defined by the above actions. A useful timeline for responding to the AC recommendations is by the time of the All Hands Meeting in July so that a community communication plan can be implemented. The next AC meeting should involve feedback on progress and should be held in September this year.

Attendees at Reverse Site Visit
Advisory Group Membership:
Cathy Constable, UC San Diego, Scripps Institution of Oceanography, Co-Chair, Geo AC
Joshua Semeter, Boston University, Co-Chair, on GEO AC
Gwen Jacobs, University of Hawaii System
Al Kellie, National Center for Atmospheric Research NCAR/AD
Cliff Mass, University of Washington (Virtual attendance)
Dan Stanzione, University of Texas, Austin (Director/TACC)
Peter Wiebe, Woods Hole Oceanographic Institution

EarthCube Leadership Council:
Kerstin Lehnert, LDEO, Chair
Mohan Ramamurthy, Unidata, UCAR
Farzad Kamalabadi, University of Illinois
Lee Allison, Arizona Geological Survey
Plus virtual attendees (David Arctur, Margie Chan, Yolanda Gil, Basil Gomez, and others)

NSF:
Eva Zanzerkia, EAR, **Earthcube Program Managers**
Amy Walton, CISE/ACI
Marge Cavanaugh – Deputy AD/Geo

Charge for the EarthCube Advisory Group
Reverse Site Visit, February 4, 2016

Background:
EarthCube is a decade-long program to develop cyberinfrastructure for the academic geosciences community to better address the scientific challenges for studying the Earth System. At this stage of development, EarthCube governance has been developed and populated by the community.

The EarthCube Leadership Council, as representatives of the community, is in the process of producing a strategic vision and roadmap for further EarthCube activities based on outcomes of two years of EarthCube funded projects, and the deliberations of interested geoscientists, CI and CS researchers, social scientists and others. These documents include the community-developed vision and goals, and short-, medium- and long-term plans for a.) discussing and developing an appropriate cyberinfrastructure architecture for EarthCube in the context of existing community resources, b.) analyzing geosciences CI capabilities and resource needs, and c.) addressing these gaps with broad community input and infrastructure development.

Charge:
The Advisory Group is asked to guide the development of EarthCube by providing external scientific advice to the EarthCube Leadership Council. Advice will be given initially through a review of the EarthCube strategic direction and roadmap during a Reverse Site Visit, February 4, 2016. This guidance will enable the EarthCube Leadership Council and Governance to revise the strategic vision and roadmap and inform their decisions about EarthCube organization and activities.

The Advisory Group will provide further guidance to EarthCube Leadership on an on-going basis through virtual and in-person meetings as agreed to by both the Leadership Council and the Advisory Group. EarthCube Leadership may request specific feedback from the Advisory Group, and the Advisory Group may request that EarthCube Leadership provide documentation or information on scientific, organizational or governance activities.

The introductory meeting of the Advisory Group will be held as a Reverse Site Visit on February 4, 2016 at NSF. This meeting will allow the Advisory Group to become familiar with EarthCube community activities and provide initial guidance to the EarthCube Leadership Council on the Strategic vision and roadmap. The agenda will include executive sessions for the Advisory Group, as well as plenary presentations and discussions with EarthCube community representatives.
During the Reverse Site Visit the Advisory Group is asked to address the following main topics:

1. Strategic Vision: How well does the strategic vision represent the scientific and technological goals of the geosciences community? Does it present an appropriate framework to guide future EarthCube community activities?

2. Roadmap: What is the quality of the roadmap? Are the short, medium and long term plans and actions appropriate for meeting the strategic vision? Are the plans achievable? How well has the roadmap, particularly the development of architecture, been informed by and incorporated existing resources, EarthCube community activities and funded projects?

3. Community Involvement: How well integrated and coordinated is the scientific community in the roadmap and other EarthCube activities? Are the plans to engage various communities appropriate? Are there domains that are missing or poorly represented?

4. Next Steps: What next steps should be taken by the EarthCube community? What are gaps or missing elements of the strategic vision and roadmap? Are there other communities or initiatives the EarthCube community should consider?

We ask that the Advisory Group write a report following the Reverse Site Visit with recommendations for the EarthCube Leadership Council. This report is intended to be an initial set of guidance to EarthCube. Further recommendations, assessments and advice will follow from future interactions between the Advisory Group and the EarthCube community. The co-chairs will organize the members of the Advisory Group to produce this report.

Reverse Site Visit Advisory Group Membership:
Cathy Constable, UC San Diego, Scripps Institution of Oceanography, Co-Chair
Joshua Semeter, Boston University, Co-Chair
Gwen Jacobs, University of Hawaii system
Al Kellie, National Center for Atmospheric Research
Cliff Mass, University of Washington
Dan Stanzione, University of Texas, Austin
Peter Wiebe, Woods Hole Oceanographic Institution