EarthCube Architecture & Implementation Plan Executive Summary

EarthCube was initiated by the National Science Foundation (NSF) in 2011 to transform geoscience research by developing cyberinfrastructure to improve access, sharing, visualization, and analysis of all forms of geosciences data and related resources.

For over 50 years, NSF has funded domain-specific geoscience activities across dozens of various research areas. The solutions needed to address the ever-increasingly complex scientific questions we face today require sophisticated integration of these seemingly disparate research areas, as well as the ability to look at the data produced by the research in fresh new ways. Currently there is no such integrated architecture. For geoscience to move towards cross-domain research, as well as the long-term societal benefits such research could provide, both existing and emerging resources of data, software, platforms, and infrastructure need to continually improve in terms of interoperability, trust, accessibility, and availability.

EarthCube seeks to transform geoscience research through improving the interoperability and integration of geosciences data resources by creating an improved “System of Systems” (SoS). An EarthCube solution needs to address both the challenge of aggregating, discovering, describing, and assessing the quality and utility of resource metadata, as well as address the need for an interoperability testing workbench infrastructure with the necessary guidance. When EarthCube launched five years ago, the challenge was to establish the functional and technical scope and role of its implementation from an incredibly vast and complex set of scientific and technical needs, all within its capital and management limits. The following summarizes the proposed architecture and plan.

EarthCube Architecture

Defining EarthCube Program Role as the Cross-Domain Portfolio Manager and Interoperability Testing Workbench

To move the architecture forward, the EarthCube Science Support Office (ESSO) commissioned a rapid solution architecture and implementation development activity that could result in a phased solution and development approach for the sustained platform to address these challenges. Multiple concepts of operations had been reviewed prior to this effort (EC-Develop SoS Operational Platform, Standards Body, Funding Body). In a short period, the team reviewed, collated, and leveraged years of effort and inputs (such as the EarthCube Architecture Final Report, Conceptual Designs, the Reverse Site Visit, GEO 2020 and the EarthCube Use Cases) which identified a comprehensive set of requirements describing the interoperability challenges to establish architecture needs. The resulting architecture proposes dual roles for EC to enable the desired progression:

- **Portfolio Management**: Standards and Assessment Portfolio Management to register, guide and assess community resource needs with a focus on interoperability, availability, accessibility, quality improvements;
- **Testing Workbench**: R&D Incubation Environment Facilitation for Resource Testing and creating value-added capabilities to directly guide, enable, and invest in resource improvements.

1. **Establish a Cross-Domain Geoscience Resource Registry through Curation Partnering**: First, by collecting inputs from partner metadata aggregators (such as DataCite, CrossRef, ORCID, etc.) EarthCube will develop and classify a supply of a discoverable, high-quality metadata (basic imported and extended detailed view) for its read-only registry of geoscience resource components. By partnering with existing aggregators, EarthCube will rely on existing established curation processes for maintaining and responding to resource

metadata needs. The registry would be discoverable via API as well as a basic search site on earthcube.org, with user traffic likely coming through “search in the wild” or partner sites.

2. **Add Guiding Usability Value to Scientists by Assessing Resources**: To assist in improving the interoperability of these resources, the registered resource components and resulting interoperability solutions will all be subject to platform and community based assessments. These assessments with supporting resource profile details and resulting actions will be EarthCube’s largest value-add to the community by helping guide the quality maturity, crowd use input, and implied use statistics to help scientists and users within the community understand the value and usability of a resource for their efforts. The assessment scores would be a substantial content benefit for partner sites.

3. **Create a Testing Workbench Environment for Interoperability Assessment and Integrating Capabilities**: To address both scientific end-user and technical interoperability developer challenges, these resources can be further investigated, developed, and tested for interoperability by deploying on an EarthCube interoperability R&D Workbench. The Workbench will be stocked with common data slices and tools to engineer or craft interoperable solutions of varying degrees of complexity. It will leverage code repositories and containerization. Newly created capabilities can be more easily shared and transferred to enhance process management, data interoperability/access, service orchestration, validation & analysis, development tools & notebook environments which cumulatively provide enhancements for the overall geoscience system of systems capabilities. After a scientist’s integrated resource solutions are tested, they could be transferred to already existing, external, scalable operational environments for full scale and repeated execution. Multiple existing options or future EarthCube sponsored funding could provide operational hosting of EarthCube developed resources. At this time EarthCube has neither the mission nor the funding to host operational data, services, platforms, or functions.

4. **Reward and Incentivize Community Engagement through Reward and Recognition**: One of the more innovative parts of EarthCube, and likely the most challenging, is fostering an environment that allows for continuous improvement in its community engagement and incentive toolsets, approaches, and capabilities. EarthCube will be able to track, recognize, and reward activity that helps achieve EarthCube goals through concepts like incentive-based tasks for rewards and recognition, cloud allocation rewards, linking author reputation profiles, and integrated communication tools. This is critical to assure engagement levels are incentivized to move towards desired outcomes such as workflow chaining, resource interoperability improvement, improved metadata based on assessment, EarthCube code and data sharing, and assessment.

EarthCube Governance

Key Additions to the Current EarthCube Governance

The implementation plan provides functional and organizational improvements to guide this effort in integrating the current EC governance teams with the business practices and the solution architecture of the EC Workbench, its supporting capabilities, and the ESSO Program Management Office. The three major functions will be: Community Engagement, Architecture Development, and Operational Performance Oversight.

1. **Establish Supplier and User Community Engagement Governance**: The outcomes of governance will be utilization of metadata-driven measurements based on maturation, metadata quality, interoperability, access, and availability improvements in the portfolio related to Process Management, Data Interoperability/Access, Service Orchestration, Validation & Analysis, and Development Tools & Notebook Environment. Supplier Engagement Governance would engage and guide suppliers through the assessment lifecycle, evaluate capability gap analysis, drive supplier incentive and recognition, and facilitate workbench efforts. User Engagement Governance

would evaluate and drive demand generation analysis, monitor user communications, and track assessment tasks, actions, and results guiding incentive activities.

2. **Govern Phased Release of Architecture**: An Architecture development Governance would oversee the implementation of the detailed Solution Architecture & Requirements as scope and qualities guidance and detailed Implementation Plan & Sequence as time and cost guidance. This group working with ESSO will guide the implementation of the EC architecture core services (Discovery, Registry, Community & Partnership Services, Assessment, and Workbench). This Governance team will be responsible for organizing, prioritizing, phasing, implementing (agile), accepting, and communicating EarthCube Architecture progress.

3. **Establish a Single Operational Performance Group to Measure EarthCube Activities and Use**: This group would regularly measure and report the performance of the EarthCube Program, Governance and Architecture including Service Operations, Resource and Workbench Use, and User Feedback.

Next Steps

Over the next three to six months, the ESSO will coordinate with NSF EarthCube leadership for acceptance and initiation of the architecture & implementation plan guided by the goal to provide an initial operating capability within 10 months from start of implementation.