EarthCube Gap Analysis Working Group Final Report - January 2016
Funded Projects Gap Analysis TAC WG (TGA): http://preview.tinyurl.com/o35hjwl

Executive Summary

The initial EarthCube Technology & Architecture Committee (TAC) Gap Analysis Working Group proposal was submitted to the EarthCube Office on November 20, 2014. Based on recommendations for revisions from the TAC, a revised Gap Analysis Working Group Proposal[1] was submitted on December 5, 2014. The proposal was approved on January 28, 2015. The working group name is Funded Projects Technology & Architecture Gap Analysis. The short name is TAC Gap Analysis (TGA).

The TGA WG made numerous contributions to but not limited to TAC meetings, working groups, presentations, research network collaborations, and technical documents. Some key TGA WG outcomes include (1) a peer-reviewed journal article in OCLC Systems & Services: International Digital Library Perspective (OSS:IDLP)[2], (2) key concepts and definitions, (3) gap analysis model, (4) funded projects questionnaire[3], (5) some analysis of funded projects questionnaire responses, and (6) recommendations of future efforts for establishing an integrated EarthCube capability.

The TGA WG was composed of committed participants from three different organizations, a chair, and a co-chair from January 2015 – November 2015. Overall, the TGA WG activities and outcomes aligned with the following articulated TAC goals:

- "Identifying gaps in coverage of needed cyberinfrastructure capabilities, and determining recommendations on how to fill them"
- "Developing recommendations for monitoring and assessing performance of EarthCube infrastructure in coordination with other EarthCube groups. Monitoring technical requirements with the goal to ensure EarthCube is meeting end user needs[4]" – Charter of the EarthCube Technology and Architecture Standing Committee (March 2015)

It is recognized that there is always a need for the identification and recommendations for addressing gaps in communication, collaboration, and capacity. The TGA WG met as needed (e.g. bi-monthly, monthly, impromptu) via Web Ex, email, and phone to discuss goals, plans, and outcomes. The informal budget of $10,000 was used to fund TGA WG members attendance

---

[1] https://drive.google.com/file/d/0B5cNNPoOTUmkeFBPNEt6VFg5b0E/view?usp=sharing
[2] Smith II, P. L., Malik, T., & Berg-Cross, G. (2015). Rediscovering EarthCube: Collaborate. Or collaborate not. There is no I. OCLC Systems & Services: International Digital Library Perspectives (OSS:IDLP), v. 32(3). [Abstract] | [Preprint]
[3] http://skc.isi.edu/wiki/earthcube/index.php/Funded_Projects_Questionnaire
[4] http://earthcube.org/sites/default/files/doc-repository/TAC-charter_Approved20150305.pdf

to the April 8-10, 2015 Tech Hands Meeting[5] or the May 27-29, 2015 All Hands Meeting[6]. Funds were used judiciously among members and funds were not used to fund the same member to two different meetings. Some members used funds from their funded projects to attend conferences. The only face-to-face TGA WG was collocated with the April 2015 Tech Hands Meeting to lower expenditures. The Smith II, Malik, & Berg-Cross (2015) article include preliminary analysis of some of the spring 2015 funded projects' questionnaire responses.

TGA WG Outcomes

The recent peer-reviewed article that includes (1) key concepts and definitions relevant to TGA WG and the (2) gap analysis model conceptualizing the interrelationships and connections between sciences drivers, use cases, prototypes within testbeds as they relate to an integrated EarthCube GeoVision 2020[7] capability are two key outcomes.

Key Concepts and Definitions

- **EarthCube (EC)** – EarthCube[8] is a community-driven, ten-year initiative aimed at transforming the conduct of geosciences research and education to develop the cyber-infrastructure for the geosciences in order to better enable transformational science within and across disciplines. EarthCube enables "fostering community-governed efforts that develop a common cyberinfrastructure for the purpose of collecting, accessing, analyzing, sharing, and visualizing all forms of data and related resources, through the use of advanced technological and computational capabilities[9]"
- **Building Blocks (BB)** – BB represented initial software component development for EarthCube.
- **Conceptual Designs (CD)** – CD represented broad architectural design for EarthCube. The idea is that understanding the conceptual designs will facilitate the scoping, alignment and coordination of EC BBs and to help manage architectural complexity by describing component interdependencies in a usable, understandable fashion.
- **Research Coordination Networks (RCN)** – "EarthCube RCNs are intended to advance geosciences cyberinfrastructure through interaction, discussion and planning between geoscientists and cyberinfrastructure experts. RCNs provide opportunities for academic geosciences communities to organize, seek input, come to consensus and prioritize data, modeling, and technology needs, as well as standards and interoperability within and across domains[10]".

---

[5] http://earthcube.org/group/tech-hands-meeting
[6] http://earthcube.org/info/about/2015-all-hands-meeting
[7] Geoscience 2020 http://earthcube.org/sites/default/files/doc-repository/GeoScience%202020.pdf
[8] http://www.nsf.gov/pubs/2013/nsf13529/nsf13529.htm
[9] http://earthcube.org/info/about
[10] Gap Analysis Model and Figure 1 from the EarthCube Strategic Technology Plan, pp. 15-16.

- **Gap** – Identification of architectural, functional and technical gaps is a key aspect of enterprise architecture efforts (Saha, 2004) and identifiable with conceptual designs. Within the scope of this paper, a gap is defined as an unmet need (e.g. architectural, functional, and technical). An example is a capability gap which is the inability to achieve a desired effect, analysis, or data process under specified standards and conditions through combinations of means and ways to perform a set of tasks.
- **Gap Analysis** – Within the scope of this paper, gap analysis is the comparison of existing capabilities to a set of required capabilities (e.g. "What we have against what we need") as defined by EarthCube Science Domains and Technical Architecture Strategy relevant to appropriate working groups. For example, as part of EarthCube software and Service development, analysis were used to document which services and/or functions have been accidentally omitted, deliberately eliminated and/or need to be developed.
- **Gap Analysis Model:** In order to understand potential gaps, the TGA WG model was used to capture the current state of EarthCube, where progress is being made to fulfill gaps, gaps that are being identified, and gaps that still need to be filled. **Figure 1** shows one such model in which the squares represent funded projects and the circles represent the class of science use cases that they are currently developing the technology for. The green lines show that the gap is being fulfilled in that some technology is being developed to address the chosen science use case. Outstanding gaps include the ability of funded projects to apply themselves to more realistic science and encompassing use cases to accomplish the EarthCube GeoVision 2020 and for funded projects themselves to expand their technical features (Test Prototypes, TestBeds, etc.) so that they can address more realistic science use cases.
- **Science Scenarios –** "Science scenarios and use cases capture specific contexts where technology can be inserted in science practice, and can therefore play a crucial role in facilitating communication and fruitful interactions between computer scientists and geoscientists[11]" (Gil et al., 2015, p. 3).
- **TestBeds –** According to National Oceanic & Atmospheric Administration (NOAA) Testbed & Proving Grounds[12], testbeds "facilitate the orderly transition of research capabilities to operational implementation through development testing in testbeds, and pre-deployment testing and operational readiness/suitability evaluation in operational proving grounds, as approved in the Guidelines and Performance Measure."
- **Use Case** – "According to the Unified Modeling Language, 'a use case shows the interaction between the system and 'actors', which may be human or other systems'" (Kelbert, 2014, p. 5). To understand researchers' needs, the science case, one may adapt the usual software oriented use case model to describe more abstract things. These

[11] http://earthcube.org/document/2015/earthcube-collaboratories
[12] NOAA Testbeds & Proving Grounds http://www.testbeds.noaa.gov/

could be system/capability boundaries, how they interact, scientific workflows involved and some idea of services offered by the system as well as the data used.

Gap Analysis Model

The TGA WG defined key concepts and created a model in which to conduct preliminary work and prepare for future gap analysis. During the course of preliminary work through working group meetings, observations of funded projects' conference presentations, and review of outputs, the gap analysis group agreed on three broad, high-level dimensions for the gap analysis. The gap analysis model (see Figure 1) conceptualizes an example of the three dimensions of gaps (features and functionality, integration requirements, and operational requirements) in practice within funded projects and in theory within testbeds environments.

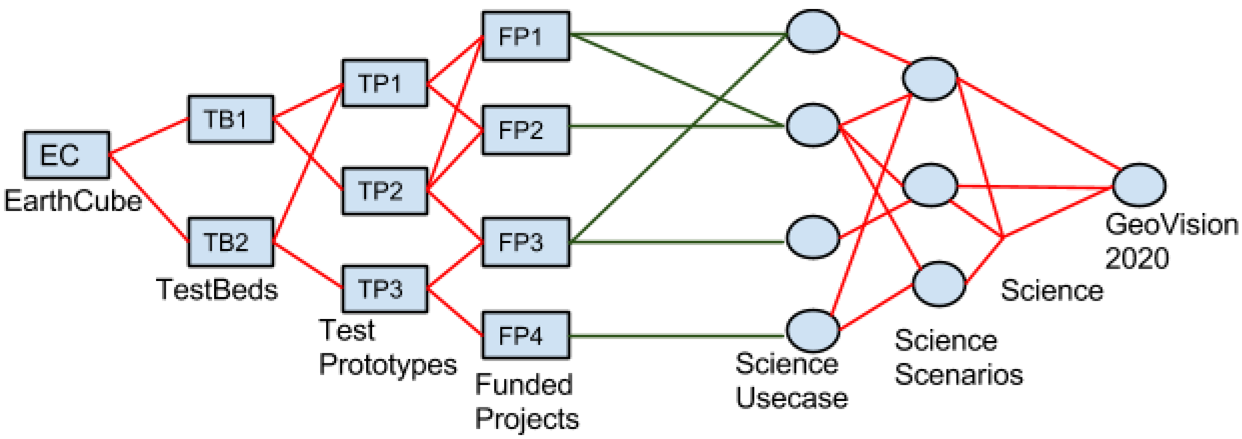

Figure 1: A model for gap analysis (Developed by Tanu Malik)

Since the survey was aimed at identifying gaps, survey questions were organized along areas of anticipated gaps. The gaps impede data integration, sharing, and research collaboration as represented in the absence of use cases, prototypes, and testbeds in which to develop capacities and infrastructure.

Gap Analysis Working Group Preliminary Framework and Model Section

The EarthCube funded projects provide vital insight into the technologies that are important for EarthCube and a vision of the architecture of a future EarthCube. Thus, there is a critical need to aggregate, coordinate, and articulate accurate and comprehensive information about the funded projects across the EarthCube community, and identify gaps in capabilities.

The TAC Gap Analysis Working Group (TGA WG) identified three dimensions of performing gap analysis, namely:

- Features and Functionality (e.g. Demo Science Use Case 1, etc.)
- Integration Requirements (e.g. TestBed 1, Test Prototype 1)

- Operational Requirements (e.g. Applying Test Prototypes and TestBeds, etc.)

The three dimensions of gaps represented common challenges among the funded projects. The desired goals for some funded projects is to increase the features and functionality, integration, and operational capacities of software APIs and tools to facilitate data access, discovery, and integration of research data within and across disciplinary domains. An example is the effective execution and performance in the use of Network Common Data Form (NetCDF), Sensor Model Language (SensorML), and WaterML data integration crosswalks within and across domains.

The Technical (Non-Functional) and Architectural Requirements gaps will include operations such as loose coupling, federation, standards (compliance) and performance requirements (response time, resource utilization and availability).

Additionally, to make the task of gap analysis easier, the TGA WG has made progress on two efforts: (1) the TGA WG surveyed funded projects for gaps and opportunities for collaboration with other funded projects and (2) coordinated with other working groups (such as the science use case and testbed working groups), to identify existing gaps from other working groups' perspectives. The TGA WG objective is to perform gap analysis that will allow the comparison of current capacities in existing applications, services, data, system, and technical architecture against a set of targeted architectural needs for those applications, services, data, and technical capabilities.

The gap analysis model (see Fig. 1) conceptualizes key components in the exploration of gaps stemming from the GeoVision 2020 plan to science drivers to science scenarios to developing use cases. One of the goals of building blocks funded projects is to create use cases while one of goals of the funded research coordination networks is to connect multiple funded projects through test prototypes as an outcome of testbeds. Through the processes and workflows from the GeoVision 2020 plan to testbeds supported through the collaborations of funded projects, conceptual designs, and research coordination networks, functional components of EarthCube are supported to advance the access, integration, and management of data across the geosciences.

Building Blocks (BB) Projects

With suggested input from a co-PI of one of the funded projects and based on understanding of the planned outcomes, the Building Blocks (BB) have been assigned into one or more of five key categories based on their projects' planned outcomes including technical, science, and/or community outcomes. The Building Blocks' planned outcomes are distinguished by the following categories: **Framework or Testbed; Standard; Tool; Report; Community Engagement.**

Project	Category	Question 4 - Please list up to five planned outcomes of your project including technical, science and/or community outcomes.
BB1-BCube	**Framework or testbed**	**1.** A cloud-based brokering **framework** that is mediating query and access across diverse data sources, serving science use cases in hydrology, oceanography meteorology and cryospheric research. **2.** A queryable catalog of Earth Science resources (data sets, and services by which data are made accessible), and the cloud-based Web crawler/reasoner that is used to populate and update this catalog. **3.** A **report on the current state of metadata and service description standards** and the impacts these have on interoperability, including recommendations for improvements. **4.** A prototype tool for scientists who seek an alternative to depositing their data in a repository and wish to advertise their data in a manner that is discoverable on the web. **5.** A cloud-based interoperability testbed.
BB2-CHORDS	**Community Engagement**	**1. Engagement** with real-time geosciences community, including real-time data providers and consumers. **2.** Demonstration of a prototype system that enables cloud-sourced, standards-based, easy access to real-time sensor data and otherwise proprietary or unavailable real-time sensor data streams. **3.** Specifications and documentation for scientists to use the prototype system.
BB3-CINERGI	**Framework or testbed, Standard, Tool, Community Engagement**	**1. Searchable inventory of geoscience resources** across domains and methodology for extending it to additional use cases, with particular focus on resources that are intended to be part of EarthCube. **2.** Metadata processing and curation pipeline, including tracking provenance of metadata records as they are processed. **3.** Validation mechanisms for metadata documents. **4.** Information model (metadata profile and resource ontology) for documenting EarthCube resources. **5. Engaging** geoscientists into inventory compilation and discovery use cases, via community resource viewers.
BB4-Cybercon nector	**Standard, Tool**	**1.** Significantly increase research productivity in the Earth science modeling community, **2.** Enable the effective use of the existing Sensor Web data and Earth Observations through **open Web interfaces and metadata standards**, **3.** Foster collaborations among Earth system modelers, geospatial information scientists, and information technologists, and **4.** Enhance infrastructure for Earth science research and education.

Table 1a: Sample of Building Blocks' planned outcomes distinguished by Framework or Testbed, Standard, Tool, Report, or Community Engagement

The Building Blocks' planned outcomes were distinguished into one or more of the key categories based on understanding of the planned outcomes such as BB1-BCube "a cloud-based brokering **framework**…" (See Table 1). This process was applied to the remaining BB resulting in categorization of BB to compare and contrast the similarities, differences, and interrelationships of planned outcomes for future gap analysis of actual outcomes. Table 1a provides the first four of the BB planned outcomes from **APPENDIX A - Building Blocks (BB) Categories, Planned Outcomes, Use, and Availability** that includes responses to survey questions 4, 5, and 7.

- Question 4 – Please list up to five planned outcomes of your project including technical, science, and/or community outcomes.
- Question 5 – Which outcomes can be used directly by the scientific community, for example, specific tools?
- Question 7 – What outcomes will be available for use by the All Hands Meeting in May from the list in Question 4?

Planned Outcomes Use by the scientific community

Below are the results of all Building Blocks' planned outcomes for use by the scientific community and available for use by the 2015 All Hands Meeting.

Building Blocks Planned Outcomes	Responses
Planned Outcomes 1, 2, 3, 4, 5 can be used directly by the scientific community	20%
Planned Outcomes 1, 2, 3, 4 can be used directly by the scientific community	13%
Three of more planned outcomes can be used directly by the scientific community (Note: varied across funded projects and not sequential)	66%
Planned Outcome 2 can be used directly by the scientific community	93%
Planned Outcome 1 can be used directly by the scientific community	60%

Table 1: Blocking Blocks (BB) Planned Outcomes that can be used directly by the scientific community

Outcomes available for use by the All Hands Meeting (AHM) in May 2015

- **73% of the BB reported at least one planned outcome available for use by the 2015 AHM**. However, the planned outcomes available varied greatly from 1 to 5 or test.

Recommendations

The TGA WG provided only preliminary exploration and investigation in the identification of gaps. It is recommended that other working groups analyze the funded projects questionnaire results from their group's perspective and then all working groups aggregate, collate, and organize all working groups' analysis.

- Based on the work of this group, further analysis of current outcomes against planned outcomes from the funded projects questionnaire is highly recommended as immediate need to assess the current efforts, collaborations, and outcomes.
- **Classifying current outcomes** as (1) Testbed of framework, (2) Standard, (3) Tool, (4) Report, or (5) Community Engagement is recommended.
- **Classifying current gaps** as (1) Scientific Reproducibility and Scientific Benchmarking Capabilities, (2) Interdisciplinary Communication, (3) Integration of Data and Models, (4) Data and Software Management, and (5) Data Discovery & Access is recommended. A project funded to explore these recommendations and develop a comprehensive gap analysis of EarthCube funded projects with integrated success metrics is needed.
- **Developing, analyzing, and prioritizing** uses cases within testbeds are needed for further gap analysis.

References:

1. Gil, Y., Arctur, D., Kelbert, A., Pearlman, J., Rubin, K., & Wood-Charlson, E. M. (2015). A CRISP Approach for EarthCube: Collaborative Resource Incubators (CRI) for a Sci-Tech Portfolio. Retrieved December 23, 2015 from http://earthcube.org/document/2015/collaborative-resource-incubators.
2. Kelbert, A. (2014), Science and cyberinfrastructure: The chicken and egg problem, Eos Trans. AGU, 95(*49*), 458–459, doi:10.1002/2014EO490006
3. Saha, P. (2004). Analyzing the open group architecture framework from the GERAM perspective. The Open Group, Tech. Rep. Retrieved November 3, 2015 from https://www.opengroup.org/architecture/wp/saha/TOGAF_GERAM_Mapping.pdf.