Rediscovering EarthCube: Research Lifecycle, Functional Areas, and Gap Analysis

All Hands Meeting 2015
May 27-29, 2015
Arlington, VA

Plato L. Smith II, CLIR/DLF Postdoc Fellow
University of New Mexico
psmith4@unm.edu

Abstract

The National Science Foundation (NSF) Public Access Policy (NSF 15-52) released March 18, 2015 and the NSF Data Management Plan Requirements activated January 18, 2011 have both significantly contributed to the development of data management and sharing policies across disciplinary domains. The research lifecycle and data curation processes model includes important components relevant to funding agencies' developing and emerging guidelines on data management and functional areas of EarthCube. As part of a Funded Projects (FPs) Questionnaire, FPs identified functional areas that support EarthCube and other FPS in which to interface (collaborate) to name two. The poster presentation will use text and images to (1) describe some research, publication, and curation processes that address data management plans and sharing requirements; (2) correlate to functional areas of EarthCube; and (3) contribute to future development of a model for gap analysis of EarthCube funded projects.

Research Lifecycle and Data Curation Processes Model

- JISC Research Lifecycle Model (2013)
- The USGS Data Lifecycle Model (2012)
- USGS Data Lifecycle Producer and Consumer Perspectives (Gunther & Govoni, 2006)
- Level 1, 2, and 3 Data Curation (Lord & Macdonald, 2003)

Functional Areas & Gap Analysis

Funded Projects'' components/products in support of **Functional Areas of EarthCube** as identified in Question #13 of Funded Projects Questionnaire:

- **Data Access (17)** - 11 BB, 1 CD, 5 RCN
- **Data Discovery (19)** – 13 BB, 1 CD, 5 RCN
- **Data Integration (17)** – 12 BB, 1 CD, 4 RCN
- **Data Management (15)** – 9 BB, 1 CD, 5 RCN
- **Modeling (12)** – 10 BB, 1 CD, 1 RCN
- **Ontologies (10)** – 9 BB, 1 CD
- **Other (11)** – 8 BB, 2 CD, 1 RCN (e.g. services, analysis, reproducibility, interoperability, logic programming, references)

Three dimensions of performing Gap Analysis:
1.Feature and Functionality (e.g. Demo Science Use case)
2.Integration Requirements (e.g. TestBed1, Test Prototype 1)
3.Operational Requirements (e.g. Applying Test Prototypes and TestBeds to funded projects)
Technical (Non-Functional) Architectural Requirements gaps will include such things as (1) loose coupling, (2) federations, (3) standards (compliance), and (4) performance requirements (response time, resource utilization and availability)

Acknowledgements

•EarthCube Office – Racheal Black and Anna Katz
•Technology and Architecture Committee (TAC) Leadership – Jay Pearlman and Yolanda Gil
•TAC Gap Analysis Working Group
•Special thanks to Tanu Malik, Gary-Berg Cross, Emily Law, Mike Daniels, and Steve Diggs

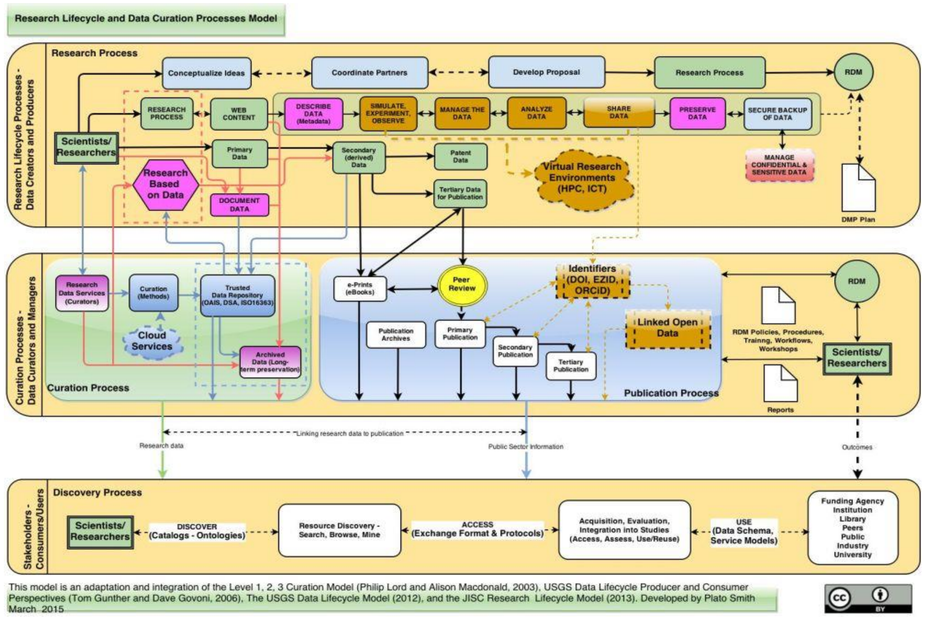
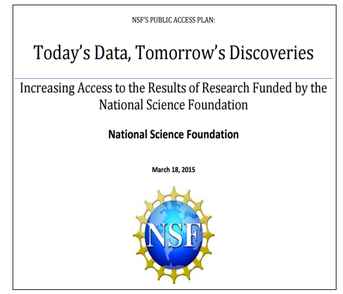


Figure 1: Research Lifecycle and Data Curation Processes Model

Project Capabilities mapped to functional areas of EarthCube:
•Data Access (17)
•Data Discovery (19)
•Data Integration (17)
•Data Management (15)
•Modeling (12)
•Ontologies (10)
•Other (11)



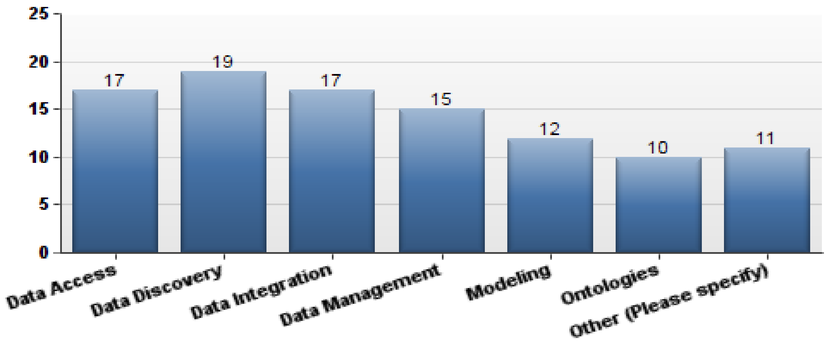


Figure 2: What functional area(s) of the EarthCube will your components support (Q13)?

#	Answer		Response	%
1	BB - BCube: A Broker Framework for Next Generation Geoscience		13	59%
2	BB - CHORDS: Cloud-hosted Real-Time Data Services		6	27%
3	BB - CINERGI: Community Inventory of EarthCube Resources for Geoscience Interoperability		19	86%
5	BB - CyberConnector: Bridging the Earth Observations and Earth Science Modeling for Supporting Model Validation, Verification, and Inter-comparison		10	45%
6	BB - DisConBB: Integrating Discrete and Continuous Data		8	36%
9	RCN - iSamplES: The Internet Samples in the Earth Sciences		8	36%
11	BB -Earth System Bridge: Spanning Scientific Communities with Interoperable Modeling Frameworks		11	50%
13	BB - Enabling Scientific Collaboration and Discovery through Semantic Connections		12	55%
14	BB -GeoDataspace: Simplifying Data Management for Geoscience Models		7	32%
16	BB - GeoDeepDive: A Cognitive Computer Infrastructure for Geoscience		8	36%
18	BB - GeoSoft: Collaborative Open Source Software Sharing for Geosciences		14	64%
19	BB - GeoLink: Leveraging Semantics and Linked Data for Data Sharing and Discovery in the Geosciences		15	68%
20	BB - GeoWS: Deploying Web Services Across Multiple Geoscience Domains		14	64%
21	BB - ODSIP: Specifying and Implementing ODSIP, A Data-Service Invocation Protocol		6	27%
22	CD - Enterprise Architecture for Transformative Research and Collaboration Across the Geosciences		8	36%
23	CD - Developing a Data-Oriented Human-centric Enterprise Architecture for EarthCube		7	32%
24	CD - A Scalable Community Drive Architecture		5	23%
25	RCN - C4P: Collaboration and Cyberinfrastructure for Paleogeosciences		7	32%
26	RCN - CReScNT: Coral Reef Science and Cyberinfrastructure Network		4	18%
27	RCN - ECOGEO: Oceanography and Geobiology Environment Omics		6	27%
28	RCN - EC3: Challenges of Field Data Collection, Management, and Integration		9	41%
29	BB - Digital Crust: An Exploratory Environment for Earth Science Research and Learning		7	32%
30	RCN - SEN: A Sediment Experimentalist Network		7	32%
31	BB - A Geo-Semantic Framework for Integrating Long-Tail Data and Models		11	50%
45	Other		0	0%
46	BB - Software Stewardship for the Geosciences - now is Geosoft		9	41%

Figure 3: Are you interested in interfacing with one or more of the EarthCube projects (Q17)?

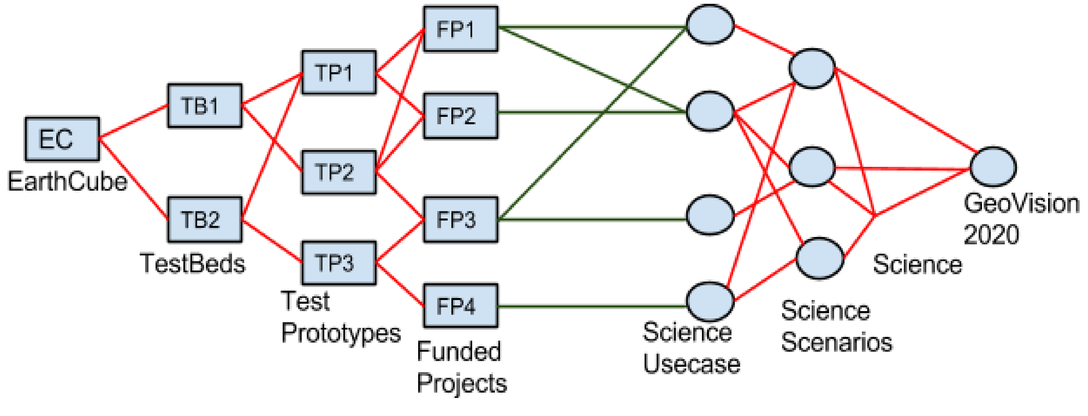


Figure 4: A model for gap analysis (developed by Tanu Malik) (Used with permission)

Potential Future Research Questions

1. **How do funded projects' components/products map between general and domain-specific research data management lifecycle (1) models, (2) ontologies, and (3) plans?**
2. **How is the data lifecycle management from data creation and curation, to interfaces and tools, trustworthiness, and long-term preservation to data policies and sustainability guided within and across funded projects?**
3. **What are the science drivers' benefits from funded projects' (1) components, (2) outcomes, and (3) products?**
4. **How does the level of maturity of funded projects' components/ products impact the development of a testbed** of (1) prototypes, (2) funded projects, and (3) use cases?