A Library Approach to Establish an Educational Data Curation Framework (EDCF) that Supports K-12 Data Science Sustainability

Benjamin D. Branch  
*Purdue University, bdbranch@gmail.com*

Kristin Wegner  
*The GLOBE Program, kwegner@globe.gov*

Steven Smith  
*Purdue University, mrssmith@purdue.edu*

Darrell G. Schulze  
*Purdue University, dschulze@purdue.edu*

Venkatesh Merwade  
*Purdue University, vmerwade@purdue.edu*

See next page for additional authors

Follow this and additional works at: [http://docs.lib.purdue.edu/lib_fspres](http://docs.lib.purdue.edu/lib_fspres)  
Part of the [Earth Sciences Commons](http://docs.lib.purdue.edu/earth-science), [Library and Information Science Commons](http://docs.lib.purdue.edu/library-and-information-science), and the [Other Teacher Education and Professional Development Commons](http://docs.lib.purdue.edu/other-teacher-education-and-professional-development)

**Recommended Citation**

Branch, Benjamin D.; Wegner, Kristin; Smith, Steven; Schulze, Darrell G.; Merwade, Venkatesh; and Jung, Jinha, "A Library Approach to Establish an Educational Data Curation Framework (EDCF) that Supports K-12 Data Science Sustainability" (2013). *Libraries Faculty and Staff Presentations*. Paper 64.  
[http://docs.lib.purdue.edu/lib_fspres/64](http://docs.lib.purdue.edu/lib_fspres/64)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
**Abstract** - It has been the tradition of the libraries to support literacy. Executive Order, Making Open and Machine Readable the New Default for Government Information, May 9, 2013, implies new roles for libraries. The library has the responsibility to support geospatial data, big data, earth science data or cyber infrastructure data that may support STEM for education. (More information can be found at http://www.whitehouse.gov/the-press-office/2013/05/09/executive-initiated in Purdue research. The EDCF may be applied to geospatial data service, engagement and outreach endeavors to augment the data science and climate literacy needs of future global citizens.

**EDUCATIONAL DATA CURATION FRAMEWORK (EDCF)**

An EDCF couples the data integration of library science and data management of faculty researcher needs, broader impacts, and its outreach delivery system as a measure of data efficiency. Here, disciplines related to STEM may benefit where their new audience or targeted growth can be expanded. Learning with an active library data management framework supports formal and informal learning. Such may be key to expanding scales of research delivery.

**Research Question:** How can libraries be involved in the data curation of geospatial data to increase the broader impact and provide open access of university research data?

**Potential Earth Science (STEM Pipeline) Library roles**

**Disciplines and subject areas where interdisciplinary efforts can support outreach and facilitate knowledge transfer with EDCF**

- Researchers
- Researchers
- Library Engagement
- Library Engagement

**Benefits of EDCF & libraries engagement in STEM disciplines, outreach and data curation activity and supporting the GLOBE program at www.GLOBE.gov.**

1. A geospatial campus community strives for outreach and data curation success as a collective. 2. GLOBE provides researchers with freedom to extend research audiences and grant collaborations. 3. Teachers can have support from a data curated lifecycle where STEM learning may be customizable for their locations. 4. The University and its communities benefit from the greater use of research data and broader impacts of the research.

**Pipeline Stimulation** is working to increase the numbers/degrees/maturatin of STEM disciplines at an earlier educational point. The theory is if you build student interest in STEM areas during middle and high school that interest will transfer to a greater number of advanced STEM degrees.

Lastly, it should be noted that the interdisciplinary collaboration and demonstration of library supported outreach partners and national organizations such the GLOBE program may best foster EDCF development. This trend in data science where library roles may emerge is consistent with NASA’s wavelength program at http://nasawavelength.org. Mr. Steven Smith, an outreach coordinator, led this Purdue University outreach activity involving the GLOBE program with support by the Purdue University Libraries GIS department. We believe one of our summer program participant comments sums it up best: "Excellent workshop. High quality materials, presenters, and setting.”

**Educational Data Curation Framework Consideration**

- Our major goal is to assist researchers to achieve sustainable data curation services.
- Libraries should partner with outreach experts to build EDCF concepts.
- Libraries should become involved in GLOBE program activity to further EDCF paradigms with their outreach partners.
- The National Science Digital Library should consider EDCF as a plausible model of engagement where libraries play a role in sustainable learning on local, state and national levels.

**Benefits**

- Established Program:
  - Pre-made Curriculum
  - Support Mechanism
  - Recruitment Factor (Wide Audience)
- Multifaceted Technologies:
  - Modules
  - Self-paced
  - Real-time data collection
  - In the Field (Portable)
  - Universal tools in one Device
  - Visualizations in the Field (DEM)

**Research findings to be further investigated**

- Over-time, a joint library, outreach, faculty approach with an integrated data curation policy and practice may augment STEM outcomes at a lower cost than regular university expenses.

**Research Challenges and Barriers**

- Continued cross disciplinary collaboration with STEM incentives
- Geospatial server set-up and prototyping services
- Open campus collaboration with library, outreach and researchers for STEM pipeline support
- Library support for geospatial data specialists, data scientists & researchers
- Recruiting teachers and forming a community of STEM learning that is university defined and driven

---

**Authors:** Benjamin D. Branch; Kristin Wegner; Steven Smith; Darrell G. Schulze; Venkatesh Merwade; Jinha Jung; Ann Bessenbacher

**National Science Foundation (NSF) and Purdue University Funding:**

- Support from NSF, GAANN (Grant 
  #DGE1258680)
- Support from NSF, IIS-1351501 (Phase 1)
- Support from NSF, IIS-1351501 (Phase 2)
- Support from NSF, PHY-1306735
- Support from NSF, IIS-1351501 (Phase 3)

---

**Acknowledgements**

- Nicole Kong, GIS Specialist and Assistant Professor, Purdue University
- Frances Christman, Reference Assistant, Purdue University

---

**Support Mechanism**

- Pre-made Curriculum
- Support Mechanism
- Recruitment Factor (Wide Audience)

---

**Multifaceted Technologies**

- Modules
- Self-paced
- Real-time data collection
- In the Field (Portable)
- Universal tools in one Device
- Visualizations in the Field (DEM)