

**Proposal for Developing USGS Courses for Online Delivery
RFP dated 4/25/14 from Office of Organizational and Employee Development (OED)**

Short Course Title: Science DM Training

Descriptive Course Title: Science Data Management (DM) Training for USGS Scientists and Data Managers

Lead SME Name(s):

Viv Hutchison (PI)

USGS Core Science Analytics and Synthesis
Denver Federal Center
Building 810; Mail Stop 302
Denver, CO 80225
303.202.4227 (office)
571.481.7799 (work cell)
Supervisor: Mike McDermott (mmcdermo@usgs.gov)
Agrees to Participate: yes

Heather Henkel (Lead SME)

USGS (Ecosystems Program, located in Coastal and Marine Geology office)
600 4th St. South
St. Petersburg, FL 33701
Office (727) 502-8028
Cell (727) 412-4180
Supervisor: Ilsa Kuffner (ikuffner@usgs.gov)
Agrees to Participate: yes

Lisa Zolly (SME: Data Lifecycle, Metadata)

USGS Core Science Analytics and Synthesis
Denver Federal Center
Building 810; Mail Stop 302
Denver, CO 80225
303.202.4152 (office)
703.628.7345 (work cell)
Supervisor: Viv Hutchison (vhutchison@usgs.gov)
Agrees to Participate: yes

Cassie Thibodeaux (SME: Publish/Share)

USGS Lafayette Branch Library
700 Cajundome Blvd
Lafayette, LA 70506-3152

337-266-8692 (office)

Supervisor: Tom Doyle (doylet@usgs.gov)

Agrees to Participate: yes

Brief Explanation of ROI to the USGS for having your course delivered via the Internet: This course is appropriate for delivery via the internet mainly due to the impressive Return on Investment. Budget numbers to support a single 2-day, in-person workshop for 8 travelers (assuming local participants do not incur travel expenses) in Denver run approximately \$8000. (Average airfare: \$250; lodging: \$149/night; Per Diem: \$66 per day; Transportation: \$24 Round Trip per traveler) Simple math shows that supporting in-person workshops across the distributed USGS landscape is not a sustainable model.

Brief Curriculum Vitae for each SME:

Viv Hutchison

Contact Information:

- USGS - Core Science Analytics and Synthesis, Denver Federal Center, Building 810; Mail Stop 302 Denver, CO 80225
- Work Phone: (303) 202-4227; Work Cell Phone: (571) 481-7799; Email: vhutchison@usgs.gov

Education:

- Virginia Tech - Falls Church, VA., Certificate in Natural Resources (2004)
- University of Maryland - College Park., M.L.S. (2002)
- Claremont Colleges - Pitzer College, B.A. Political Science (1991)

Experience:

- 2013-present: Acting Science Data Management Branch Chief, USGS Core Science Analytics and Synthesis program
- 2010-present: Data Management Coordinator, USGS Core Science Analytics and Synthesis program
- 2012-present: USGS Data Management Policy Team
- 2010-present: Co-lead for USGS Community for Data Integration (CDI) Data Management Working Group; CDI Coordinating Committee
- 2009-present: Co-lead for DataONE Community Engagement and Education Working (CEE) Group (funded by NSF)
- 2002-2010: Metadata Program Coordinator, USGS National Biological Information Infrastructure program

Relevant Publications:

- Viv Hutchison, Thomas Burley, Michelle Chang, Thomas Chatfield, Robert Cook, Heather Henkel, Carly Strasser, Lisa Zolly. 2013. USGS Data Management Training Modules. USGS Community for Data Integration. <http://dx.doi.org/10.5066/F7RJ4GGJ>
- Faundeen, John L.; Burley, Thomas E.; Carlino, Jennifer A.; Govoni, David L.; Henkel, Heather S.; Holl, Sally L.; Hutchison, Vivian B.; Martín, Elizabeth; Montgomery, Ellyn T.; Ladino, Cassandra C.; Tessler, Steven; Zolly, Lisa S. USGS Open-File Report: 2013-1265, <http://pubs.er.usgs.gov/publication/ofr20131265>
- Heather Henkel, Viv Hutchison, Carly Strasser, Stacy Rebich Hespanha, Kristin Vanderbilt, Lynda Wayne, DataONE Educational Modules, <http://www.dataone.org/education-modules>

- San Gil, I; Hutchison, V; Palanisamy, G; Frame, M. 2010. Metadata Activities in Biology, *Metadata for Scientific Data, Journal of Library Metadata*; vol 10, issue 2: 99-118.
- Hutchison, Vivian. 2009. "Spatial Data Management through Metadata: Global Concepts, Formats, Tools and Requirements." *Spatial Complexity, Informatics, and Wildlife Conservation*. Ed. Samuel A. Cushman and Falk Huettman. New York: Springer. 223-232.

Heather Henkel

Contact Information:

- USGS - St. Petersburg Coastal and Marine Science Center, 600 Fourth St. South, St. Petersburg, FL 33701
- Work Phone: (727) 502-8028, Cell Phone: (727) 412-4180, Email: hhenkel@usgs.gov

Education:

- Carnegie Mellon Univ., M.S., Information Technology Management (2007)
- Univ. of Wisc.-Madison, B.S. Geology (1997)

Experience:

- 2012-present: USGS Data Management Policy Team
- 2011-present: DataONE Community Engagement and Education Working (CEE) Group (funded by NSF)
- 2011-present: Chair and Southeast Regional Data Coordinator for USGS Science Data Coordinator Network (SDCN)
- 2010-present: Co-lead on USGS Council for Data Integration (CDI) Data Management Working Group
- 2004-present: Data manager for "Everglades Depth Estimation Network" (EDEN)
- 2003-present: USGS Council for Data Integration (CDI) and CDI Coordinating Committee
- 1998-present: Principal Investigator: "South Florida Information Access (SOFIA) Metadata, Data Delivery; and Everglades Depth Estimation Network (EDEN) Geospatial System, and Website in Support of Greater Everglades Restoration" project

Recent Publications:

- Viv Hutchison, Thomas Burley, Michelle Chang, Thomas Chatfield, Robert Cook, Heather Henkel, Carly Strasser, Lisa Zolly. 2013. USGS Data Management Training Modules. USGS Community for Data Integration. <http://dx.doi.org/10.5066/F7RJ4GGJ>
- Faundeen, John L.; Burley, Thomas E.; Carlino, Jennifer A.; Govoni, David L.; Henkel, Heather S.; Holl, Sally L.; Hutchison, Vivian B.; Martín, Elizabeth; Montgomery, Ellyn T.; Ladino, Cassandra C.; Tessler, Steven; Zolly, Lisa S. USGS Open-File Report: 2013-1265, <http://pubs.er.usgs.gov/publication/ofr20131265>
- Heather Henkel, Viv Hutchison, Carly Strasser, Stacy Rebich Hespanha, Kristin Vanderbilt, Lynda Wayne, DataONE Educational Modules, <http://www.dataone.org/education-modules>

Cassie J. Thibodeaux

Contact Information:

- USGS, National Wetlands Research Center, Lafayette Branch Library, 700 Cajundome Blvd., Lafayette, LA 70506
- Work Phone: (337) 266-8692; Cell Phone: (337) 849-2671; Email: thibodeauxc@usgs.gov

Education:

- Louisiana State University - Baton Rouge, LA., MLIS (2009)

- University of Louisiana at Lafayette - B.S., Biology (2007)

Experience:

- 2014-present: Lead, USGS Lafayette Branch Library
- 2013-present: Project Lead, SE Data Management Community of Practice
- 2013-present: Technical Information Specialist, USGS Lafayette Branch Library
- 2010-present: Member of USGS Community for Data Integration (CDI) Data Management Working Group
- 2009-2013: Librarian, USGS National Wetlands Research Center Library
- 2006-2010: Content Manager of Central Southwest / Gulf Coast Information Node, USGS National Biological Information Infrastructure

Lisa S. Zolly

Contact Information:

- USGS Core Science Analytics & Synthesis Program, Denver Federal Center, Bldg. 810, MS 302, Lakewood, CO 80123
- Work Phone: (303) 202-4157, Email: lisa_zolly@usgs.gov

Education

- Associate's Certificate in Project Management, ESI (2010)
- MSc, Information Sciences, University of Tennessee, USA (1998)
- MA, English, Virginia Tech, USA (1994)
- BA, English, Virginia Tech, USA (1989)

Experience

- USGS Community for Data Integration (CDI) and CDI Data Management Working Group
- Association for Information Science & Technology (ASIST)
- Organization of Fish & Wildlife Information Managers (OFWIM)
- Data lifecycle management; Data and information retrieval; Data and information management;
- Metadata; Terminology systems development and applications; Information systems design; Digital libraries; Interface design and usability; Project management; Library reference services; Technical writing and documentation.
- 2014 - Co-PI, A Survey of USGS Scientists and Data Managers on the State of Data Management in the Bureau. CSAS-funded project for FY14.
- Sept. 2013 - Guest Lecturer, University of Tennessee, School of Information Sciences, "The USGS Data Lifecycle: Processes and Applications for Science Data Management." IS 590: Issues in Information Science - Federal Libraries.
- 2012-2013 - Collaborator, Development of Training Modules for Data Management Best Practices. CDI Funded Project for FY13. <<http://dx.doi.org/10.5066/F7RJ4GGJ>>.
- 2011-present - Co-investigator, Development of a USGS Data Management Web Site. CDI Funded Project for FY12. <<http://www.usgs.gov/datamanagement/>>.
- Oct. 2012 - Guest Lecturer, University of Tennessee, School of Information Sciences. "USGS Science Data and Data Products." IS 534: Government Information Sources.
- June 2012 - Guest Lecturer, University of Tennessee, School of Information Sciences. "Science Data Management, Analysis and Synthesis at USGS." IS 590: Issues in Information Science - Federal Libraries.

Recent Publications

- Faundeen, J, Burley, T, Carlino, J, Govoni, D, Henkel, H, Holl, S, Hutchison, V, Martin, E, Montgomery, E, Ladino, C, Tessler, S, **Zolly, L**, (2014). United States Geological Survey Science Data Lifecycle Model. Reston, VA, United States Geological Survey, Open File Report, 9 pp.

Problem Statement: USGS science data are among the most valuable assets of the organization. It is critical that we ensure our scientists and staff produce and manage data in such a way that at the completion of a project, the data continues to be accessible in useable formats, documented so it can be understood, and preserved properly for future uses. To work efficiently in the era of big data, the Survey needs to ensure the ability to re-use and analyze USGS data in ways not originally envisioned. Scientists will need to combine USGS data with other data in, for example, supercomputing environments, to perform world-class science and achieve the mission of the USGS. Therefore, the purpose of this online course is to continue to educate our scientists and our staff about best practices for science data management.

Course Summary/Description:

Description: Understanding and implementing best practices associated with the science data lifecycle are fundamental to basic principles of research. This course will introduce fundamental concepts of the science data lifecycle, including its definition and standards for USGS, the value of managing data according to a lifecycle, and best practices. The course will provide examples for implementing the best practices, in addition to helpful 'how-to' videos to demonstrate practical application of the lifecycle in the science research workflow.

Objectives:

Participant will be able to:

- define the components of the science data lifecycle
- describe the value in implementing best practices in data management
- specify fundamental data management best practices for working with science data

Target Audiences: Scientists, researchers, data managers, data miners, science technicians, and any staff working with science data.

Provide as much information as possible regarding course content and learning outcomes along with their applicability to online delivery: Training materials for this project will include content that will directly support the USGS Data Management Website and the USGS Science Data Lifecycle. These modules will build upon the existing three data management modules, and further explore the Science Data Lifecycle. Each module will introduce the student to a new topic and educate them on key points, best practices, and tools. Outcomes include:

- Informing and encouraging broadest possible application of data Management best practices
- Providing a consistent message to all scientists about data management practices

- Facilitating compliance with recent mandates from OMB and OSTP regarding open data access to well managed, federally funded research

Indicate any desired live events anticipated in your course: No live events will be included in the modules.

Indicate any student assessments: Each of the 3 modules will include video segments, captured as screencasts, which allow the user to learn how to execute a particular best practice in the data lifecycle. Student exercises will be included with the modules to reinforce learning.

Indicate if the Course Has Been Taught Before in the classroom: A version of this course has been taught in the classroom in 2011 in collaboration with the NSF-funded DataONE project. (Location: Santa Barbara, CA; Year: 2011; Participation: approximately 20 graduate students; Presenters: Viv and Heather were part of teaching team). There were many lessons learned, including the challenges in creating online modules intended for independent download, and the overlap of content when presenting the modules contiguously in a live classroom environment.

List Existing Course Materials: Existing course materials include two major sources:

- 1) DataONE data management training materials (developed as part of NSF-funded DataONE project, working group led by Viv Hutchison; Heather Henkel on content development team). Format: PowerPoint files with notes section annotated to allow anyone using the slides to present the course. <http://www.dataone.org/education-modules> .
- 2) USGS Data Management modules (based on DataONE modules, refined for use at USGS, enhanced with testimonials from USGS scientists), available at <http://www.usgs.gov/datamanagement/training/modules.php>.

TEL Format: Powerpoint slides converted into online videos. Slides, resources, and glossary terms available for download, similar to content available at <http://www.usgs.gov/datamanagement/>.

Intended Audience: Scientists, researchers, data managers, data miners, science technicians, and any staff working with science data are our primary audience. It is difficult to determine the number of potential course participants three years out, however, the three existing USGS online data management modules have received a tremendous amount of positive responses with requests for more training. For example, a recent email received from outside USGS commented,

“As I indicated at the Open Data Meeting last week, I believe that your data management modules are the single best contributions to DOI data stewardship that have been produced thus far. In my view, DOI needs to work from the ground up to change agency culture and make data management a permanent part of the scientific method. Your modules do just that. Thank you for developing them.”

-- Douglas R. Clark, Ph.D.; Bureau of Reclamation, TSC

Additionally, based on user statistics from the existing materials, the DataONE modules have been downloaded over 2,100 times since December 2012, and the USGS modules have been accessed over 1,000 times since their launch in October 2013.

Anticipated Results and Benefits: Restrictions on travel and conference attendance have severely limited the ability for employees to attend in-person training. By providing access to online modules, we will not only be able to reach a greater audience, but also provide training that is often not available, and when it is, carries a substantial cost. Additionally, in-person data management courses often involve coordination of multiple subject-matter experts to teach the course, affecting travel costs and other coordination factors.

Applicability of Course to Multiple USGS mission areas: This course has broad applicability for USGS science and will provide support for those in all Mission Areas and regions, as data management supports research and enables data integration by providing a foundation to the entire data lifecycle.

In this era of proliferating sensors for making observations and hardware for storage, the USGS will have to contend with large volumes of many complex data files ("big data"). To prepare for the onslaught of big data, the Survey needs to ensure the ability to re-use and analyze USGS data in ways not originally envisioned. Additionally, scientists will need to combine USGS data with other data in, for example, supercomputing environments, to perform world-class science and achieve the mission of the USGS. Therefore, it is imperative that we continue to educate our scientists and our staff about best practices for data management and begin the move to a USGS standard procedure.

As noted in a USGS Office of Water Quality Technical Memorandum, "the future value of environmental data is not fully known by those who collect it" (U.S. Geological Survey, 2008). This concept applies to all USGS Mission Areas answering research questions and conducting science for the Bureau.

Further, on February 24, 2013, the White House Office of Science and Technology Policy (OSTP) released the Executive Order, "Increasing Access to the Results of Federally Funded Scientific Research", and on May 9, 2013, the White House released the Executive Order, "Making Open and Machine Readable the New Default for Government." This Order, built upon the earlier OSTP Executive Order, and was accompanied by an Office of Management and Budget (OMB) Policy M-13-13 "Open Data Policy- Managing Information as an Asset". These memos call for increased support for public access to the results of research funded by the Federal Government. This includes any results published in peer-reviewed scholarly publications that are based on research that directly arises from Federal funds, and the data associated with that research.

Describe How the Course Will Support the USGS Science Strategy: USGS data are one of the most valuable assets of the organization, and it is critical that we ensure our scientists and staff produce and manage data in such a way that at the completion of a project, the data continues to be accessible in usable formats, documented so it can be understood, and preserved properly for future uses. In 2007, the Survey's Science Strategy Team put forth a 10-year vision for the agency to better serve the Department of the Interior (DOI) and the Nation through comprehensive integration of USGS capabilities. Clear, vetted standards, best practices, and procedures are necessary to govern the data

lifecycle and, when implemented across the organization, to support rapid data discovery, integration of data, and data provenance. In addition, the ability to respond rapidly to emergent needs for comprehensive data from disparate sources can play a critical role in investigating and understanding environmental challenges and allow the USGS to quickly respond to disasters. There is a critical need for better data management within USGS to realize the 2017 vision of integrated science, a premise that relies upon integrated data.

Timetable:

(Please note: the dates listed below assume a start date of July 1st. If the project is approved at a later date, then the dates will be adjusted accordingly.)

Activity	Start Date
Project awarded	July 1 (if proposal accepted)
Review recent USGS data management survey results. Determine which new modules should be created from existing materials based on community feedback.	July 1
Refine and streamline content from existing materials. Include best practice examples (existing) and refine for USGS as needed.	July 7
Develop scripts for each slide.	August 4
Identify where content benefits from video how-to segments and create strategy for development of video segments.	September 2
Determine and develop storyboard for content in how-to video clips based on examples.	September 8
Create how-to video clips	September 22
Incorporate slides, scripts, videos into TEL format.	November 10
Complete audio voiceovers for scripts in each slide	November 24
Develop supplementary documentation for all modules to be available for download.	January 5
Quality control modules and make refinements	January 20
Post modules to TEL course webpage and USGS Data Management Website	February 9

Estimated Project Budget:

	Hours*	Trips*	Nights*	Days PD*	Rate	Total
Henkel Salary	80				\$50.84	\$4,067.20
Airfare+baggage		2			\$425.00	\$850.00
Hotel			6		\$156.00	\$936.00
Per diem				7	\$66.00	\$462.00
POV Mileage for 2 trips, 50 miles X 4 legs X .56 per mile						\$112.00
Rental Car		2			\$265.45	\$530.90
					Travel:	\$2,890.90
					Travel OH:	\$544.93
					Travel total:	\$3,435.83
					Salary OH:	\$766.67
					Salary total:	\$4,833.87
					Travel and salary:	\$8,269.70

* Notes: Hours are based upon funding for 50% of 2 PP totaling 80 hours. Travel costs are based upon traveling in the morning, working 2 1/2 days, departing on 4th day. (Example, travel Monday morning, work Monday afternoon-Weds afternoon, return Thurs.) This is 3 nights per trip, 6 nights total for 2 trips. Per diem (PD) is 2 days plus 2* .75 for a total of 3.5 days of per diem for each trip, for a total of 7 for two trips.

Estimated Number of Course Modules/Lessons: Three (3) modules will be created for the project to follow the components of the science data lifecycle not already complete.

Estimated Frequency for Course Updates: Course content will be reviewed annually for needed updates.

Written Supervisory Approval: As instructed, included as attachment