The Potential Value and Impacts of a Data Decadal Survey [1]

Submitted by erinmr on Thu, 2013-06-06 14:28   Wednesday, July 10, 2013 - 10:30 to 12:00
Event: Summer Meeting 2013 [2]
Session Type: Plenary [3]
Collaboration Area: Data Preservation [4]
Abstract/Agenda:

A Decadal Survey is a broad study of a topic or area that is coordinated by the National Research Council (NRC). It provides community consensus on research goals and priorities for moving forward in a particular focus area over the next ten years. The studies are requested and funded by government agencies and other organizations and are used to prioritize research areas and focus efforts and resources accordingly. The decadal survey process has been shown to be a robust method for developing goals and to support objectives in pursuit of high priority science questions. The NRC has done a number of these studies, including its first decadal survey for Earth science in January 2007 at the request of NASA, NOAA, and the USGS.

Scientific data collected or modeled by government agencies are a public investment and should be stewarded appropriately to maximize the return. Today's science is increasingly collaborative. Important research questions increasingly span projects, discipline domains, and other boundaries. There is a growing interest in the repurposing of data far from the point of collection. Data collected long ago can become important today. Current data management and stewardship practices are not sufficiently meeting these pressures. Scientists today regularly cite the 80/20 rule for working with data: 80% of their effort is spent finding, understanding, acquiring, and putting the research data in a usable format, and 20% doing actual science.

Members of ESIP and representatives from the NRC Board on Research Data and Information (BRDI) have organized a cluster to investigate the need for and feasibility of conducting a Data Decadal Survey [ESIP DDS wiki] [5]. The survey would address overarching issues and research priorities in scientific data management and stewardship. Improved practices in this area could ultimately enhance scientific knowledge by increasing the meaningful availability of data and redirecting resources previously required for data discovery, acquisition, and formatting to performing actual science. When data sets can be easily analyzed and combined in novel ways then new scientific insights are more likely to occur and more quickly. Such a survey could address at the broadest level gaps in data management knowledge and practices that hold back scientific progress.

At the ESIP summer meeting we are organizing a panel discussion around this idea. We are inviting panelists to give us their vision of future data developments, and to discuss data management and stewardship (DMAS) topics and issues.

Questions for the Panel

- Do you have ideas for science that could be done with improved data management and stewardship (DMAS) that cannot be done now? If so, what?
- Do you know of serious gaps in DMAS that negatively impact science? If so, tell us about some.
- Do you see a need for a Data Decadal Survey? If so, why? If not, why not?
- What is your vision for the future regarding scientific data? Please be bold and include fanciful, idealistic, lofty, and even utopian ideas.

If in support of a survey:

- What do you think are the big questions around DMAS that must be addressed? What are the highest priorities?
- Is a Decadal Survey the right vehicle, or should it be something else?
- What do you think should be the scope of the survey
- Given that either extreme of depth or breadth is of less general use, what exactly should we target?
- Earth Science only or broader? For example, does it make sense to start "small" in the Earth Sciences and then generalize? Or, as data management problems across all domains are basically similar, should we start with the more general and perhaps go into greater detail later?
- Data? Software? Methodologies?
- What would be metrics for assessing survey success?
- A Data Decadal Survey is risky because the topic is extremely broad, the community of data users is vast and heterogenous (and can include commercial interests), and the outcome will not be a focused mission, facility, or research initiative. How should that risk be managed?

The panel will include:

- Stan Ahalt, Renci
- Dan Baker, LASP
- Michael Tiemann, Red Hat
- Todd Vision, Nescent

The panel will be moderated by Anne Wilson, LASP and chair of the Data Decadal Survey cluster

Stan Ahalt

Stanley C. Ahalt became RENCI director in September 2009 after serving as executive director of the Ohio Supercomputer Center (OSC) from 2003 to 2009 and as a professor in the department of electrical and computer engineering at The Ohio State University for 22 years. In addition to directing RENCI, Ahalt is a professor in the department of computer science at the University of North Carolina at Chapel Hill. Since coming to RENCI, Ahalt has increased RENCI's sponsored research portfolio and solidified RENCI's partnerships with the UNC School of Medicine, the UNC School of Information and Library Sciences, UNC's department of computer science, and various research units at North Carolina State and Duke universities. He is a member of the Board for National Lambda Rail, a major network for advanced research and innovation, and a member of Microsoft's Technical Computing Advisory Committee. He will begin a term as president of the Board of the Great Lakes Consortium for Petascale Computation (GLCPC) in fall 2011 and currently chairs the GLCPC Allocation Committee. Ahalt chairs the subcommittee on regional computing centers for the National Science Foundation Taskforce on High Performance Computing and was a key contributor to the NSF Data and Visualization and Campus Bridging Task Force reports, two of the six reports that comprise the NSF-wide Advisory Committee for Cyberinfrastructure reports published in April, 2011. He chaired the Coalition for Academic Scientific Computation (CASC) in 2009 and 2010 and has been a member of the Council on Competitiveness High Performance Computing Advisory Committee since 2004. While at OSC, Ahalt launched several model programs, including Blue Collar Computing, a national program to bring high performance computing to a wide spectrum of industries and applications, and OSCnet, a leading high-speed research network for K-12 schools, higher education and economic development. He also served as co-chair of the Ohio Broadband Council, the coordinating body for the state's initiative to extend the reach of the Broadband Ohio Network. Ahalt's research expertise involves neural networks, high performance computing, signal/image/video processing and object identification. He has authored or co-authored more than 120 technical papers and been principal investigator or co-principal investigator on research grants totaling nearly $17 million. Ahalt also served as the academic lead in the area of signal and image processing for the Department of Defense High Performance Computing Modernization Program. As a member of the Ohio State faculty, Ahalt co-founded the Information Processing Systems Laboratory. He received the OSU Lumley Research Award in 1997 and the OSU College of Engineering Research Award in 1999. A native of Virginia, Ahalt holds a Ph.D. in electrical and computer engineering from Clemson University and master's and bachelor's degrees in electrical engineering from Virginia Polytechnic Institute and State University.

Daniel N. Baker

Dr. Daniel Baker is Director of the Laboratory for Atmospheric and Space Physics at the University of Colorado-Boulder and is Professor of Astrophysical and Planetary Sciences and Professor of Physics there. He holds the Broad Reach Chair of Space Sciences at CU. His primary research interest is the study of plasma physical and energetic particle phenomena in planetary magnetospheres and in the Earth's vicinity. He conducts research in space instrument design, space physics data analysis, and magnetospheric modeling. Dr. Baker obtained his Ph.D. degree with James A. Van Allen at the University of Iowa. Following postdoctoral work at the California Institute of Technology with Edward C. Stone, he joined the physics research staff at the Los Alamos National Laboratory, and became Leader of the Space Plasma Physics Group at LANL in 1981. From 1987 to 1994, he was the Chief of the Laboratory for Extraterrestrial Physics at NASA's Goddard Space Flight Center. From 1994 to present he has been at the University of Colorado. Dr. Baker has published over 700 papers in the refereed literature and has edited eight books on topics in space physics. He is a Fellow of the American Geophysical Union, the International Academy of Astronautics, and the American Association for the Advancement of Science (AAAS). He is an elected member of the National Academy of Engineering. He currently is an investigator on several NASA space missions including the MESSENGER mission to Mercury, the Magnetospheric MultiScale (MMS) mission, and the Radiation Belt Storm Probes (RBSP) mission. He has won numerous awards for his research efforts and for his management activities including recognition by the Institute for Scientific Information as being "Highly Cited" in space science (2002), being awarded the Mindlin Foundation Lectureship at the University of Washington (2003) and being selected as a National Associate of the U.S. National Academies (2004). Dr. Baker was chosen as a 2007 winner of the University of Colorado's Robert L. Stearns Award for outstanding research, service, and teaching. In 2010, he was awarded the University of Colorado's Boulder Faculty Assembly Distinguished Research Lecturer Award. Dr. Baker was the 2010 winner of the American Institute of Aeronautics and Astronautics (AIAA) James A. Van

Allen Space Environment Award and Medal. Dr. Baker recently served on several national and international scientific committees including the Chairmanship of the National Research Council Committee on Solar and Space Physics. Dr. Baker served as President of the Space Physics and Aeronomy section of the American Geophysical Union (2002-2004) and he presently serves on advisory panels of the U.S. Air Force and the National Science Foundation. He was a member of the National Research Council's 2003 Decadal Survey Panel for solar and space physics and he was a member of the 2006 Decadal Review of the U.S. National Space Weather Program. Dr. Baker just completed service as chair of the National Academies 2013-2022 Decadal Survey in Solar and Space Physics.

Michael Tiemann

Michael Tiemann is a true open source software pioneer. He made his first major open source contribution more than two decades ago by writing the GNU C++ compiler, the first native-code C++ compiler and debugger. His early work led to the creation of leading open source technologies and the first open source business model. In 1989, Tiemann's technical expertise and entrepreneurial spirit led him to co-found Cygnus Solutions, the first company to provide commercial support for open source software. During his ten years at Cygnus, Tiemann contributed in a number of roles from President to hacker, helping lead the company from fledgling start-up to an admired open source leader. When Cygnus was acquired by Red Hat in 2000, Tiemann became Red Hat's Chief Technical Officer (CTO) before becoming its first Vice President of Open Source Affairs. In that role Tiemann provides technology, strategy, and policy advice to executives in the public and private sectors. Tiemann graduated from the Moore School at the University of Pennsylvania (Class of 1986) with a BS CSE degree, and later did research at INRIA (1988) and Stanford University (1988-1989).Tiemann has served on a number of boards crucial to the success of open source, including the Open Source Initiative (where he retired as President in April 2012), the Eclipse Foundation (where he was a founding member), the Embedded Linux Consortium (where he was a founding member), and the GNOME Foundation Advisory Board. Tiemann also provides financial support to organizations that further the goals of software and programmer freedom, including the Free Software Foundation and the Electronic Frontier Foundation. As part of his interdisciplinary approach to furthering the understanding and practice of open source, he accepted an appointment as a Visiting Scholar at the School of Infomration and Library Science at UNC Chapel Hill (2004-2005). He was also a founding member of the Board of Advisors for the Center for Environmental Farming Systems (2006-present). Tiemann has also remained active in the Creative Commons community, as both a sponsor of projects and promoter of the cause.

Todd Vision

"How many sizes are needed to fit all research data?"

 Todd Vision is Associate Professor of Biology at the University of North Carolina at Chapel Hill, where he has been since 2001, and serves as Associate Director for Informatics at the National Evolutionary Synthesis Center since 2006. He received his PhD from Princeton University in 1998 and did postdoctoral work at Cornell University and the US Dept. of Agriculture studying ancient patterns plant genome evolution using sequence data. In addition to his ongoing research in computational evolutionary biology, he has become increasingly involved in initiatives related to scholarly communication, particularly regarding data. He has been closely involved in the development of Dryad (datadryad.org [6]), a repository that works with a diverse array of journals to archive the data associated with the published scientific literature, and currently serves on its Board of Directors. He also serves on National Science Foundation Advisory Committee for Cyberinfrastructure, the leadership team of

DataONE (dataone.org [7]) and the Board of Directors of ORCID (orcid.org [8]).

Notes:
Link to one set of notes on session by Denise Hills (view and edit by all): http://goo.gl/exz9o [9]

The Potential Value and Impacts of a Data Decadal Survey [10] from ESIPFed [11] on Vimeo [12].

Attachments/Presentations: 0_panel intro.pdf [13]
1_ESIP_2013_Baker_0.pptx [14]
2_Ahalt_ESIP-preso-7.10.13-v2.pptx [15]
3_vision_esip_2013.pptx [16]
4_ESIP DDS Comments - Michael Tiemann.pdf [17]
Notes takers: **Name:** Denise Hills [18]

Organization(s): Geological Survey of Alabama [19]
Email: dhills@gsa.state.al.us [20]

Creative Common License: Creative Commons Attribution 3.0 License
Accepted:

Source URL: http://commons.esipfed.org/node/1536

Links:
[1] http://commons.esipfed.org/node/1536
[2] http://commons.esipfed.org/taxonomy/term/651
[3] http://commons.esipfed.org/session-type/plenary
[4] http://commons.esipfed.org/collaboration-area/data-preservation
[5] http://wiki.esipfed.org/index.php/Data_Decadal_Survey
[6] http://datadryad.org/
[7] http://dataone.org/
[8] http://orcid.org/
[9] http://goo.gl/exz9o
[10] http://vimeo.com/70358020
[11] http://vimeo.com/esipfed
[12] https://vimeo.com
[13] http://commons.esipfed.org/sites/default/files/0_panel%20intro_0.pdf
[14] http://commons.esipfed.org/sites/default/files/1_ESIP_2013_Baker_0_0.pptx
[15] http://commons.esipfed.org/sites/default/files/2_Ahalt_ESIP-preso-7.10.13-v2_0.pptx
[16] http://commons.esipfed.org/sites/default/files/3_vision_esip_2013_0.pptx
[17] http://commons.esipfed.org/sites/default/files/4_ESIP%20DDS%20Comments%20-%20Michael%20Tiemann_0.pdf
[18] http://commons.esipfed.org/node/1466
[19] http://commons.esipfed.org/taxonomy/term/352
[20] mailto:dhills@gsa.state.al.us