

# Earth and space science informatics infrastructure

Tom Narock · Melissa Cragin

Received: 28 October 2009 / Accepted: 11 November 2009 / Published online: 27 November 2009  
© Springer-Verlag 2009

## Introduction

Opportunities for countless scientific discoveries are anticipated from data intensive research and the application of computational methods and visualization tools to vast and growing data stores. These opportunities allow scientists to generate new questions, expose unseen or novel patterns, and to answer questions that challenge our global societies (Newman et al. 2003; Kelling et al. 2009). Data intensive research depends on the acquisition, organization and long-term management of these data collections. Such research also depends on the development and implementation of tools and systems for data integration, retrieval, and analysis. The emergent fields of informatics and data curation aim to meet these challenges, and each will require a skilled, professional workforce to meet the needs of the 21st century scientific enterprise.

Informatics is “the science of information,” its focus of study is “the representation, processing, and communication of information in natural and artificial systems” (Fourman 2002). An emerging application area for informatics research is within the Earth and space sciences—sometimes referred to as geoinformatics. The American Geophysical Union (AGU), a conglomerate of 50,000 Earth and space science researchers, has developed a focus group “concerned with

issues of data management and analysis, large-scale computational experimentation and modeling, and hardware and software infrastructure needs, which ultimately provide the capability to change data systems into knowledge systems that support the range of Earth and space science interests.”<sup>1</sup>

In August of 2009 a group of graduate students and established researchers met for a three-day geoinformatics workshop. There were dual goals for the workshop; first, there was an emphasis on further identifying and extending the geoinformatics community. The expertise needed for such a community to flourish ranges from domain experts in the Earth and space sciences, to computer science, library science, and information systems. To this end, the workshop had a strong focus on bringing together as diverse a group of attendees as possible. Of particular interest for participation were graduate students who would serve as the next generation of geoinformatists. The second goal of the workshop was to identify the infrastructural needs and common problems facing the geoinformatics community. We feel that the workshop was successful in achieving both of these goals, and the identified needs and problems are briefly outlined in subsequent sections. This special issue highlights selected presentations from the workshop, and the contributions of these papers are summarized in the following sections.

T. Narock (✉)  
Goddard Earth Sciences and Technology Center,  
University of Maryland,  
Baltimore County, MD, USA  
e-mail: Thomas.W.Narock@nasa.gov

M. Cragin  
Graduate School of Library and Information Science,  
University of Illinois at Urbana-Champaign,  
Urbana-Champaign, IL, USA

## Informatics as design science

### Software as an instrument

In the Earth and space sciences, instruments are built to test and validate scientific theories. This instrumentation is

<sup>1</sup> <http://essi.agu.org>

deployed on spacecraft, and at various locations around the world, in order to collect data relevant to a particular hypothesis. Weigel (2009) argues that software is a scientific product analogous to an instrument. In geoinformatics, and more generally in informatics as a whole, software allows us the ability to test not only *how* something might work, but also *why* it works. Just as an Earth scientist needs to understand an instrument in order to make sense of its data, a geoinformatist needs to understand the software that was deployed in a given environment. Weigel concludes that prevailing viewpoints need to change, that software needs to be considered research, and that methods need to be put in place for software to live on long after a project ends.

Such thinking echoes the Design Science research of Hevner et al. (2004) and March (March and Smith 1995) and such an approach to geoinformatics should be encouraged. The geoinformatics field is not only technical, but also socio-technical. One must understand IT systems and also the environment in which they operate. Building, evaluating, and justifying—concepts familiar to Earth and space scientists—need to be brought to the forefront in software development and deployment.

#### Software reuse

Fundamental to the survival of geoinformatics is the reuse of established knowledge and practices. A key theme repeated throughout the workshop was of software reuse. This theme involves the necessity of software repositories, documentation, and an understanding of how and why software works. To this end, Marshall et al. (2009) advocates for software readiness levels and demonstrates this approach in an Earth science software reuse portal.

#### Data infrastructure, trust, and organizational implications

Downs and Chen (2009) argue for better integration between data providers and the existing infrastructure. One solution they recommend is that data providers and archives must work together early and often. This echoes the general agreement across the information professions that new kinds of collaborations must happen at the local and institutional levels (Green and Gutmann 2007), as well as at the national level (Choudhury et al. 2009), which is now being undertaken by the National Science Foundation's DataNet initiative.<sup>2</sup>

<sup>2</sup> <http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm>

#### Emerging education initiatives

Another significant theme evident throughout the workshop was education, and the need for new programs to train the informatics and data curation workforce. Branch et al. (2009) presented research showing that gaps in current undergraduate education, particularly in areas related to geo-spatial technologies, are harming the efficiency and flow of the scientific enterprise. This undergraduate education has gaps in skill areas that then must be acquired on the job. Borne's (2009) discussion of "astroinformatics", presented the links between research and education, and a new program on Data Science. Importantly, it was noted that faculty are using a great deal more "real" data in their teaching. Librarians engaged in current data curation research projects have also identified this change.

The need for specialized programs to address current and emergent problems in informatics is congruent with the need for similar educational programs in data curation. Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Many organizations have begun to address these needs, and graduate programs are now being established. One such program is the data curation specialization in the Master of Science degree program at the Graduate School of Library and Information Science (GSLIS) at the University of Illinois.<sup>3</sup> The Data Curation Education Program is a specialized curriculum that focuses on data collection and management, knowledge representation, digital preservation and archiving, data standards, and policy. Data curation includes not only data archiving and digital preservation, but also active management and appraisal of data over the life-cycle of scientific interest. Students in the program are expected to enter the workforce ready to take responsibility for assimilation and management of data in ways that add value and promote sharing across institutions and disciplinary specializations.

#### Summary and conclusions

The workshop brought together people with a wide range of experiences and research efforts, and out of this diversity emerged several fundamental issues and overlapping objectives. The geoinformatics community is still nascent, yet we hope that through the workshop we have begun to identify the key infrastructural challenges and research needs, as well as some of the people who will address them. In this special issue we report on the state of geoinformatics infrastructure from a set of diverse viewpoints, and we

<sup>3</sup> [http://www.lis.illinois.edu/programs/ms/data\\_curation.html](http://www.lis.illinois.edu/programs/ms/data_curation.html)

believe that the collaborations, discussions, and research that resulted from the workshop will pave the way for addressing future challenges. We encourage readers to view all presentations from the workshop at: <http://essi.gsfc.nasa.gov/presentations.html>

**Acknowledgements** We would like to thank the program committee, attendees, and presenters of the aforementioned workshop for enabling a thorough and engaging discussion of geoinformatics topics. In addition, we are indebted to the authors, reviewers, and editors who enabled this timely special issue. Moreover, we would like to thank the University of Maryland, Baltimore County Department of Information Systems for hosting the workshop, as well as Tetherless World Constellation at Rensselaer Polytechnic Institute for generously providing financial support.

## References

- Branch D, Maybin B, Frink K (2009) Implications of advancing informatics pipeline. Earth and Space Science Informatics Workshop, University of Maryland, Baltimore County, August 3–5, 2009, presentation available at: <http://essi.gsfc.nasa.gov/pdf/Branch.pdf>
- Borne K (2009) Astroinformatics: A 21st Century Data-Oriented Approach to Astronomy Research and Education. Earth and Space Science Informatics Workshop, University of Maryland, Baltimore County, August 3–5, 2009, presentation available at: <http://essi.gsfc.nasa.gov/pdf/Borne.pdf>
- Choudhury GS, DiLauro T, Evans M (2009) Sustainable data curation infrastructure. Proceedings of the Indo-US Workshop on International Trends in Digital Preservation, pp. 119–122, March 24–25, Pune, India.
- Downs R, Chen R (2009) Designing Submission Services for a Trustworthy Digital Repository of Interdisciplinary Scientific Data, Earth and Space Science Informatics Workshop, University of Maryland, Baltimore County, August 3–5, 2009, presentation available at: <http://essi.gsfc.nasa.gov/pdf/Downs.pdf>
- Fourman M (2002) Informatics. Informatics Research Report EDI-INF-RR-0139. Available at: <http://www.informatics.ed.ac.uk/publications/online/0139.pdf>
- Green AG, Gutmann MP (2007) Building partnerships among social science researchers, institution-based repositories and domain specific data archives. OCLC Syst Serv: International Digital Library Perspectives 23(1):35–53
- Hevner A, March S, Park J, Ram S (2004) Design science in IS research. MIS Quarterly 28(1):75–105
- Kelling S, Hochachka WM, Fink D, Riedewald M, Caruana R, Ballard G, Hooker G (2009) Data-intensive science: a new paradigm for biodiversity studies. BioScience 59(7):613–620
- March ST, Smith GF (1995) Design and natural science research on information technology. Decis Support Syst 15(4):251–266
- Marshall J, Downs R, Samadi S (2009) Relevance of reuse in building advanced data processing systems, earth and space science informatics workshop, University of Maryland, Baltimore County, August 3–5, 2009, presentation available at: <http://essi.gsfc.nasa.gov/pdf/Marshall.pdf>
- Newman HB, Ellisman MH, Orcutt JA (2003) Data-intensive e-science: frontier research. Communications of the ACM 46(11):68–77
- Weigel B (2009) “VxOware: Software for managing scientific metadata”, Earth and Space Science Informatics Workshop, University of Maryland, Baltimore County, August 3–5, 2009, presentation available at: [http://essi.gsfc.nasa.gov/pdf/Weigel\\_VxOware.pdf](http://essi.gsfc.nasa.gov/pdf/Weigel_VxOware.pdf)