



National Snow and Ice Data Center

449 UCB
University of Colorado
Boulder, CO 80309-0449
Phone: +1.303.492.6199
Fax: +1.303.492.2468
Email: nsidc@nsidc.org
<http://nsidc.org>

IIS-GEO preparatory statement: Data Curation and Informatics research connecting computer and geo science

Lynn Yarmey, Toni Rosati, Ruth Duerr, Colleen Strawhacker, and the NSIDC Informatics Team

Bridging between the geo and computer sciences, Data Curation and Informatics research offer a unique contribution to innovation in intelligent and information systems research. Lessons learned, often socio-technical in nature, from progress in informatics and data curation can inform future development efforts.

From a computer science perspective, the possibilities and value of data and metadata brokering have been showcased in diverse venues over recent years. By honoring both the distributed nature of local expertise and infrastructure as well as the need for global standards, brokering mediates diverse content and offers new avenues for data reuse in effective and efficient ways. Web crawling for science data moves the community closer to ubiquitous data discovery. With relevant data found, reading and analysis tools support data use and integration for new scientific discoveries.

All of these advances would not have been possible without the years of challenging grassroots work to change cultural norms and develop policies and infrastructure to promote and implement open data sharing practices. Only when data and metadata are shared and aggregated can we learn through comparisons. As federal data management and sharing mandates mature and evolve, new technologies can better support the now more centralized work of discovering, aggregating, translating, processing, analyzing, and visualizing scientific geodata to advance our understanding of changing global systems. Trends include but are certainly not limited to: shifting from centralized to distributed models supported by web services, collaboratively co-designed cyberinfrastructure partnering scientists with cyberinfrastructure experts, user-centered design and usability testing, and interdisciplinary data integration supported by semantically-enabled tools and standards.

That said, there is still considerable work to do. While cloud-based services like Dropbox, Skype, iPython, and Google Earth Engine have improved virtual collaboration among science teams, data management is still a very manual and resource-intensive exercise. Data are much more discoverable and accessible than in the past, however challenges such as a lack of indicators describing the trustworthiness and relevance of data still hamper reuse. The metadata necessary to reach easy and accurate reuse require further cultural shifts and development of new expertise. Updated reward systems remain necessary to fully realize the potential of open data. Together, we can continue moving the scientific enterprise forward.



University of Colorado at Boulder

Cooperative Institute for Research in Environmental Sciences

Member, World Data System

