Session: C21 Poster: T256A

Managing very diverse data for complex, interdisciplinary science

Mark Parsons[†]; ÿystein God¯y; Ellsworth LeDrew; Taco de Bruin; Bruno Danis; Scott Tomlinson; David Carlsonn
[†] University of Colorado, USA
Leading author: parsonsm@nsidc.org

Much attention has been given to the challenges of handling massive data volumes in modern data intensive science. This paper examines an equally daunting challenge - the diversity of interdisciplinary data, notably research data, and the need to interrelate these data to understand complex systemic problems such as environmental change and its impact. We use the experience of the International Polar Year 2007-2008 (IPY) as a case study to examine data management approaches seeking to address issues around complex interdisciplinary science. We find that while technology is a critical factor in addressing the interdisciplinary dimension of the data intensive science, the technologies developing for exa-scale data volumes are not the same as what is needed for extremely distributed and heterogeneous data. Research data will continue to be highly heterogeneous and distributed and will require technologies to be much simpler and flexible. More importantly there is a need for both technical and cultural adaptation. We describe a vision of discoverable, open, linked, useful, and safe collections of data, organised and curated using the best principles and practices of information and library science. This vision provides a framework for our discussion and leads us to suggest several short and long-term strategies to facilitate a socio-technical evolution in the overall science data ecosystem.