Geophysical Research Abstracts
Vol. 18, EGU2016-10482, 2016
EGU General Assembly 2016

Challenges and Research of Data Intensive Infrastructures That Enhance Geoscience Knowledge

Emily Law, Dan Crichton, and Thomas Huang
NASA JPL, United States (emily.law@jpl.nasa.gov)

Geoscience digital infrastructures today are primarily organized around the capture and stewardship of data from specific missions, campaigns, and/or research efforts, addressing fundamental digital infrastructure challenges with regard to common standards, tools, access patterns, and discovery.
The Data Science Team at JPL believes that the usability of data, and how to transform geoscience data from bits to information and then to knowledge is important for supporting scientific research. A large gap in current infrastructures has been the focus on localized data management and analysis versus a holistic consideration of cross domain systems and how they can better advance the broader geoscience community. The integration of data from multiple sources (e.g., insitu, remote sensing, etc.,) independent of their geographic locations around the world, in particular, in moving from isolated data analysis to knowledge discovery through the use of a big data analytics and cross domain data integration presents additional significant challenges. These include making data available from multiple sources and integrating those data using intelligent algorithms and methods, and presenting the knowledge to end users (e.g., scientists, decision makers, etc.,) through innovative visualization. The process of provisioning and analysis of geoscience data that are voluminous, highly distributed and heterogeneous, and extracting the meanings from the analysis, is highly complex and requires new breakthrough approaches to overcome their massive scale and geographic distribution.
This presentation further illustrates these challenges and discusses some of our prior works and research of digital infrastructures that address these challenges. It also presents an overview of their core components including knowledgebase system, cloud-based big data platform, novel data search and discovery, as well as data science framework.