EXECUTIVE SUMMARY: EARTHCUBE REAL-TIME WORKSHOP RESULTS
June 17 and 18th, 2013
Boulder Colorado

Organizing Committee:
Mike Daniels, NCAR (Chair)
V. Chandrasekar, Colorado State University
Sara Graves and Sandra Harper, University of Alabama - Huntsville
Branko Kerkez, University of Michigan
Frank Vernon, Scripps Institution of Oceanography/University of California - San Diego

Workshop Breakout Teams:
Tim Ahern, Incorporated Research Institutions for Seismology
Jennifer Arrigo, CUAHSI
Janet Fredericks, Martha's Vineyard Coastal Observatory/Woods Hole Oceanographic Institution
Alexandria Johnson, Purdue University
Kate Keahey, Argonne National Laboratory / University of Chicago
Charlie Martin, NCAR
Jim Moore, NCAR
Mohan Ramamurthy, UCAR Office of Programs
Siri Jodha Singh Khalsa, NSIDC
Greg Stossmeister, NCAR

Earth Cube Workshop Title: Integrating Real-time Data into the EarthCube Framework

Introduction: The primary findings of this workshop contend that real-time data have the potential to transform the geosciences by enabling adaptive, feedback-driven experimentation and improved societal decision making. Once the cyberinfrastructure is in place, the ability to act on and analyze data as it is collected will enable the discovery of scientific phenomena that may otherwise go unobserved and unexplained. Specifically, instantaneous feedback from sensing devices will enable real-time hypothesis testing, improving the quality of ongoing interdisciplinary experiments through adaptive reaction, while facilitating unaliased observations of space/time scales. Continuous, real-time data will lead to new, potentially unanticipated data discoveries as scientists respond to emerging, in-situ phenomena. Furthermore, real-time data is relevant and engaging to both scientists and the public. The utility of real-time data to societal decision makers cannot be understated, as it will enable new suites of operational support and disaster mitigation systems.

There exists a significant opportunity to realize this potential through improved sensors, better networks, advanced algorithmic techniques, and on-demand availability of computer resources. A concerted effort is required across the geoscience and cyber infrastructure communities to define the unique nature of real-time data streams and their role in the future of *EarthCube* and broader NSF initiatives. In summary, the integration and development of real-time capabilities will have significant transformative effects on Earth science in the next five to fifteen years.

The EarthCube real-time data workshop took place June 17th and 18th in Boulder, Colorado, and involved 76 participants from a variety of state and federal agencies, academic institutions and industry. A broad spectrum of geosciences was represented, including but not limited to Hydrology, Oceanography, as well as the Earth, Atmospheric, Space, Polar and Cyberinfrastructure sciences. To motivate breakout sessions and to provide use-case ideas, prominent experts from these domains presented real-world examples of the need for real-time data in their experimental campaigns. The major scientific and technical challenges behind the integration of real-time data into the EarthCube framework, as well as motivating use case scenarios were outlined by the workshop participants and are summarized below.

SCIENCE ISSUES AND CHALLENGES

1. **Important science drivers and challenges:** Participants identified several high-priority science questions that will be the focus of interdisciplinary efforts during the next 5-15 years (list 3 to 6).

    - How can we better use real-time data to understand the processes of high impact events or phenomenon and translate that knowledge to better response procedures? Examples of critical cases include, but are not limited to:
        - Improved hurricane track and intensity forecasting; prediction and response to coastal inundation and shoreline breaches
        - Better understanding of tornado and severe convective storm genesis and warning
        - Earthquake and tsunami prediction
        - Better understanding, predicting, and managing of Hydrologic Extremes, e.g. flash floods
        - Early detection of harmful algae blooms
        - Prediction of large solar flare events for assessment of damage satellites.

    - How can we better understand scientifically compelling phenomenon with adaptive real-time, feedback-driven science? The following strategies optimize the scientific values of our measurements and enable new discoveries:
        - Dynamic sampling strategy to collect, analyze, and respond to real-time data
        - Response examples: Moving platforms, changing scan strategies, adjusting flight patterns, automatic adjusted of instrument signal processes, deployment of additional instruments, etc.
        - Using models in conjunction with adaptive strategies to improve sampling
        - Real-time awareness of instrument status to support rapid response to issues and improve data quality
        - Instrument validation (is it responding to its environment and can we adjust the instruments to improve the response?)
        - Tools that enable broad communication and collaboration during real-time mission oriented research
        - Detection and discovery of new, unexpected phenomena that need to be explored further
        - Tracking and sampling of transient phenomena

2. **Current challenges to high-impact, interdisciplinary science:** Several themes emerged as consistent challenges faced within/across the involved discipline(s) (list 3 to 6).

- Interoperable streaming protocols and metadata (including consistent and accurate time stamping and spatial coverage) for real-time data streams across the geosciences domain do not exist.

- There are few, if any, mechanisms and processes in place to assess the quality of real-time geosciences data.

- Visualization tools for interdisciplinary real-time data of varying spatial and temporal coverage need to be developed.

- Valuable real-time data streams are often not integrated with downstream decision support systems used by emergency managers, etc.

- Real-time data streams need better connections to prediction models and/or systems that produced derived products.

- The scientific community generally does not properly address real-time data at the same level of archival data or Big Data in terms of data management plans and other data-focused initiatives.

TECHNICAL INFORMATION/ISSUES/CHALLENGES

1. **Desired tools, databases, etc. needed for pursuing key science questions with brief elaboration:**

    - Improved community infrastructure: access to improved communication infrastructure, on-demand computing and protocols for data exchange

    - Metadata generation for real-time data streams and tracking of provenance

    - Real-time signal processing, calibration, and quality control: existence of standardized software libraries

    - Real-time computing: software that provides the ability to process, produce, and transmit derived products in real time.

    - Tools for integrating and assimilating real-time observations: from differing geospatial and temporal resolutions

    - Playback tools for re-creation and analysis of phenomena and the observed environment of past experiments

    - Frameworks and secure mechanisms for remote operation of instruments

    - Real-time visualization of observations made at different temporal and spatial scales

- Data discovery and access including data subsetting of large bandwidth streams
- Rendering of observations with widely different time scales for real-time displays
- Decision support tools and integration with tools for emergency management
- Engine/middleware/platform that will combine all these capabilities for the community
- Developing networks for dissemination - including social media, apps, and user driven interfaces/portals including citizen science, crowd sourcing and open data access
- Mechanisms to discover software and hardware for real-time acquisition and processing and to provide guidelines in implementing real-time capabilities (eg, SUB/PUB real-time streams, buffering data for remote access, etc), education

COMMUNITY NEXT STEPS

1. **List of what your community needs to do next to move forward how it can use EarthCube to achieve those goals:**

- Community Development:
  - Assess system interoperability between geosciences real-time data providers and users
  - Establish best practices with scientists and CI experts across domains engaged in real-time data systems
  - Share knowledge, tools and approaches to real-time data experts across the geosciences
  - Refine requirements needed for real-time data streams to connect to downstream decision-making tools and processes
  - Increase awareness of the real-time data streams that are in existence among the geosciences community to facilitate new uses of these data

- Prototyping:
  - Pilot projects, demonstration testbeds, identification and development of real-time data capabilities
  - Build real-time stream translation tools across geosciences disciplines
  - Develop of a prototype framework for real-time control of instruments that can be more generally applied to the geosciences
  - Respond to missing capabilities in Section 2 above such as real-time quality control mechanisms, real-time metadata standards or real-time visualization tools that span geosciences data of varying spatial and temporal domains
  - Begin work to develop a “universal real-time infrastructure” where data streams are captured, organized and made quickly available, making it much more likely that they will be adopted by stakeholders who have noticed a new phenomena and want to examine it in the context of current events

- Capacity Building:

- Work with undergraduate and graduate students to engage them and popularize real-time science and data
- Develop the next generation workforce by exposing students to real-time data and its importance to science
- Explore and engage the private sector to meet the collective needs of the real-time geosciences data community