Chair's Comments
2nd Meeting of the
Big Data Task Force

June 28, 2016

Goals of this meeting

1. Hear from the Goddard projects for data archiving and large-scale computing.

2. Continue discussions of the NSF Big Data Hubs project

3. Converge on topics for TF studies over the next 18 months.

4. Generate findings and recommendations for the BDTF report to the Science Committee (July 26).

Chair's report to the BDTF

June 28, 2016

Chair's report to the BDTF

- Attended the Science Committee meeting in March 10-11 and gave the TF's 1st report.
- Visited Goddard Institute for Space Science in NYC – discussed their "big data" needs.
- Visited the PI and Executive Director of the North East Big Data Hub at Columbia University. Discussed how NASA PIs can participate in their project.
- Considered topics for BDTF projects.
- Tuned in to two "Big Data" webinars.

Introduce Dr. Larry Smarr

Dr. Larry Smarr
Director, CalIT^2

http://www.calit2.net

"Creating a Science-Driven Big Data Superhighway"

Lists of Topics

1. From Erin Smith's briefing to the BDTF Feb 2016:

<u>Big Data motivations</u>
- NASA has been charged with making NASA Earth observations data interoperable with other agencies such as NOAA and USGS.
- Providing interoperable data sets facilitates potentially new science, better use of data for decision support and potentially actionable science information.
- Making data more accessible can result in its use for completely new purposes.
- Large data sets demand understanding of architecture and tools for current and future needs.

<u>Usability of Data, Datasets and Databases</u>
- Management and Access
- Utilization (including near real-time)
- Interoperability (e.g. integration of heterogeneous datasets, harmonization of data with different timeframes)

<u>Analysis</u>
- Data mining/analysis of large datasets
- Algorithm and statistics development

<u>Storage</u>
- Data curation
- Archiving tools and technology

<u>Visualization (e.g. hyperwall)</u>

<u>Utilizing State-of-the-Art IT Systems and Tools</u>

More topics

2. Suggested Topics from presentations to the BDTF in Feb:

- Joint science return from WFIRST, LSST, & Euclid
- Data management plans for ROSES research proposals
- Evolution of NASA's extended data stores

3. Subjects for topics that have come across my desk since the Feb. meeting:

- Workflows for Earth system modeling
- Data analytics architectures for climatology studies
- Need for improved data/science analysis methodology

From my email – June 8th, 2016

- My goal is to have by noon of the 30th a list of 3-4 topics that will we begin pursuing. I'll ask for volunteers to take the lead (or co-lead) for each topic.
  - The first duty after the meeting will be to draft a short (1-2 pages) note defining the topic, the approach the TF should take for addressing the topic and the outcome that you would like to see if the results of our work go to the Science Committee and NASA.
  - We will finalize these drafts at our fall meeting.

- Please look at the list above. Select one or two that interest you and be prepared to advocate for those topics at our next meeting in three weeks. If you have suggestions for another topic, please send me a short note with a title and a couple sentences describing the topic. And I will add it to the list.