

[Visioning for the Science Data Enterprise](#) [1]

Submitted by Krbm on Wed, 2014-01-08 10:20 Wednesday, January 8, 2014 - 10:30

Event: [Winter Meeting 2014](#) [2]

Session Type: [Plenary](#) [3]

Notes:

Panel: Jeff de La Beaujardiere, Sky Bristol, George Stawn, Jeff Walter

Jeff de La Beaujardiere NOAA

5 Trends

1. can't solve problems in isolation
 2. rapid growth in data volumes
 3. users who want answers not data
 4. work anywhere with anyone
 5. things need to be reliable
- need to move to cloud computing
 - bring science to data
 - current cost model (GB/day) – need public/private partnership

Sky Bristol, USGS

- what is the science data enterprise?
 - We are always behind the curve with technology
 - USGS reorganized away from geology, hydrology.... to ecosystem, climate, land-use change...
- o Modular science framework
- We have developed and surpassed Star Trex' tricorder – but apps need documented efficiency
 - Need to not fight over who's domain is what/who has data
 - Define success as whole of earth science (get people working together)
 - Read more science fiction to make it reality tomorrow

Jeff Walter, NASA

- Physical infrastructure
- o NASA has data located in physical center for different expertise
- Library or curation aspects of data management
- o Long term preservation, metadata and documentation, formats

Visioning for the Science Data Enterprise

Published on Commons (<http://commons.esipfed.org>)

- o Provenance and linkage between data, persistent identifiers, citations
- o Publication paradigm is changing to incorporate data and code
- o Dark data problem
- o Old data vs. new data, known unknowns vs. unknown unknown
- How do we realize the full potential of all the data

George Strawn, NCO/NITRD

- Started group because presidential mandate of Big Data as an important topic
- It is now standard for federal data to be public (mainly administration data) - open government data
- Scientific publications - need to be publically available in an open format if funded by federal government
- Scientific data that is collected and stored for federally funded research will be shared as default
- o These 3 new mandates will be difficult to implement
- Big data group will be active in next few years - set up a series of seminars/colloquia where groups can come together to discuss how to meet OSTP requirements for data

OPEN QUESTION PERIOD

- Karl Benedict - what keeps you up at night that can be address by data science?
- o Sky - everything is miscellaneous concept - not yet scaled how to track provenance of how data is used in the lifecycle

§ How we quantify uncertainty in data

- o Jeff D.- better and trustable automation
- o Jeff W. - data interoperability - content is much deeper than what meets the eye - what kind of science goes into resampling or conversion to combine data sets
- o George - usually talking about a subset of science - overall issue is very large and cross boundaries
- Peter Fox - what do you see as the existing mechanisms for innovations and where they are going (both private and academic communities) (how to balance desire for innovation and operating existing - disturbing by innovation)
- o Jeff D. - NOAA has this problem (ex. Weather information) - modernize in place - cost is an issue
- o Jeff W. - ability to integrate disruptive technology is proportional to risk tolerance (big enterprises) - ability is limited by budget
- o Sky - seen interesting disruptions out of small projects - able to provide better capability - small things, handled appropriately can have substantive changes
- Open Access Science publication mandate - what is the time lines, what will the agencies have

planned, what can ESIP do with

- o George - agencies have already turned in short description of how
- o Jeff W. - defer to Martha - NASA worked as a team (across agency) to put in there description - more comprehensive long tail data put in archive
- o Sky - panel Thurs at 10:30, ESIP is small in publishing world - committed to getting publishers to change there access
- Kevin Ashley - public/private partnerships - how well prepared are you to get benefits for both groups?
- o Jeff D. - existing partnership with weather service - make available data and have industry do the interesting work with - have systems where people pay for access to the data (to cover the maintenance) - but service is still free
- o Sky - like seeing different uses of USGS data crop up - need formal or informal partnerships to show what can happen/does with the data
- o Jeff W. - goal is to get data to the point - OSTP wants to create spark of "what can you do with our data"
- ? - At what point to we benefit from incorporating the National Archive in this discussion?
- o George - Archive is a member of NITRD
- o Jeff D. - NOAA has data centers that are on the 100+ yr data scale
- Data study working group - what would be the best way to move forward in a significant manner
- o Jeff W. - not sure if NRC study is the right answer - goal is something agencies can hang hat on to drive agenda - not sure of any other agency to accomplish the goals
- o Sky - USGS has launched line item program based on NRC study
- o Jeff D. - implementation (may not be better than NRC study), if it doesn't work out then just shut it down - risk free (except for money) - lets try things
- o Karl - ESIP came out of an NRC study
- ? - How to balance pressure from private sector products vs agency innovation
- o Jeff D. - RFI will be interesting because it will give insight into this issue
- o Sky - need innovative ways to get our data out - don't want to be scared of innovation because it is part of the agency mission - need to stick to mission
- o Jeff W. - agree - keep eye on private sector to avoid redundancy
- o George - continuous balancing act in congress of who should do what and who should pay or it - long term/life changing research is not done well in the private sector

Session Leads:

Name: [Karl Benedict](#) [4]

Organization(s): [University Libraries](#) [5], [University of New Mexico](#) [6]

Presenters:

Name: [Jeff deLaBeaujardiere](#) [7]
Organization(s): [NOAA](#) [8]
Email: jeff.delabeaujardiere@noaa.gov [9]

Name: [Sky Bristol](#) [10]
Organization(s): [U.S. Geological Survey](#) [11]

Name: [Jeff Walter](#) [12]
Organization(s): [NASA Goddard Space Flight Center](#) [13]
Email: jeff.walter@nasa.gov [14]

Name: [George Strawn](#) [15]
Organization(s): [NCO/NITRD](#) [16]

Notes takers:

Name: [Kelly Monteleone](#) [17]
Organization(s): [University of New Mexico](#) [6]
Email: krbm@unm.edu [18]

Creative Common License: Creative Commons Attribution 3.0 License
Accepted:

Source URL: <http://commons.esipfed.org/node/2004>

Links:

- [1] <http://commons.esipfed.org/node/2004>
- [2] <http://commons.esipfed.org/taxonomy/term/1029>
- [3] <http://commons.esipfed.org/session-type/plenary>
- [4] <http://commons.esipfed.org/node/297>
- [5] <http://commons.esipfed.org/taxonomy/term/1469>
- [6] <http://commons.esipfed.org/taxonomy/term/297>
- [7] <http://commons.esipfed.org/node/1403>
- [8] <http://commons.esipfed.org/taxonomy/term/242>
- [9] <mailto:jeff.delabeaujardiere@noaa.gov>
- [10] <http://commons.esipfed.org/node/323>
- [11] <http://commons.esipfed.org/taxonomy/term/397>
- [12] <http://commons.esipfed.org/node/719>
- [13] <http://commons.esipfed.org/taxonomy/term/246>
- [14] <mailto:jeff.walter@nasa.gov>
- [15] <http://commons.esipfed.org/node/2003>
- [16] <http://commons.esipfed.org/taxonomy/term/1232>
- [17] <http://commons.esipfed.org/node/430>
- [18] <mailto:krbm@unm.edu>