

# News

BESSIG members should be able to add new News items. To do so, from the column on the left of any page, click through Edit -> New -> News. Please let Anne or Ransom know if this does not work for you.



[Subscribe to the news RSS feed](#)



[The Future of BESSIG](#)

[Soren Scott](#) posted on Feb 11, 2016

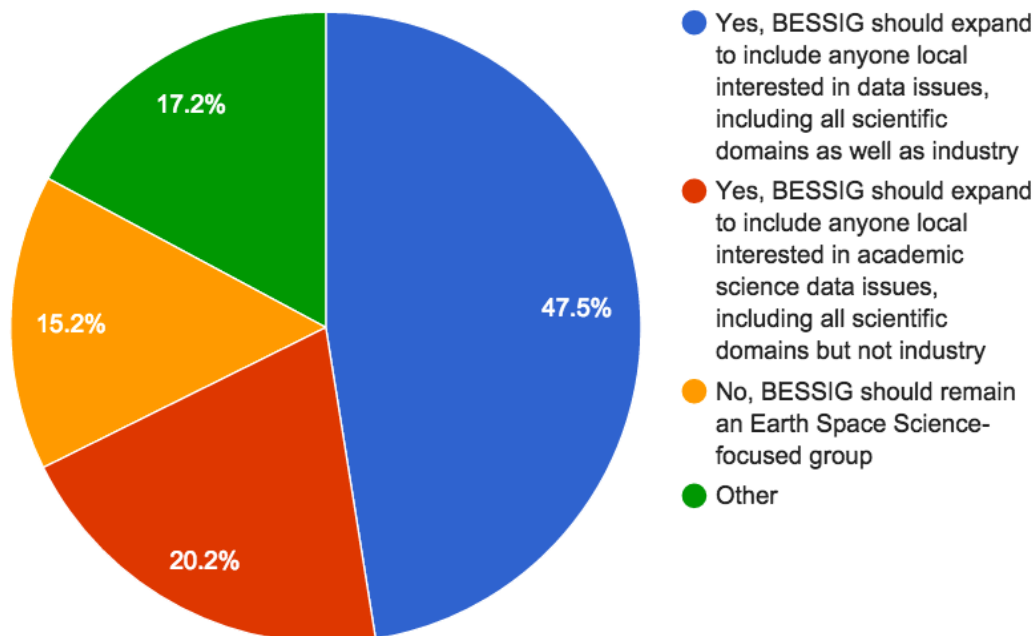
Thank you all for participating in the recent Future of BESSIG survey (and thank you, Lynn, for putting it together). We had a great discussion at the January meeting and a good response to the survey.

For those that missed it, we talked about three areas: the scope of the group, the name of the group, and a possible funding approach. A number of BESSIG members learned through the AGU ESSI group that the ESSI acronym isn't well-known or well-understood in the larger geoscience community. Locally, the question came up - do we face the same issues with the BESSIG brand and what would be potential alternatives. As we considered a potential name change, we also discussed expanding or changing the scope of the group to include data science and informatics beyond the current ESSI scope.

Finally, Anne once again raised the question of community funding support, brought to the forefront by the new requirements put forth by our previous meeting place and the difficulties in finding an acceptable meeting space in Boulder.

On to the results!\*

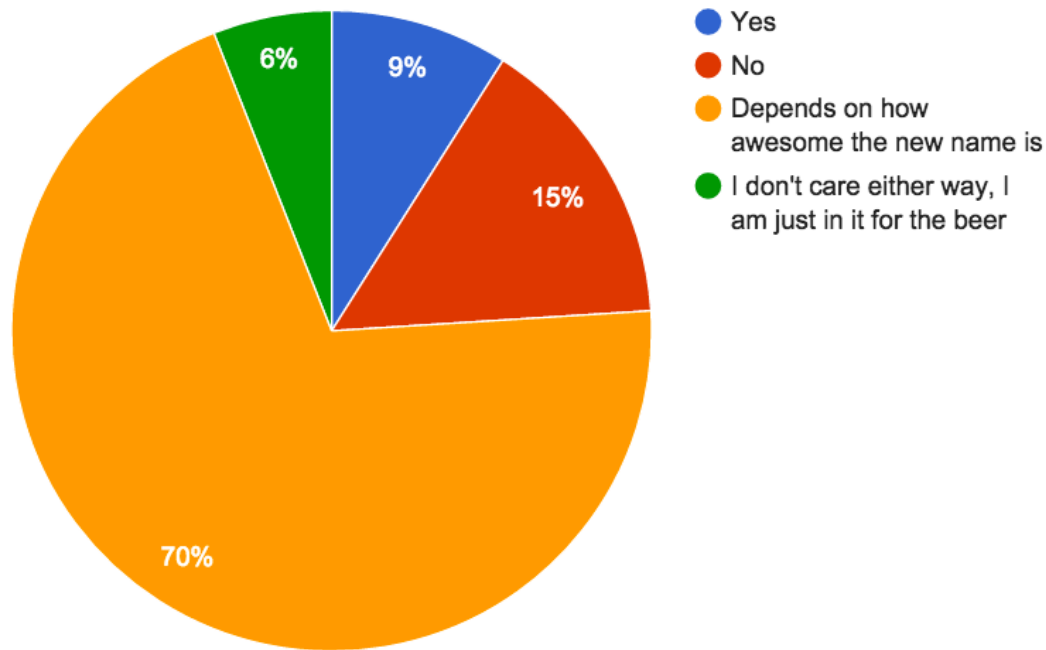
## Should BESSIG expand to include domain interests beyond Earth and Space Science?



Response	Count	Percent
Yes, BESSIG should expand to include anyone local interested in data issues, including all scientific domains as well as industry	16	47
Yes, BESSIG should expand to include anyone local interested in academic science data issues, including all scientific domains but not industry	7	20
No, BESSIG should remain an Earth Space Science-focused group	5	15
Other	6	17

This is a pretty solid vote in favor of expanding the scope of the group, although we may want to refine the nature of that expansion somewhat.

## Should the BESSIG name change?



Response	Count	Percent
Yes	3	9
No	5	15
Depends on how awesome the new name is	24	70
I don't care either way, I am just in it for the beer	2	6

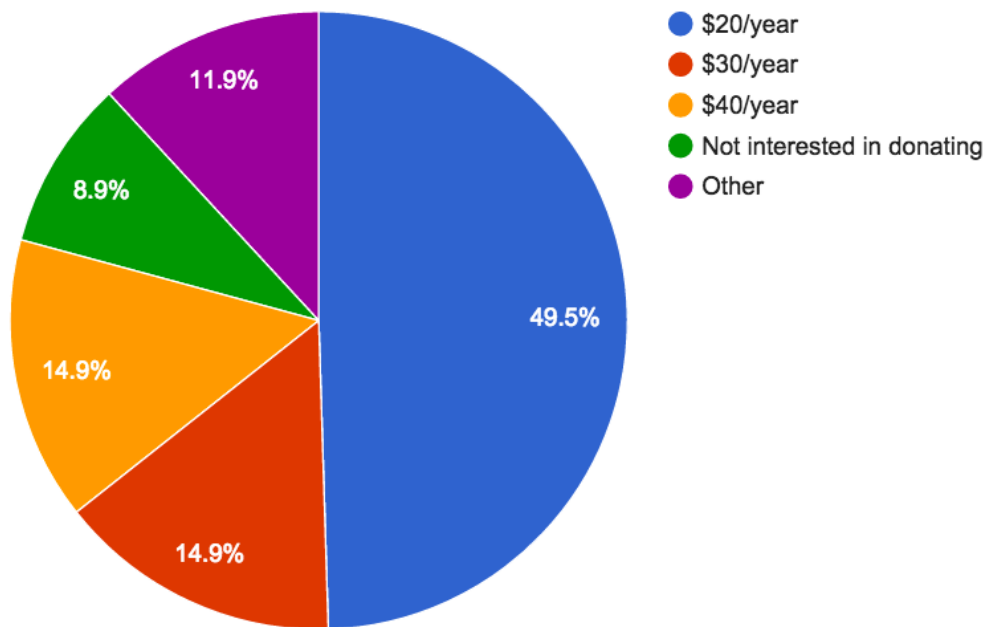
So we're not sure about the name change. Some of the suggestions:

- Boulder Planetary Data or Boulder Earth Data
- BDIG (Boulder Data on the Ground)
- BADTAP (Boulder Area Data in Theory and Practice)
- B DISH (Boulder Data Interoperability Starts Here; or flipped as DISH B)
- BADCOP (Boulder Area Data Community-of-Practice)
- Boulder Data Masters
- Boulder Data Group
- Boulder Data Whisperers
- FREDs (Front Range Earth Data Society)
- FREIDA (Front Range Informatics and Data Association)
- BADS Group (Boulder Area Data Science Group)
- BAFOD (Boulder Area Friends of Data)
- DIGBA (Data Interest Group, Boulder Area)
- BESDIG (Boulder Earth Science Data Interest Group)
- BAD ASS (Boulder Area Data Art and Sciences Society)

A word of caution here - there are a number of existing data science meetups in Boulder and we need to take into consideration the names of those groups before selecting a new name here. BESSIG is pretty unique; Boulder Data Science is not.

And another valid comment - it depends on the outcome of the scope question. If we expand beyond Earth and Space Sciences, shouldn't the name reflect that?

## Would you be willing make an annual donation to keep BESSIG running? If so, how much?



Response	Count	Percent
\$20/year	17	50
\$30/year	5	15
\$40/year	5	15
Not interested in donating	3	9
Other	4	12

Here, a majority of the respondents are willing to contribute some funds to support BESSIG. (This is voluntary and in no way reflects a desire to impose dues.)

Finally, from the set of additional comments space, we see a couple of trends. First, the meeting time is an issue for some folks. And second, there's a hesitation in expanding the scope outside of academics or the sciences, i.e. what do we lose if the scope is too broad?

We've learned a lot from just the three questions. The most important takeaway is that there's a lot of support to keep this community going and growing. And thank you, Anne, for keeping it going for the past five years!

(For anyone interested in the complete survey results, they're available as a [Google sheet](#).)

\* Since we are, at heart, a bunch of data geeks, the percents in the tables and in the charts vary due to Fun With Rounding across the two visualization options. The bins are big enough not to quibble, yes?

- [business](#)
- [survey](#)



[BESSIG Meeting Monday, August 31, 2015, 4:00 - 6:00, Gondolier](#)

Unknown User (wilson) posted on Aug 24, 2015

★ We're meeting at the Gondolier, <http://www.gondolierboulder.com/>, who are providing us this space for free. In searching for a venue, I have learned how rare and special it is for a business to provide this. Please come and order something so that the Gondolier can continue to offer us this space. ★

# Leveraging Internet Identity for Scientific Collaborations

**Ken Klingenstein, Evangelist, Digital Identity & Privacy, Internet2**

In the last several years there has been rapid development of an identity layer for the Internet. Efforts in government, R&E, businesses and among social identity providers are creating an infrastructure of identity and attributes that is being leveraged to access supercomputers, social sites, health care providers, federal research agencies, instrumentation and databases, cloud based storage and compute services, etc.

The two major areas in this work are federated identity, which allows local identities, authentication and attributes to be used Internet-wide, and collaboration platforms, which allow virtual organizations and other multi-institutional efforts to build on federated identity and seamlessly use a growing pool of collaboration applications (wikis, listservs, file sharing, code management tools, command line apps, etc).

This talk will discuss the current state of federated identity, including international inter-federation and US government activities, and how federated identities are being used in leading-edge US science communities. It will then present the emergence of collaboration platforms, and their ability to integrate access control and group management across collaboration applications using open standards. Demos might happen; interruptions and comments most welcome.

## Schedule (more or less)

4:00 - 5:00 Presentation

5:00 - 6:00 Social



BESSIG Meeting Thursday, July 9, 4:00 - 6:00, Gondolier

Unknown User (wilson) posted on Aug 24, 2015

## Identifiers & Relationships

### Joe Hourcle, Solar Data Analysis Center, Goodard

At a recent meeting, I came to realize that there are quite a few people who have more recently come into the field of data informatics, and have missed out on much of the discussions over the last decade on data identifiers. In the last few weeks two papers were published by some of the same co-authors that took a contrary position on the presentation of identifiers, although that was not a focus on either of the papers.

I will give an overview of some of the issues regarding identifiers for data (both those that I think are resolved and not), the need for vocabulary and standards to describe what is being identified, and the implications for data citation and describing other data relationships.

#### Bio:

Joe Hourcle is (as of the time this was written) a programmer/analyst for the Solar Data Analysis Center at Goddard Space Flight Center, working as a programmer / DBA / sysadmin / cataloger / whatever else on the Virtual Solar Observatory. He has an interest in classifying things and naming concepts -- he has been working with Todd King on a (still unpublished) vocabulary to discuss data systems (<http://virtualsolar.org/vocab>), and back before he knew anything about ontologies & controlled vocabularies, added the topics to [fark.com](http://fark.com). He would also like to remind you that the crew neck means that most t-shirts qualify as a 'shirt with a collar'.



BESSIG Meeting Monday, October 6, 4:30 - 6:00, Gondolier

Unknown User (wilson) posted on Oct 01, 2014

★ We'll be meeting at the Gondolier, <http://www.gondolierboulder.com>. The restaurant is in the Meadow's Shopping Center, on the southwest corner of Baseline and Foothills. We'll meet in their back room.

The Gondolier is providing us this space for free. In searching for a new venue, I have learned how rare and special it is for a business to provide this. Please come and order something so that the Gondolier can continue to offer us this space.

## What's an Ontology and What Should I Do With It?

### Beth Huffer, Lingua Logica

The word "ontology" is used to refer to a variety of different artifacts, from controlled vocabularies that serve as glossaries, to formal ontologies that serve as data schemas for graph databases and/or deductive reasoning systems. This talk will focus on use cases for formal ontologies, with a demonstration of and presentation on ODISEES (Ontology-Driven Interactive Search Environment for Earth Science) which was recently released in beta by the Atmospheric Science Data Center at NASA Langley Research Center. ODISEES provides a parameter-level search environment for discovering ASDC data resources, enabling users to specify a precise set of criteria and get a set of results that exactly match those criteria. Following an overview of the technology behind ODISEES, Beth will discuss additional use cases for formal ontologies of the sort driving ODISEES.

## Schedule (mostly)

4:30 - 5:30 Presentation

5:30 - 6:00 Social



BESSIG Meeting Wed, May 28, 3:30 - 5:00, Gondolier

Unknown User (wilson) posted on May 22, 2014

★ We have a new venue! We'll be meeting at the Gondolier, <http://www.gondolierboulder.com>. The restaurant is in the Meadow's Shopping Center, on the southwest corner of Baseline and Foothills. We'll meet in their back room.

The Gondolier is providing us this space for free. In searching for a new venue, I have learned how rare and special it is for a business to provide this. Please come and order something so that the Gondolier can continue to offer us this space.

## Who's Afraid of File Format Obsolescence?

### Evaluating File Format Endangerment Levels and Factors for the Creation of a File Format Endangerment Index

Heather Ryan, University of Denver Library and Information Science

Much digital preservation research has been built on the assumption that file format obsolescence poses a great risk to the continued access of digital content. In an endeavor to address this risk, a number of researchers created lists of factors that could be used to assess risks associated with digital file formats. My research examines these assumptions about file format obsolescence and file format evaluation factors with the aim of creating a simplified file format endangerment index.

This study examines file format risk under the new lens of 'file format endangerment,' or the possibility that information stored in a particular file format will not be interpretable or renderable in human accessible means within a certain timeframe. Using the Delphi method in two separate studies, this exploratory research collected expert opinion on file format endangerment levels of 50 test file formats; and collected expert opinion on relevance of 28 factors as causal indicators of file format endangerment.

Experts expressed the belief that generally, digital information encoded in the rated file formats will be accessible for 20 years or more. This indicates that file format experts believe that there is not a great deal of short-term risk associated with encoding information in the rated file formats, though this does not preclude continued engagement with preservation activities for these and other file formats. Furthermore, the findings show that only three of the dozens of file format evaluation factors discussed in the literature exceeded an emergent threshold level as causal indicators of file format endangerment: 'Rendering Software Available,' 'Specifications Available,' and 'Community/3rd Party Support.' Consequently, these factors are ideal candidates for use in a simple file format endangerment index that can be used to assess endangerment levels of any file format.

The findings of this study have implications for further exploration of file format endangerment in specific digital information creation domains. In particular, applying this model to file formats created by and used in the Earth and Space Science communities will both strengthen the model and will produce valuable insight into format-centric Earth and Space Science data creation and management practices. This insight can then be applied to risk assessment and subsequent actions to support continued access to datasets over time.

Come join us!



BESSIG Meeting Wed, April 16, 4:15 - 6:00, Outlook

Unknown User (wilson) posted on Apr 09, 2014

We're still at the Outlook through April 2014. We seek an alternative venue for May and beyond. Please see New Venue Desires below and keep them in mind as you move around Boulder.

★ Note that we'll start at 4:15 this month due to our speaker's schedule.

★ This month marks the 3rd anniversary of the BESSIG!

## An Easy Bake Semantic Metadata Repository for Scientific Data

Mik Cox, Tyler Traver, Anne Wilson, Doug Lindholm, Laboratory for Atmospheric and Space Physics (LASP), Don Elsborg, CU Faculty Affairs

This presentation will discuss the use of open source tools and the tasks that remained to create a semantically enabled metadata repository.

The LASP Interactive Solar Irradiance Data Center, LISIRD, is a web site that serves the lab's solar irradiance and related data products to the public. LISIRD provides information about the data it offers as part of its web page content, embedded in static HTML. At the same time, other LASP web sites also provide the same information, such as sites pertaining to specific missions or education and outreach. Keeping data set information updated and in sync across web sites is a problem. Nor is the information interoperable with emerging search and discovery tools.

To address this and other issues, we created a semantically enabled metadata repository that holds information about our data. In conjunction, we prototyped a new implementation of LISIRD that dynamically renders page content, pulling metadata from the repository and including in the page current, vetted metadata from a single, definitive source. Other web pages can similarly pull this information if they choose. Additionally we can now offer new semantic browse and search capabilities, such as search of data sets by type (currently spectral solar irradiance, total solar irradiance, and solar indices) or over a particular spectral range provided by the user.

We can also render the metadata in various formats understandable to other communities, such as SPASE for the heliophysics community and ISO for the international community. This will allow us to federate with sites that use those formats, allowing broader discovery of our data.

To date, metadata management at LASP has generally been done on a per project, ad hoc basis. We are building applications on top of the repository that provide CRUD (create, read, update, delete) capabilities for metadata records to metadata 'owners' and 'curators'. We expect this to help data managers to store and manage their metadata in a more rigorous fashion should they choose to use it.

We heavily leveraged existing open source tools to create the repository. In this talk we'll talk about using VIVO to create a semantic database, LaTiS to fetch data and metadata, and AngularJS to write dynamic, testable JavaScript. We'll describe our experiences extending two existing ontologies to meet our space physics domain needs.

With these tools and some student time (though our students are exceptional) we are achieving significantly increased capabilities at a relatively low cost. We believe this tool combination could help projects with limited resources achieve similar capabilities to manage and provide access to metadata.

And, if that's not easy-bake enough for you, try this PC EZ-Bake Oven, made especially for geeks: <http://www.thinkgeek.com/stuff/41/ezbake.shtml>. Link to the presentation: [http://prezi.com/89us-xb64leu/?utm\\_campaign=share&utm\\_medium=copy](http://prezi.com/89us-xb64leu/?utm_campaign=share&utm_medium=copy)

#### **Schedule (mostly)**

4:15 - 5:xx presentation

5:xx - 6:00 social

#### **New Venue Desiresments**

Free, or cost based on attendance

Can purchase food and beverages, or within walking distance of such

Easy to get to, easy to park, in Boulder

Separate room

Projection capability

Internet connectivity

hours 4:00 - 6:00 Tu or Wed, 2nd, 3rd, or 4th week of the month, flexible



BESSIG Meeting \*Tuesday\*, Mar 18, 4:00 - 6:00 PM, Outlook

Unknown User (wilson) posted on Mar 12, 2014

We're still at the Outlook through April 2014. We seek an alternative venue for May and beyond. Please see New Venue Desiresments below and keep them in mind as you move around Boulder.

★ Note that we're meeting on a Tuesday rather than a Wednesday this month due to room availability. We're back in the Chatauqua room at the Boulder Outlook Hotel.

## **Earth System CoG and the Earth System Grid Federation: A Partnership for Improved Data Management and Project Coordination**

**Sylvia Murphy, Cecelia DeLuca, Allyn Treshansky, NOAA/CIRES, Luca Cinquini, JPL/NOAA**

The Earth System CoG Collaboration Environment, led by a NOAA ESRL/CIRES team, is partnering with the DOE-led Earth System Grid Federation (ESGF) data archive to deliver a capability that will enable users to store, federate, and search scientific datasets, and manage and connect the projects that produced those datasets.

ESGF is an international network of data nodes that is used to host climate data sets, including the model outputs from the Coupled Model Intercomparison Project (CMIP), which supported the Intergovernmental Panel on Climate Change (IPCC) assessment reports. ESGF data nodes are federated, so that all data holdings are visible from any of the installation sites. An ESGF data node is now installed at NOAA's Earth System Research Laboratory (ESRL's). It currently hosts data from the Dynamical Core Model Intercomparison Project (DCMIP) and Twentieth Century Reanalysis data from ESRL's Physical Sciences Division.

CoG is a collaboration environment and connective hub for networks of projects in the Earth Sciences. It hosts software development projects, model intercomparison projects, and short university-level courses. It includes a configurable search to data on any ESGF node, metadata collection and display, project-level wikis, and a host of other capabilities. There are 74 projects currently using the system.

CoG is partnering with the international Earth System Model Documentation (ES-DOC) project, funded by both NOAA and the EU's Infrastructure for the European Network for Earth System Modeling (IS-ENES) project. ES-DOC is developing tools that capture, display, and compare Earth system model metadata. This information can be linked directly from a CoG project or attached to specific datasets in the ESGF node.

This presentation will provide an overview of both CoG and ESGF, demonstrate data discovery and download, and key CoG capabilities using relevant example projects.

CoG: <https://earthsystemcog.org/>

ESRL ESGF data node: <http://hydra.fsl.noaa.gov/esgf-web-fe/>

### Schedule (mostly)

4:00 - 5:xx presentation

5:xx - 6:00 social

### New Venue Desiresments

Free, or cost based on attendance

Can purchase food and beverages, or within walking distance of such

Easy to get to, easy to park, in Boulder

Separate room

Projection capability

Internet connectivity

hours 4 - 6:00 Tu or Wed, 2nd, 3rd, or 4th week of the month, flexible



BESSIG Meeting Wed, Feb 19, 4:00 - 6:00 PM, Outlook

Unknown User (wilson) posted on Feb 12, 2014

We're still at the Outlook through April 2014. We seek an alternative venue for May and beyond. Please see New Venue Desiresments below and keep them in mind as you move around Boulder.

★ Note that this meeting will be held in the Panorama Room of the Outlook Hotel instead of our usual Chataqua room. This means that we won't have a server and food and drinks must be ordered in the restaurant.

## Accessing Data Instead of Ordering Data: A New Normal

**Michael Little**, the Advanced Development Systems Engineer at the Atmospheric Science Data Center (ASDC)

Mike will describe how the new generation of research objectives will need to avoid staging data locally from multiple modeling and observational repositories. Rather, new access methods will present a machine-to-machine interface which permits codes and software applications to retrieve small increments of data continuously as part of the processing. The ASDC's Data Access architecture will be described with a particular emphasis on iRODS as one of the most promising tools for remote access to data held in earth science data centers.

Mike's slides for this talk are available here: [DataDistributionArchitecture\\_0.4.3.pptx](#).

### Schedule (mostly)

4:00 - 5:xx presentation

5:xx - 6:00 social

## New Venue Desirements

Free, or cost based on attendance

Can purchase food and beverages

Easy to get to, easy to park, in Boulder

Separate room

Projection capability

Internet connectivity

hours 4 - 6:00 Tu or Wed, 2nd, 3rd, or 4th week of the month, flexible



[BESSIG Meeting Wed, Jan 22, 4:15 - 6 PM, Outlook](#)

Unknown User (wilson) posted on Jan 17, 2014

[We're still at the Outlook through April 2014.](#) We seek an alternative venue for May and beyond. Please see New Venue Desirements below and keep them in mind as you move around Boulder.

## Deep Carbon Observatory - Data Science and Data Management Infrastructure Overview and Demonstration

**Patrick West**, Rensselaer Polytechnic Institute

The Deep Carbon Observatory (DCO) brings together hundreds of organizations and individuals from all around the world, spanning a great many scientific domains with a focus on Carbon. The DCO Data Science team is anticipating the generation of terabytes of information in the form of documents, scientific datasets from level 0 to data products and visualizations, information about events, people, and organizations, and more. So how do we keep track of all of this information, manage the information, and disseminate the information?

In order to organize all of this information and provide the research community the tools necessary to collaborate and do their research, the DCO Data Science team is putting together a suite of tools that will integrate all of these components in a seamless, distributed, heterogeneous environment. This presentation and demonstration will provide an overview of the work that we, the DCO Data Science team, are doing to provide such an environment.

Due to Patrick's schedule, we'll plan on starting at 4:15 instead of 4:00.

Here are Patrick's slides: <http://tw.rpi.edu/web/doc/DCO-DS-Overview-Demonstration-BESSIG>.

## Schedule (mostly)

4:15 - 5:xx presentation

5:xx - 6:00 social

## New Venue Desirements

Free, or cost based on attendance

Can purchase food and beverages

Easy to get to, easy to park, in Boulder

Separate room

Projection capability

Internet connectivity

hours 4 - 6:00 Tu or Wed, 2nd, 3rd, or 4th week of the month, flexible



[BESSIG Meeting Wed, Nov 20, 4 - 6 PM](#)

Unknown User (wilson) posted on Oct 16, 2013

Our meeting this month is a special event for several reasons. Copies of Andrew's book will be available to the first 50 attendees, and the HDF Group will be providing refreshments for us. Also, this may be our last meeting at the [Boulder Outlook Hotel](#), as the hotel has been sold. So, please join us **in the Crown Rock room (not our usual room)** at the Outlook for:

## Improving Science with Open Formats and High-Level Languages: Python and HDF5

**Andrew Collette**, Laboratory for Atmospheric and Space Physics (LASP)



This talk explores how researchers can use the scalable, self-describing HDF5 data format together with the Python programming language to improve the analysis pipeline, easily archive and share large datasets, and improve confidence in scientific results. The discussion will focus on real-world applications of HDF5 in experimental physics at two multimillion-dollar research facilities: the Large Plasma Device at UCLA, and the NASA-funded hypervelocity dust accelerator at CU Boulder. This event coincides with the launch of a new O'Reilly book, *Python and HDF5: Unlocking Scientific Data*, complimentary copies of which will be available for attendees.

As scientific datasets grow from gigabytes to terabytes and beyond, the use of standard formats for data storage and communication becomes critical. HDF5, the most recent version of the Hierarchical Data Format originally developed at the National Center for Supercomputing Applications (NCSA), has rapidly emerged as the mechanism of choice for storing and sharing large datasets. At the same time, many researchers who routinely deal with large numerical datasets have been drawn to the Python by its ease of use and rapid development capabilities.

Over the past several years, Python has emerged as a credible alternative to scientific analysis environments like IDL or MATLAB. In addition to stable core packages for handling numerical arrays, analysis, and plotting, the Python ecosystem provides a huge selection of more specialized software, reducing the amount of work necessary to write scientific code while also increasing the quality of results. Python's excellent support for standard data formats allows scientists to interact seamlessly with colleagues using other platforms.

#### Schedule (more or less)

4:00 - 5:00 presentation  
5:00 - 6:00 social



BESSIG Meeting Wed, Oct 23, 4 - 6 PM

Unknown User (wilson) posted on Oct 04, 2013

Regridding of data is a common problem faced by many scientific software developers. If regridding is part of your world, this talk may be of interest to you. Come join us at the [Boulder Outlook Hotel](#) for this month's talk:

## There is more to conservative interpolation--- interpolating edge and face centered fields in the geo-sciences

Alexander Pletzer, Tech-X

Interpolation is one of the most widely used postprocessing tasks, according to a survey of Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT) users. Most geo-postprocessing tools (UV-CDAT, NCL, Ferret, etc) support a choice of both bilinear and conservative regridding with conservative interpolation guaranteeing that the total amount of "stuff" (energy, water, etc) remains unchanged after regridding. The SCRIP and ESMF are examples of libraries implementing these interpolation methods.

We argue that the type of interpolation is dictated by the type of field and that cell centered fields require conservative interpolation whereas nodal fields require bilinear (or higher order) interpolation. Moreover, the wind velocity fields used by finite-volume atmospheric codes, which are neither cell-centered nor nodal but face-centered (Arakawa D staggering), require different interpolation formulas. Interpolation formulas of face-centered and edge-centered (Arakawa C) fields have been known as Whittney forms since 1957 and are widely used in electromagnetics. We present interpolation methods new to the geo-sciences that conserve flux and line integrals for Arakawa D, respectively Arakawa C, staggered fields.

This talk should be of interest to anybody in need to regrid velocity and other vector fields whose components are staggered with respect to each other.

#### Schedule (mostly)

4:00 - 5:00 Presentation  
5:00 - 6:00 Social



BESSIG Meeting Wed, Sep 18, 4 - 6 PM

Unknown User (wilson) posted on Sep 12, 2013

*"Code without tests is bad code. It doesn't matter how well written it is; it doesn't matter how pretty or object-oriented or well-encapsulated it is. With tests, we can change the behavior of our code quickly and verifiably. Without them, we really don't know if our code is getting better or worse." [FEATHERS]*

A strong statement, but it does bring home the vital role of testing in software development. Join us at the [Boulder Outlook Hotel](#) for:

## Strategies, motivations, and influencing adoption of testing for scientific code

Ian Truslove, Erik Jasiak, NSIDC

Computation and programming are increasingly inescapable in modern Earth Sciences, but scientists and researchers receive little or no formal software engineering or programming training. At the same time, research into the reproducibility of other academic papers exposing disappointingly low rates of repeatability and high-profile retractions due to computational or data errors increase the onus on researchers to write repeatable, reliable, even reusable programs; in other words, "write better code".

Software engineering has plenty to say on the matter of "better code": metrics, methodologies, processes, tools... Of course, none are indisputable and none provide absolute guarantees. One seemingly obvious technique - testing - has enjoyed a renaissance in incarnations such as unit testing, and with approaches such as test-driven development (TDD) and behavior-driven development (BDD).

Based on our experience at the National Snow and Ice Data Center (NSIDC) with unit testing, TDD and BDD, we present a set of recommendations to scientific and research programmers about some techniques to try in their day to day programming, and possibly provide some inspiration to aim for more comprehensive approaches such as BDD. We will highlight some use cases of various types of testing at the NSIDC, discuss some of the cultural and management changes that occurred for programmers, scientists and project managers to consider and adopt processes such as TDD, make recommendations about how to introduce or expand rigorous code testing practices in your organization, and discuss the likely benefits in doing so.

**[Scroll down to see post presentation references and material.]**

## Schedule

4:00 - 5:00 Presentation

5:00 - 6:00 Social

All are welcome.

## Post Presentation References and Material

### Presentation slides

Wilson et al, "[Best Practices for Scientific Computing](#)", highly recommended!

Merali, "[Computational science: ... Error](#)" in Nature. "As a general rule, researchers do not test or document their programs rigorously, and they rarely release their codes, making it almost impossible to reproduce and verify published results generated by scientific software, say computer scientists."

Good books for unit testing, TDD, and higher-level tests

Freeman and Pryce, [Growing Object-Oriented Software, Guided by Tests](#)

Beck, [Test Driven Development: By Example](#)

Fowler, [Mocks Aren't Stubs](#) - Martin Fowler on the terminology and usage of mocks, stubs, test doubles - all those "fake collaborators"

A couple of Bob Martin's books are particularly noteworthy for covering lots and lots of desirable attributes for code

Martin, [Agile Software Development, Principles, Patterns, and Practices](#)

Martin, [Clean Code: A Handbook of Agile Software Craftsmanship](#)

Some test frameworks

[JUnit](#) - the original (Java)

[RSpec](#) (Ruby)

[Behave](#) (Python)

[Jasmine](#) (JavaScript)

[mgunit](#) (IDL)

[pfUnit](#) (FORTRAN)

[Cucumber](#) (Acceptance tests, lots of languages)

Also

Feathers, Michael C., [Working Effectively with Legacy Code](#), Prentice Hall, 2005, p. xvi.

Snowden, Cynefin: [Wikipedia on Cynefin](#), [David Snowden introducing Cynefin \(video\)](#) - applicable to knowledge management, cultural change, and community dynamics, and has also involved issues of organizational strategy.

Snowden, Boone, 2007 "[A Leader's Framework for Decision Making](#)" (must pay for access from Harvard Business Review, though perhaps available elsewhere)



[BESSIG Meeting Wed, Aug 21, 4 - 6 PM](#)

Unknown User (wilson) posted on Jul 22, 2013

NSIDC's and ultimately Boulder's loss of Mark Parsons is RDA's gain. But maybe that's better for the world as a whole. Join us at the [Boulder Outlook Hotel](#) on August 21 to hear about the Alliance Mark has joined.

## The Research Data Alliance: Creating the culture and technology for an international data infrastructure

**Mark Parsons**, Managing Director, Research Data Alliance/U.S.

All of society's grand challenges -- be it addressing rapid climate change, curing cancer and other disease, providing food and water for more than seven billion people, understanding the origins of the universe or the mind -- *all* of them require diverse and sometimes very large data to be shared and integrated across cultures, scales, and technologies. This requires a new form and new conception of infrastructure. The Research Data Alliance (RDA) is creating and *implementing* this new data infrastructure. It is building the connections that make data work across social and technical barriers.

RDA launched in March 2013 as a international alliance of researchers, data scientists, and organizations to build these connections and infrastructure to accelerate data-driven innovation. RDA facilitates research data sharing, use, re-use, discoverability, and standards harmonization through the development and adoption of technologies, policy, practice, standards, and other deliverables. We do this through focussed Working Groups, exploratory Interest Groups, and a broad, committed membership of individuals and organizations dedicated to improving data exchange.

What data sharing problem are you trying to solve? Find out how RDA can help.

[Schedule](#)

4:00 - 5:00 Presentation

5:00 - 6:00 Social

Please join us!



[BESSIG Meeting Wed, July 24, 4 - 6 PM](#)

Unknown User (wilson) posted on Jul 15, 2013

In July we're meeting in the 4th week of the month, rather than the 3rd. Please join us at the [Boulder Outlook Hotel](#) for our own Ted talk:

## HDF and The Earth Science Platform

**Ted Habermann**, The HDF Group

Interoperable data and understanding across the Earth Science community requires convergence towards a standard set of data formats and services, metadata standards, and conventions for effective use of both. Although large legacy archives still exist in netCDF3, HDF4, and many custom formats, we have achieved considerable convergence in the data format layer with the merger of the netCDF4 and HDF5 formats. The way forward seems clear as more groups in many disciplines join the HDF5 community. The data service layer has experienced similar convergence as OGC Service Standards are adopted and used in increasing numbers and connections across former chasms are deployed (ncWMS, ncSOS, netCDF/CF as OGC Standards). Many data providers around the world are in the process of converging towards ISO Standards for documenting data and services. Connections are also helping here (ncISO). Many groups are now working towards convergence in the conventions layer. The HDF-EOS and Climate-Forecast conventions have been used successfully for many datasets spanning many Earth Science disciplines. These two sets of conventions reflect different histories and approaches that provide a rich set of lessons learned as we move forward.

[Schedule](#)

4:00 - 5:00 Presentation

5:00 - 6:00 Social

Stop on by!



[BESSIG Meeting Wed, June 19, 4 - 6 PM](#)

Unknown User (wilson) posted on Jun 13, 2013

This month we'll meet again the [Boulder Outlook Hotel](#) for:

## Py in the Sky: IPython and other tools for scientific computing

**Monte Lunacek**, Application Specialist, CU Research Computing

**Roland Viger**, Research Geographer, USGS

Python offers a rich toolkit that is useful for scientific computing. In this talk, we will introduce the IPython package and discuss three useful components: the interactive shell, the web-based notebook, and the parallel interface. We will also demonstrate a few concepts from the Pandas data analysis package and, time permitting, offer a few tips on how to profile and effortlessly speedup your python code. This talk will describe and illustrate these tools with example code. If Python is not your favorite programming language, this overview might change that.

[Schedule](#)

4:00 - 5:00 Presentation

5:00 - 6:00 Social

Come on by!



BESSIG Meeting Tue, May 21, 4 - 6 PM

Unknown User (wilson) posted on Apr 11, 2013

Note that this month we are meeting on a Tuesday instead of a Wednesday!

Please join us at the [Boulder Outlook Hotel](#) for a presentation and demo of:

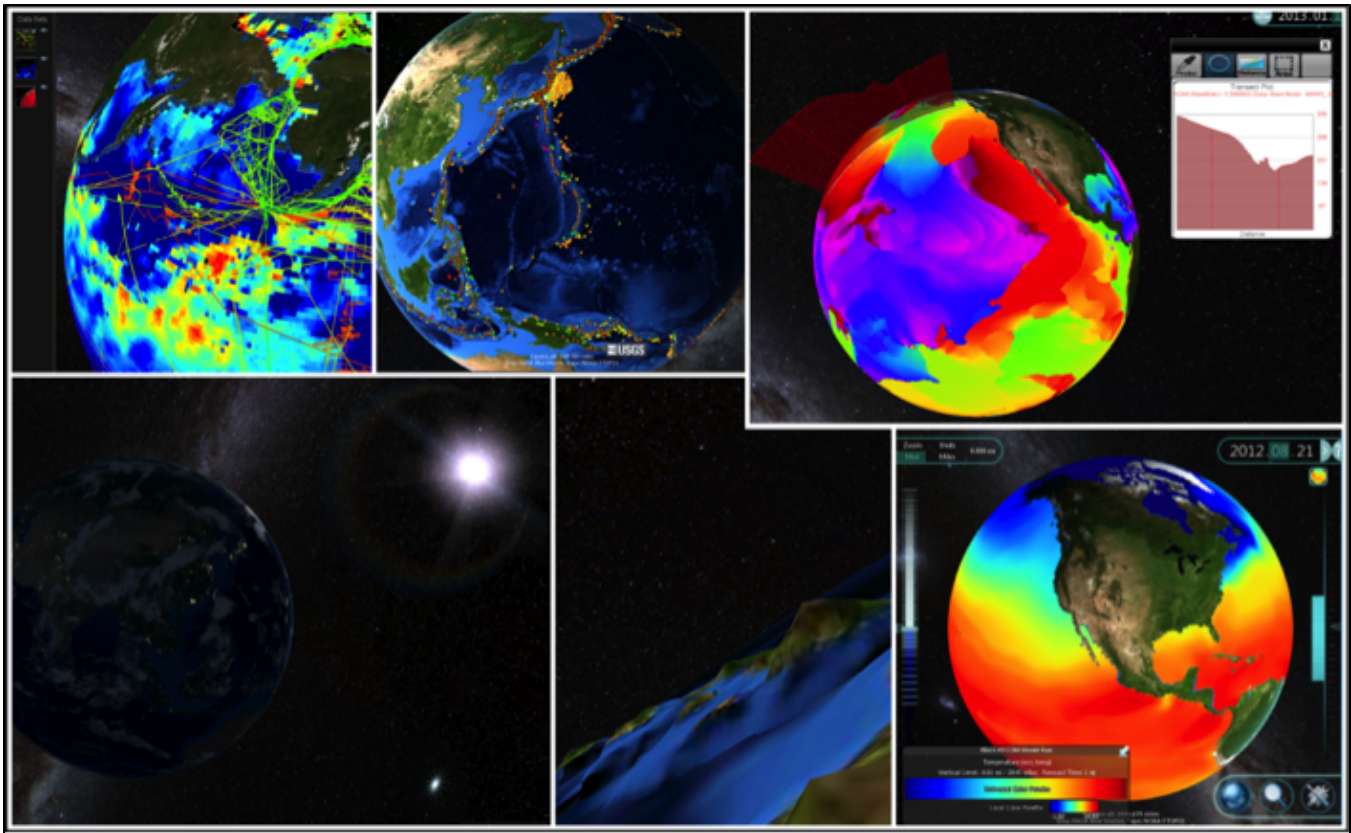
*NOAA Earth Information Services and TerraViz*

**Eric Hackathorn, Julien Lynge, Jeff Smith, TerraViz, NOAA**

**Jebb Stewart, Chris MacDermid, NEIS, NOAA**

The NOAA Earth Information Services (NEIS) is a framework of layered services designed to help the discovery, access, understanding, and visualization of data from the past, present, and future. It includes a visualization component named TerraViz that is a multi-platform tool, running on desktops, web browsers, and mobile devices. The goal is to ingest "big data" and convert that information into efficient formats for real-time visualization. Designed for a world where everything is in motion, NEIS and TerraViz allow fluid data integration and interaction across 4D time and space, providing a tool for everything NOAA does and the people NOAA affects.

TerraViz is built using the Unity game engine. While a game engine may seem a strange choice for data visualizations, our philosophy is to take advantage of existing technology whenever possible. Video games are a multibillion-dollar industry, and are quite simply the most powerful tools for pushing millions of points of data to the user in real-time. Our presentation illustrated displaying environmental data in TerraViz at a global scale, visualizing regional data in "scenes" such as the flooding of the Washington DC area or rotating a coastal ecosystem in three axes, and developing environmental simulations/games like exploring the ocean floor in a submarine.



The NEIS backend similarly takes lessons from private industry, using Apache Solr and other open source technologies to allow faceted search of NOAA data, much as sites like Amazon and Netflix do.





We believe that to have an impact on society, data should be easy to find, access, visualize, and understand. NEIS simplifies and abstracts searching, connectivity, and different data formats, allowing users to concentrate on the data and science.

Please contact us if you want to explore including your environmental data within NEIS/TerraViz or if you want to talk to us about developing custom visualizations or educational simulations to showcase your important data.

NOAA / Earth System Research Lab / Global Systems Division, Boulder, Colorado

NEIS/TerraViz: [NEIS/TerraViz](#)

### Schedule

4:00 - 5:00 Presentation

5:00 - 6:00 Social



[BESSIG Meeting Wed, Apr 17, 4 - 6 PM](#)

Unknown User (wilson) posted on Apr 05, 2013

This month marks the two year anniversary of the BESSIG! Please join us at the [Boulder Outlook Hotel](#) for a remote presentation:

**Chris Lynnes**, Chief Systems Engineer of the Goddard DAAC, NASA, "*The Earth Science Collaboratory*"

The Earth Science Collaboratory is a proposed framework for supporting the sharing within the Earth science community of data, tools, analysis methods, and results, plus all the contextual knowledge that go with these artifacts. The likely benefits include:

- access to expert knowledge about how to work with data safely and efficiently
- full reproducibility of results
- efficient collaboration within multi-disciplinary and/or geographically distributed teams
- a social network to bring together researchers and data users with common interests

Currently, there are some nascent efforts to construct such a collaboratory. However, by its very (inclusive) nature, this construction is likely to be most successful as an emergent process, evolving from many point-to-point connections to an eventual ecosystem of cooperating components supporting collaboration.

In particular, **we are actively seeking scientists and other potential users of such a collaboratory to provide an end user perspective of system functionality.** Would you find such a collaboratory helpful? Do you have ideas about how it could be better? Would you like to influence its design? Those that are actively engaged will be heard and could end up with a tool that particularly suits their needs. If this role interests you, please attend this talk and/or otherwise let us know of your interest.

### Schedule

4:00 - 5:00 Presentation

5:00 - 6:00 Social

Drop on by!

### Post presentation material

The slides for the talk are available here: [ESC BESSIG slides](#).

The recorded version of the talk is available [here](#). Please note that the talk actually starts 21 minutes into the recording, as the first 15 minutes were intended to be for testing. (Sorry, we had serious technical difficulties at the hotel! It will be better next time!)



[BESSIG Meeting Wed, March 20, 4 - 6 PM](#)

Unknown User (wilson) posted on Mar 07, 2013  
Please join us at the [Boulder Outlook Hotel](#) for:

**Doug Lindholm**, LASP, "*LaTiS: a data model, an API, a web service AND a floor wax*"

LaTiS is a data model, a data analysis API, and a REST-ful web service for accessing scientific data via a common interface.

The LaTiS data model provides a scientific domain independent, unifying, mathematical foundation for describing datasets that captures the functional relationships between parameters. The Scala implementation of this model provides an API for reading data directly from their native source, the ability to compute with high level abstractions appropriate for the task at hand, and options for filtering, transforming, and writing data in various formats.

This talk will discuss how these capabilities are used to enable a modular web service framework that can easily be installed and configured by a data provider, and that allows users to dynamically reformat a dataset, including its time representation, storage format, missing values, etc.

This talk will be a preview (i.e. beta release) of the talk I will give at UCAR Software Engineering Assembly Conference in April.

### Schedule

4:00 - 5:00 Presentation

5:00 - 6:00 Social

Come on by!

*\* [Open on suburban kitchen, Wife and Husband arguing]*

*Wife: New LaTiS is a floor wax!*

*Husband: No, new LaTiS is a data model!*

*Wife: It's a floor wax!*

*Husband: It's a data model!*

*Wife: It's a floor wax, I'm telling you!*

*Husband: It's a data model, you cow!*

*Spokesman: [enters quickly] Hey, hey, hey, calm down, you two. New LaTiS is both a floor wax and a data model! Here, I'll spray some on your mop. [sprays LaTiS onto mop] ..and some for your data server. [sprinkles LaTiS onto laptop]*

*[Husband computes while Wife mops]*

*Husband: Mmmmm, works great!*

*Wife: And just look at that shine! \*\**

\*\* with apologies to SNL



[BESSIG Meeting Wed, Feb 13, 4 - 6 PM](#)

Unknown User (wilson) posted on Jan 29, 2013

**Due to the constraints of our speaker, we're meeting the 2nd week of February instead of the 3rd.**

Yet more around semantics! Please join us at the [Boulder Outlook Hotel](#) for:

**Beth Huffer**, Lingua Logica, "*ODISEES: An Ontology-Driven Interactive Search Environment for Earth Sciences*"

As part of an on-going effort at NASA Langley's Atmospheric Science Data Center, and in cooperation with the Computational & Information Sciences & Technology Office at the Goddard Space Flight Center, we have developed a semi-automated method for finding and comparing equivalent data and climate model output variables across disparate datasets. We will demonstrate an ontology-driven variable matching service that provides an automated mapping among comparable variables from multiple data products and climate model output products. The interactive user interface is driven by a queriable ontological model of the essential characteristics of data and climate model output variables, the products they occur in, the atmospheric parameters represented in the data, and the instruments and techniques used to measure or model the parameters. Queries of the ontology and triple store are used to match comparable variables by enabling users to search for those that share a user-specified set of essential characteristics.

The application addresses an emerging need among Earth scientists to compare climate model outputs to other models and to satellite observations, and addresses some of the barriers that currently make such comparisons difficult. In particular, the application

- Eliminates the need for users to be familiar with the multiple data vocabularies and standards that exist within the Earth sciences community; and

- With a few mouse clicks, provides ready access to the information needed by scientists to understand the similarities and differences between two or more data or climate model products, enabling them to quickly determine which products best suit their requirements.

#### Schedule

4:00 - 5:00 Presentation

5:00 - 6:00 Social

Come on by!



[BESSIG Meeting Wed, Jan 16, 4 - 6 PM](#)

Unknown User (wilson) posted on Jan 08, 2013

More on semantics! Please join us at the [Boulder Outlook Hotel](#) for:

**Stephen Williams**, Office of Faculty Affairs, CU Boulder, "*VIVO, VITRO, DataStar, and Beyond - The VIVO Project*"

The VIVO project was started at Cornell University in 2003 as a faculty profiling system for Mann Library. The profiling system that is VIVO was designed in two parts, VITRO the semantic engine that is ontology agnostic and VIVO the ontology specific pages and data for presenting faculty profiles. This concept of a two tier system was taken into the third tier with location specific changes (Cornell and CU-Boulder) and ontologies that build upon VIVO (data star). This talk will focus on the VIVO project as a whole, its history, its ancillary projects, and its future. We'll also try to cover difficulties and lessons in semantic programming and the experiences of building ETL tools for semantic data.

#### Schedule

4:00 - 5:00 Presentation

5:00 - 6:00 Social



[BESSIG Meeting Tue, Nov 27, 4 - 6 PM](#)

Unknown User (wilson) posted on Nov 09, 2012

**Note: this month we're meeting on a Tuesday!**

Please join us back at the [Boulder Outlook Hotel](#) for:

**Nate Wilhelmi**, NCAR/CISL, "[Experiences Using RDF and a Triple Store for Metadata Storage and Search](#)"

4:00 - 5:00 presentation

5:00 - 6:00 social

See you there!



[BESSIG Meeting Wed, Oct 10, 4 - 6, NEON HQ](#)

Unknown User (wilson) posted on Sep 12, 2012

### "Informing Science Policy: the Role for Scientists and Engineers"

#### A Panel Discussion

**Note: This meeting will be held at NEON Headquarters, instead of our usual location. Details below.**

This month Brian Wee, Steve Aulenbach, and I are delighted to have representatives from law, government and science come together to discuss various aspects of science policy. We've asked them to consider questions like these:

- What does "science policy" mean to you? To your organization? What impact does it have?
- What are the roles in science policy and what impacts do they have? Who are the main players?
- How have you or your organization tried to impact science policy? What worked and what did not work? What did you learn?
- How does one prepare for a science policy discussion? Any do's and don'ts?
- Scientists and engineers are trained to think and communicate in certain ways. Should those same skills be applied to policy discussions?
- If someone wanted to move more heavily into science policy, how would you advise them? What career moves would be good? Any bad career moves?

#### Panelists

##### Peter Backlund

Director, NCAR External Relations and the Integrated Science Program  
Director, Research Relations, NCAR

## Dan Baker

Professor of Astrophysical and Planetary Sciences  
Director, Laboratory for Atmospheric and Space Physics

## Alice Madden

Wirth Chair in Sustainable Development, UC Denver  
Colorado House Representative (2001 - 2010), Majority Leader (2004 - 2008)  
Climate Change Adviser, Deputy Chief of Staff for Gov. Ritter  
Senior Fellow on Climate Change, Center for American Progress

## Andy Schultheiss

District Director at Office of Congressman Jared Polis  
Campaigns Director at League of Conservation Voters  
Boulder City Council (2003 - 2007)

## Logistics

This meeting is being hosted by NEON, Inc. It will be held at NEON Headquarters, 1685 38th Street #100, Boulder, CO.

4:00 - 5:15 Panel Discussion

5:15 - 6:00 Social

Beverages will be provided courtesy of NEON. **Thank you Brian and NEON!**

Very light snacks to be covered by donations.

Please join us for this interesting discussion!

The discussion will be available via Web Ex, info to follow.



**BESSIG Meeting Wed, Sept 19, 4 - 6 PM**

Unknown User (wilson) posted on Sep 12, 2012

Another BLAST! Many thanks to Anna and Dave for their willingness to share ideas with us this month.

**Anna Milan**, NOAA/NESDIS/NGDC, *Metadata for the Archive: Transition to ISO, Approaches, Challenges, and Opportunities*

**Dave Fulker**, President, OPeNDAP, Inc., *A (Very) Rough Idea: Raster Binning and Masking Services* (see attachment)

I'll sketch my idea for a new type of data query/response service built (perhaps for EarthCube) around a standardized space-time raster that has a dual function. Tentatively dubbed "Raster Binning & Masking Services" or RBinMasks, users would gain a (potentially standard) way to specify (irregular) space-time regions of interest and a (potentially standard) way to gain information about the space-time distributions of pertinent data, without -- or before -- retrieving actual values.

Come join us at the [Boulder Outlook Hotel!](#)

4:00 - 5:00 Presentations

5:00 - 6:00 Social

[2 Comments](#)



**BESSIG Meeting Wed, Aug 15, 4 - 6 PM**

Unknown User (wilson) posted on Aug 03, 2012

**Note: Meeting on Wednesday this month!**

**Brian Wee**, NEON, Inc., *"NEON: A continental-scale research and operations platform for the environmental sciences"*

As NEON, Inc.'s Chief of External Affairs, Brian is the organization's liaison to Congress, US Federal agencies, and other scientific organizations. He also represents the informatics needs of the large-scale environmental sciences before the computer science and Federal data community. Brian joined the NEON Project Office at the American Institute of Biological Sciences in 2004 as a post-doctoral associate, then became a staff scientist before transitioning to the role of Administrative Director. Previously he worked for Andersen Consulting (now Accenture) designing and implementing IT solutions and then served as Senior Instructional Designer leading instructional design, knowledge management, business-process redesign, and web development projects.

Brian holds a Ph.D. in Ecology, Evolution, and Behavior from the University of Texas at Austin, a M.Sc. degree in Computer Science – Artificial Intelligence at Northwestern University, Evanston, IL and a B.Sc. in Information Systems and Computer Science from the National University of Singapore. His M.Sc. studies focused on designing and implementing computer augmented learning solutions for high-school classrooms and corporate training at the Institute for the Learning Sciences. His Ph.D. focused on investigating the relative effects of behavioral, physiological and landscape barriers on the genetic structure of insect populations by integrating genetic, behavioral, and GIS analyses.

Come join us at the [Boulder Outlook Hotel!](#)



4:00 - 5:00 Presentation

5:00 - 6:00 Social



[BESSIG Meeting Tue, July 24, 4 - 6 PM](#)

Unknown User (wilson) posted on Jul 12, 2012

**NB:** This is a slightly different scheduling than we've had in past! Tuesday instead of Wednesday, 4th week of the month instead of the third.

**Jeffrey Morisette**, USGS, *"Developing a common modeling framework for the Department of Interior's North Central Climate Science Center"*

This month we welcome Jeff Morisette, visiting us from USGS in Fort Collins to talk about, among other things, his experience with VisTrails. From the page [VisTrails Overview](#), "VisTrails was designed to manage rapidly evolving scientific workflows and provenance that support simulations, data exploration and visualization." Jeff's group added a package to that software: SAHM, Software for Assisted Habitat Modeling.

Jeff is currently the director of the DOI North Central Climate Science Center where he manages and conducts research on how natural and cultural land management can respect the non-stationary nature of climate. A current research theme is how dynamic species distribution models can contribute to vulnerability assessment and adaptation planning.

We'll meet at the [Boulder Outlook Hotel](#). Come on by!

#### Schedule

4:00 - 5:00 Presentation

5:00 - 6:00 Social



[BESSIG Meeting Wed, June 20, 4 - 6 PM](#)

Unknown User (wilson) posted on Jun 13, 2012

More around semantics!

#### Who

SiriJodha Khalsa, NSIDC, *"Modeling the Model - the Semantics of the CCSM4 Sea Ice Model "*

Don Elsborg, LASP, *"Applied Semantic Web Technology - A use case with Semantic Mediawiki"*

#### Where

[Boulder Outlook Hotel](#)

#### Schedule

4:00 - 6:00 PM

Please come!



[BESSIG Meeting Wed, May 16, 4 - 6 PM](#)

Unknown User (wilson) posted on May 10, 2012

This month we continue exploring ontology and semantic-related areas.

#### Who

Stephan Zednick, Rensselaer Polytechnic Institute (RPI) *"Data Models and Ontologies, describing structure and classification"*

#### When

Wednesday, May 16, 4:00 - 6:00 PM

#### Where

[Boulder Outlook Hotel](#)

#### Schedule

4:00 - 5:00 Presentation and discussion

5:00 - 6:00 social

Please join us!



[BESSIG Meeting Wed, April 18, 5PM - 7PM](#)

Unknown User (wilson) posted on Apr 15, 2012

**Note the time change: 5:00 instead of 4:00!**

This month we'll review the recent UCAR data citation workshop, then make a foray into ontology and semantic-related areas.

In May and June we'll continue with speakers on ontology and semantic-related topics. The site [calendar](#) contains more information. If you have experiences in this area that you are willing to share, please contact Anne.

### Speakers

Matt Mayernik, NCAR Library, "[UCAR Workshop Review - Bridging Data Lifecycles: Tracking Data Use via Data Citations](#)." (Note: Many of these slides were taken from the workshop presentations posted at [http://library.ucar.edu/data\\_workshop/](http://library.ucar.edu/data_workshop/). Original slide authors are noted in red text in the top left of the slides.)

Ruth Duerr, NSIDC, "[Early Experiences in Semantics](#) "

**When:** Wednesday, April 18, 5:00 - 7:00

**Where:** [Boulder Outlook Hotel](#)

### Schedule

5:00 - 6:00: speakers

6:00 - 7:00: social



[BESSIG Social Hour, Tuesday, April 3, 5,30PM, Outlook](#)

Unknown User (wilson) posted on Mar 29, 2012

Organized by Mike Daniels and Matt Mayernik, UCAR's EOL/CDS is hosting a BESSIG social hour in the Atrium at the [Boulder Outlook Hotel](#), on Tuesday, April 3, 2012, at 5:30. The occasion is the visit of our out of town friends and ESIP colleagues Carol Meyer and Erin Robinson. EOL/CDS will provide hors d'oeuvres and non alcoholic drinks. Alcoholic beverages will be available for purchase.

In particular, this is an opportunity to discuss ESIP activities. Of course, any topic is fair game. Please join us!

Thank you, UCAR, EOL/CDS, Mike, and Matt!!



[Research Data Support Person - Job Posting](#)

Unknown User (kstacey) posted on Mar 19, 2012

### RESEARCH DATA SUPPORT

#### POSITION DESCRIPTION

**University of Colorado at Boulder Research Computing**

**Research Data Support (Part-Time/Temporary 15 hours per week for 6 months)**

**Research Computing**

**POSITION SUMMARY:** The Research Data Support position is part-time, reporting to the Research Data Manager. Primary job duties include website content creation for the new collaborative portal for Research Data Services. Research Data Services is a collaborative partnership between many campus groups though initially the primary groups are Libraries and Research Computing. The position will work with the Research Data Manager and Metadata Librarian to write and organize content to be used in the Drupal web site. The Research Data Services group and website will act as a nexus for coordinating activities to support researchers with a multitude of research data needs.

#### JOB RESPONSIBILITIES

1. Research and write up information about various topics for use as content for the Research Data Services website.
1. Assist with design and organization of the Research Data Services website following standards and best practices for information architecture.
1. Perform other related duties as assigned.

#### QUALIFICATIONS

1. Effective research and communication skills.

1. Knowledge of data management, familiarity with various stages of the data lifecycle (e.g., collection, access, preservation), and awareness of ongoing national/international discussions and literature on these topics.
1. Strong technical skills and ability to learn to use new software quickly.
1. Drupal development knowledge/experience desired but not required.

#### To Apply or Ask Questions:

Send an e-mail to Kimberly Stacey, Research Data Manager with Research Computing at [kimberly.stacey@colorado.edu](mailto:kimberly.stacey@colorado.edu)

Include a resume and/or relevant information related to job qualifications.



#### UCAR-NCAR Workshop Announcement

Matt Mayernik posted on Mar 15, 2012

Bridging Data Lifecycles: Tracking Data Use via Data Citations

April 5-6, 2012 at the University Corporation for Atmospheric Research in Boulder, CO.

Please join us for a workshop on data citation and data use geared to the interests of data manager specialists and librarians. Data citations are increasingly seen as being critical to enabling scientific results to be traced back to their underlying data. Data citations also promote the transparency of scientific work by making data more discoverable, and enable scientists, data managers, and data centers to be credited for producing useful data. Presentations will focus data lifecycles, current options for data citation and data reuse, and will highlight pilot projects in implementing citation standards and practices. Dr. Tim Killeen, Assistant Director for Geosciences at the National Science Foundation will provide the keynote address. Other presenters include Joan Starr of the California Digital Library and Mark Parsons of the National Snow and Ice Data Center. Additional speakers are listed on the workshop web page.

Limited space is available for attendees. More information and the registration form are at [http://library.ucar.edu/data\\_workshop/](http://library.ucar.edu/data_workshop/). Please register by March 23 to ensure your participation. Once registered, participants will be provided with logistical details and any updates to the agenda.

Please feel free to share this information with your colleagues. Contact Matt Mayernik at [mayernik@ucar.edu](mailto:mayernik@ucar.edu) or Karon Kelly at [kkelly@ucar.edu](mailto:kkelly@ucar.edu) if you have questions.



#### BESSIG Meeting Wed, March 21

Unknown User (wilson) posted on Mar 10, 2012

This month we are pleased to have some out of town colleagues speaking about a variety of topics related to content management.

#### Speakers:

Jerry Pan, Oak Ridge National Laboratory (ORNL), "[Geoscience Data Repository in Digital Object Model and Open-Source Frameworks: Provenance Applications](#)"

Ajinkya Kulkarni, University of Alabama, Huntsville, "[Utilizing Drupal, an Open Source Web Framework to Support Science 2.0 Projects](#)"

Erin Robinson, ESIP Foundation, "[ESIP Commons: Publishing Non-Traditional Content](#)"

**When:** Wednesday, March 21, 4:00 - 6:00

**Where:** [Boulder Outlook Hotel](#)

#### Schedule:

4:00 - 5:00: speakers

5:00 - 6:00: social



#### BESSIG Meeting Wed, February 22, 5 PM

Unknown User (wilson) posted on Jan 27, 2012

We are dovetailing with the UCAR SEA 2012 Software Engineering Conference, which allows us to have these speakers from out of town. This is why **we are meeting at 5:00 instead of our usual 4:00**. (And also why we are meeting on the 4th Wednesday of the month instead of the 3rd.)

#### Speakers:

Jay Alameda, NCSA, "[Bringing Modern Tools to High Performance Computing - with a Spotlight on XSEDE and Blue Waters](#)"

Chris Mattmann, JPL, "[Understanding how to Best Leverage Open Source Data Management Software: A Roadmap](#)"

**When:** Wednesday, February 22, 5:00 - 7:00

**Where:** [Boulder Outlook Hotel](#)

#### Schedule:

5:00 - 5:30 gather, social

5:30 - 6:00 Jay Alameda

6:00 - 6:30 Chris Mattmann

6:30 - 7:00 social



**BESSIG Meeting Wed, January 18, 4PM**

Unknown User (wilson) posted on Jan 02, 2012

### **BLAST (BESSIG Lively Activists Speaking Topically)**

#### **Speakers:**

Mike Daniels, NCAR, *"From Sensor to Archive: Data Flow, Tools, and Management of Observational Data at NCAR's Earth Observing Laboratory"*

Ruth Duerr, NSIDC, *"Stewardship and Access - Activities of the Federation of Earth Science Information Partners (ESIP) "*

**When:** Wednesday, January 18, 4:00 - 6:00

**Where:** [The Boulder Outlook Hotel](#)

**Schedule:** 4:00 - 5:00 talks, 5 - 10 minute business meeting, 5:00 - 6:00 social (more or less)



**BESSIG Meeting Wed, November 16, 4PM**

Unknown User (wilson) posted on Nov 08, 2011

### **Panel Report and Discussion on Recent NSF EarthCube Charette**

According to the EarthCube web site, <http://earthcube.ning.com/>, "The goal of EarthCube is to transform the conduct of research by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the Geosciences."

During November 1 - 4, 2011, the NSF held the EarthCube Charette in Washington, D.C. The Charette was intended to be "the first stage in an iterative process to build a community-based cyberinfrastructure," and an "opportunity for face-to-face interaction and feedback among groups that are suggesting designs for EarthCube and a broad, diverse set of the geosciences and cyberinfrastructure communities."

Panel members for our meeting are Charette attendees who will share their thoughts and experiences. Perhaps we'll finally understand the answer to the question, "What is a charette, anyway?"

#### **Panel members:**

**Michael Wiltberger**, HAO/NCAR

**Russ Rew**, Unidata/UCAR

**Seth McGinnis**, CISL/NCAR

**Scott Peckham**, CSDMS/INSTAAR/CU Boulder

**Chris MacDermid**, CSU CIRA Fort Collins/NOAA ESRL

#### **Location:**

This meeting will be held at the [Boulder Outlook Hotel and Suites](#) in the Chatauqua room. The Hotel is near Aurora and 28th St Frontage Road.

#### **Schedule:**

4:00 - 5:00: panel and discussion

5:00 - 6:00: social

If you attended the Charette and are interested in participating in the panel please contact me ASAP! The only requirement is a willingness to tell us a little about your experience at the Charette. Thank you!



**BESSIG Meeting Wed, October 19, 4PM**

Unknown User (wilson) posted on Oct 13, 2011

### **NetCDF for the Masses**

*Ed Hartnett, Unidata, UCAR*

"Why should anyone consider using netCDF to store their data?"

**NOTE THE \*NEW\* NEW LOCATION:** This meeting will be held at the [Boulder Outlook Hotel and Suites](#) in the Chatauqua room. The Hotel is near Aurora and 28th St Frontage Road\*.

We are trying this new location that offers a private room with food and beverages in a non smoking environment. The bar menu includes starters, soups, salads, sandwiches, burgers, burger alternatives, steaks, desserts, and 9 beers on tap, most of them local. The Outlook has graciously offered to let us use the room for free for this meeting. If we can generate sufficient sales for them, they might let us come back.

Schedule:

4:00 - 5:00: talk and discussion

5:00 - 6:00: social

Hope to see you there!

\*If driving from 30th and Colorado, don't try to turn left from Colorado onto the 28th St Frontage Road, it is illegal. Instead head south on 30th and turn right on Aurora.



#### [GEOSS Workshop XLIII Sharing Climate Information and Knowledge](#)

Unknown User (wilson) posted on Sep 19, 2011

**Place:** NCAR Center Green, Boulder Colorado, USA

**Time:** Friday, 23 September 2011, 8:00am-5:30pm

**Fee: Free** - all are welcome (registration requested, see below)

This one-day workshop is focused on how the Global Earth Observation System of Systems (GEOSS) can contribute to studying climate change, how GEOSS can enable broader community access to climate information and knowledge; and how scientists can manage and work with integrated data sources having different levels of quality.

This workshop is free and open to the public.

To confirm attendance, please register here: [https://portal.opengeospatial.org/public\\_ogc/register/110923geoss.php](https://portal.opengeospatial.org/public_ogc/register/110923geoss.php)

For the complete agenda, please visit: <http://www.ieee-earth.org/event/geoss-workshop-xliii-climate>

(Posted for Siri Jodha Khalsa)



#### [BESSIG Meeting Wed, September 21, 4PM](#)

Unknown User (wilson) posted on Sep 16, 2011

### **Panel discussion on Peer Review of Scientific Data**

"Data centers have few established practices for peer review of data. Indeed there is no clear definition of what peer review of data really means. Is it a review of data accuracy or validity, or is it a review of data documentation to ensure complete description of uncertainty and context? Despite these challenges, scientists and data managers have a professional and ethical responsibility to do their best to meet the data publication goals asserted by AGU." [Parsons, M. A., R. Duerr, and J.-B. Minster (2010), Data Citation and Peer Review, *Eos Trans. AGU*, 91(34), 297, doi:10.1029/2010EO340001.]

**Wednesday, September 21, 2011, 4:00 - 6:00, Millenium Hotel's Boulder Creek Living Room.**

Note that the Millenium is graciously providing us this room including wireless connectivity without charge. Please attend this meeting and consider purchasing a drink or appetizer to support the Millenium and our continued use of this room and its amenities.

**Who:**

Mark Parsons, Program Manager, Senior Associate Scientist, NSIDC

Matt Mayernik, Research Data Services Specialist, NCAR library

Eric Kihn, Deputy Director of the National Geophysical Data Center, NOAA

Karen Simmons, Experiment Manager for Galileo UVS/EUV instruments and archivist for data from Voyager and Galileo. Also, current archive specialist on an archiving grant through NASA for very old (1969, 71, 75 etc) planetary data., provider to NASA PDS, LASP

Rob Wilson, Scientist studying magnetospheres of the outer planets, using PDS data, LASP

**Topics for discussion include:**

- Should data peer review differ from conventional document review? If so, how and why? What does peer review of a dataset really mean?
- How are peer-reviewed datasets certified? Everyone knows the peer reviewed journals and has some notion of their impact. How does that work for data?
- Who conducts the review? Who is the "editor" to coordinate the review?
- Does peer review imply more responsibility or accountability for the data "publisher"?
- One critical discussion should be about what facility is accepting the data and how they operate, what requirements they have for peer review and discussion of how well that's worked in the past. Their perspective is often quite different from that of a scientist.

Please join us in this discussion!

The discussion will be followed by a short business meeting for feedback and planning purposes.



**BESSIG Meeting Wed, August 17, 4PM**

Unknown User (wilson) posted on Aug 01, 2011

**NOTE THE NEW LOCATION:** This meeting will be held at the **Millenium Hotel** in the Boulder Creek Living Room. This private room is adjacent to a bar where food and beverages can be purchased and brought into the meeting.

Note that the Millenium is graciously providing us this room including wireless connectivity without charge. Please attend this meeting and consider purchasing some comestible fare to support the Millenium and our continued use of this room and its amenities.

**Who:**

Gary Strand, *The NCAR Community Earth System Model and the IPCC Fifth Assessment Report*. [Gary's slides: [Gary Strand Aug 17 2011 BESSIG presentation](#)]

Siri Jodha Khalsa will give a short summary of a workshop he led at this summer's INSPIRE conference titled "What is Interoperability and How do we Measure it?" followed by an open discussion of the topic. (INSPIRE is the panEuropean spatial data infrastructure initiative. Workshop presentations are available at: [INSPIRE conference presentations](#).)

Leonard Sitongia, *Challenges in Legacy Software and Scientific Methods*, followed by Steve Olding from the NASA Earth Science Data Systems Technology Infusion Working Group (via Skype), *A Toolbox for Organization-wide Infusion of Data Systems Technologies*.

**When:** Wednesday, August 17, 2011, 4:00 - 6:00

**Where:** The Boulder Creek Living Room of [The Millenium Hotel](#). Please see the web site for directions to the hotel. The Boulder Creek Living Room is found by entering the lobby, passing through the doorway on the right, then left through another doorway into the room.

Food items available include: appetizers, salads, a burger, wraps, and entrees, ranging in cost from \$6 to a \$29 entree. Most items are \$10 or less.



**BESSIG Meeting Tue, July 19, 4PM**

Unknown User (wilson) posted on May 20, 2011

**Another BLAST!**

A time to listen and discuss as members present their work and areas of interest.

**Who:**

- Doug Fowler, *GLAS and MODIS Data Management and Services at NSIDC*
- Jeff McWhirter, *A RAMADDA Blast*
- Anne Wilson, *Report on Earth Science Information Partners (ESIP) Federation Summer Meeting*

If anyone else is interested in speaking at this meeting, please contact Anne.

After the presentations we'll have a short business meeting for planning purposes. At the meeting we'll discuss the possibility of holding future meetings at the Millenium Hotel.

**When:** Tuesday, July 19, 2011, 4:00 - 6:00 PM

**Where:** Room A200, LASP Space Technology Building, 1234 Innovation Drive, Boulder, Colorado

LASP is located on the University of Colorado's East Campus Research Park. A map and directions are available on the [LASP Address and Directions](#) page. The building sits just off Colorado Avenue, immediately adjacent to CU's new, very large biotechnology building that is currently under construction. Look for the tallest cranes in town, we're next door to the East. Free parking is available.



**BESSIG Meeting Thu, June 23, 4PM**

Unknown User (wilson) posted on May 20, 2011

**BLAST: Bessig Lively Activists Speaking Topically**

Come listen and discuss as Vince, Jeff, Steve, Peter, Doug, and Kimberly present their work and areas of interest. Followed by 10 - 15 minute business meeting to discuss future planning. Snacks will be provided.

**When:** Thursday, June 23, 2011, 4:00 - 6:00 PM

**Where:** Room A200, LASP Space Technology Building, 1234 Innovation Drive, Boulder, Colorado

LASP is located on the University of Colorado's East Campus Research Park. A map and directions are available on the [LASP Locations and Directions](#). The building sits just off Colorado Avenue, immediately adjacent to CU's new, very large biotechnology building that is currently under construction. Look for the tallest cranes in town, we're next door to the East. Free parking is available.



#### [May 18 Planning Meeting Gets Items on Calendar](#)

Unknown User (wilson) posted on May 20, 2011

At the BESSIG planning meeting on Wednesday we scheduled meetings for Thursday, June 23 and Tuesday, July 19, 4:00 PM. (See our [Calendar](#)). These meetings will be BLASTs: BESSIG Lively Activists Speaking Topically. That means people are on deck to present their work, area of interest, or whatever. The goal is to for us to get to know each other and our areas of expertise better, and to allow for discussion. Thus, speakers will present for as short or as long as they wish as long as it's "not too long" (TBD). Some speakers may get bumped until later, and that's okay. We'll also work in some time for socializing.

We have six speakers that will be ready on June 23, see the [Calendar](#) item to see who they are. We're looking for people to sign up for July 19. If you're interested in sharing with us, go to the that calendar event and put your name on the list under the "Description" section.

We also updated the table of [Meeting Topics](#) with some ideas to incubate. We found it particularly liberating to nominate topic organizers that weren't actually present at the meeting. Feel free to update the table with your ideas about meeting topics and organizers.

Also, the group paid it forward by contributing \$40 to a kitty (including pledges) for snacks for our next meeting. Thanks, guys!!

An important meeting topic was: beer. There is interest in having beer at our meetings, but institutions have severe limitations on alcohol. Here at CU a recent update of the alcohol policy muddled the waters sufficiently that what is and is not allowed is unclear and is under investigation by our staff. More on this to follow. Suggestions?



#### [BESSIG Meeting Wed, May 18, 4PM](#)

Unknown User (wilson) posted on May 16, 2011

**Purpose:** Plan meeting topics for upcoming year, and hobnob

**When:** Wednesday, May 18, 2011, 4:00 - 6:00 PM

**Where:** Room A200, LASP Space Technology Building, 1234 Innovation Drive, Boulder, Colorado

As we've not identified a talk topic for this meeting, we'll use our time to plan meeting topics and, as much as possible, a schedule for the upcoming year. Please come to the meeting with your calendar and specific ideas about meeting topics you would like to hear about or might be interested in presenting. We'll also get to know each other a little better over some light snacks. If all goes well we could be done early.

I have started a [Meeting Topic and Organizer List](#), check it out.

If you can't make this meeting but would like to provide input, send an email to [bessig.info@lasp.colorado.edu](mailto:bessig.info@lasp.colorado.edu) (which currently comes to me, Anne).

LASP is located on the University of Colorado's East Campus Research Park. A map and directions are available on the [LASP locations page](#). The building sits just off Colorado Avenue, immediately adjacent to CU's new, very large biotechnology building that is currently under construction. Look for the tallest cranes in town, we're next door to the East. Free parking is available.



#### [BESSIG Meeting planned for Wed, May 18](#)

Unknown User (wilson) posted on Apr 21, 2011

At our most recent meeting we decided to meet again on Wednesday, 5/18 from 4:00 - 6:00 with this schedule:

4:00 Talk (TBD)

5:00 Business meeting

The group agreed to discuss talk topics and speakers via the web site.

In retrospect, I think this is a very tight schedule, as in general it would be good to provide a speaker with a two week lead time. Plus, we have to find someone that can come at the particular hour. I'm a little nervous about making this goal.

Please contribute to our discussion about identifying a speaker and topic for this upcoming meeting in the Discussion section of the web site. Thank you!

If anyone is available to sponser snacks or beverages for this meeting please contact me.

Anne



#### [Launch Meeting Witty, Erudite](#)

Unknown User (wilson) posted on Apr 21, 2011

I'm really pleased about our launch meeting last night. Thank you to all who attended and also those who showed interest but could not attend.

We had roughly 45 people in attendance from 10 organizations: LASP, NDP LLC, NEON, NOAA, NSIDC, CU Boulder Research Computing, Tech-X Corporation, UNAVCO, UCAR, and UCAR/Unidata.

Ted led us off with a talk about "Convergence and Trust in Earth Space Science Data Systems": [Habermann\\_ConvergenceAndTrust](#). "Convergence" refers to convergence towards standards and best practices that simplify exchange and use of data. Simplification of that aspect of science is helping us move from data to information to knowledge and finally to wisdom. Generally data producers are involved in transforming data to information, data consumers are involved in transforming information to knowledge, and it takes the community to translate knowledge to wisdom. This path leads to the

question of trust: given that we're in an age where people use data and may not know who the data provider was, let alone have communicated with them, how do we say with certainty that our results are correct? This leads to the issue of trust. The topic of trust, being an over arching issue, played a big role in our subsequent discussion at our business meeting. Discussion notes from that meeting will be available soon.

Dave Fulker entertained us with a witty, erudite joke followed by a talk on "Standard Mechanisms for Data Exchange", practices that became "standard" by virtue of being commonplace: [Fulker\\_StandardMechanismsForDataExchange](#). Dave made a comparison between the history of artifacts and what those artifacts enabled, e.g., writing instruments enabled numerical symbols, scrolls/books/libraries enabled tables and relations, etc. Currently, computers in science have enabled data management, analysis, and visualization as library functions or APIs, e.g., library functions for data formatting. Regarding data analysis and visualization, the power of a system is directly related to the generality and richness of its underlying data model. And now, the web has enabled near real time data flows and remote access to data.

These talks were relevant in our business meeting where we brainstormed about what this group might be or do, what our goals might be, who is available for contributing and what to do next. In Ted's terms, we are in a stage of ferment, trying to focus our energy as a community. Some common themes emerged, such as serving as a resource for those trying to learn how to build good systems. Another theme was trust: how is it currently earned, how can we ensure its viability in our escience world? Are there projects for us there? Notes from this section of meeting are available off the Discussion section.

We decided to meet again in month, on Wednesday, 5/18, from 4:00 - 6:00. (If anyone is available to sponser snacks or beverages for this meeting please contact me.) The first hour will be a talk. The second hour will be a business meeting. The group is tasked with identifying a speaker for the 4:00 time slot.

I really appreciate the strong interest and support I received in launching this group. I'd like to stress here again that this is a community effort. The more people contribute, the more successful we'll be as a group. The good news is that there is a good number of us such that each individual effort does not have to be huge. Please think about what you are able to contribute.

Watch our site as news about the upcoming 5/18 meeting develops.



[BESSIG Launch Meeting](#)

Unknown User (wilson) posted on Feb 18, 2011

**Purpose:** 1) To bring together people interested in scientific data in the Boulder, Colorado area for the purpose of sharing ideas and information, and 2) to gauge interest in this group and discuss its future.

**When:** Wednesday, April 20, 2011, 5:00 PM

**Where:** Room A200, LASP Space Technology Building, 1234 Innovation Drive, Boulder, Colorado

**Schedule:**

5:00 Arrive

5:15 **The Role of Standards for Data and Metadata**, by *Ted Habermann and Dave Fulker*

A hallmark of science in the Internet age has been the increasing importance of means for exchanging digital data. Indeed the significance is so great that Turing Award winner Jim Gray dubbed this data-intensive pattern the "Fourth Paradigm of Science." Tightly linked with effective data exchange is a critical need for standards such that data providers and data users are, in essence, talking the same language. The authors will cover two aspects of this need: 1) Habermann will discuss important progress, international in scope, on *standards for metadata* (i.e., information about data) that are being widely adopted; 2) Fulker will discuss software and protocols that have, through patterns of use, served to significantly standardize the *mechanisms for data exchange*, including how complex data structures are digitally represented and serialized.

6:15 Dinner

6:45 Business meeting: Group Planning, Goals

7:30 End

8:00 Building vacated

Dinner and beverages will be provided. Thanks to the Foundation for Earth Science for the food! **Please RSVP to Anne Wilson so the appropriate amount of food can be ordered.**

LASP is located on the University of Colorado's East Campus Research Park. A map and directions are available on the [LASP locations page](#). The building sits just off Colorado Avenue, immediately adjacent to CU's new, very large biotechnology building that is currently under construction. Look for the tallest cranes in town, we're next door to the East. Free parking is available.

Please note that we will need to lock our front doors at 7:00 PM. If you need to arrive after that, please contact Anne Wilson to make alternative arrangements.