

## [EarthCube Workflow Session](#) [1]

Submitted by superadmin on Fri, 2012-06-29 21:13 Wednesday, July 18, 2012 - 14:00 to 15:30

**Event:** [Summer Meeting 2012](#) [2]

**Session Type:** [Breakout](#) [3]

**Expertise Level:** [Beginner](#) [4]

**Identifier:** doi:10.7269/P36Q1V54

**Collaboration Area:** [Climate Education Working Group](#) [5]

[Earth Science Collaboratory](#) [6]

[Education](#) [7]

[Information Technology and Interoperability](#) [8]

### **Abstract/Agenda:**

This WG breakout meeting will continue a series of existing virtual participation meetings and F2F for the NSF EarthCube Workflows team. The efforts to date have focused on developing a strategic plan/vision, and series of vignettes for the ways that workflows, i.e., representations of control/data flow, semantics, provenance, data archiving, dissemination and other associated capabilities, play into the ever growing EarthCube community and suite of projects and systems. Our goals will involve audience participation, collection of vignettes, planning for the next meeting, and provision of input to the NSF.

### **Notes:**

#### **EarthCube Workflows**

- Supposed to be an open discussion regarding the EarthCube Workflow group
- NSA wants this to be a community effort

Goals of NSF:

- This of future initiatives from point of view from Geoscientist point of view
- What kind of cyber infrastructure will improve science
- Productivity could be improved

EarthCube Charrette:

- NSF decided to form four groups:
  - Data: mining, management, discovery
  - Workflows:
  - Semantics/ontologies
  - Governance

Workflows came up as an important topic

Scientific workflows are used to describe compose, and execute ensembles of scientific computations on distributed resources

- improve productivity and accelerate research

At Charrette:

- Workflow discovery
- Workflow metadata and provenance
- Workflow execution management

- Data management within workflows

### Question/Comment

Workflow discover, looking at workflows themselves so it's assumed that the workflow is stored and curated.

- provenance of workflow and provenance of data

EarthCube formed a steering committee

- some people had never worked with workflows and other had

Initial Roadmap is available on website:

- Understanding what are the requirements
- Living document
- Identifies grand challenges of next decade

Workflow questionnaire:

- putting together a questionnaire
- put together a number of questions that would be useful to understanding workflows in geosciences
- learned that people are very aware and continue to write scripts to transform data
- thought about when running larger models or cleaning, transferring, all of these steps take lots of time

What are workflows?

- used to describe computation
- prescriptive and descriptive
- can publish the provenance of results
- useful to capture pipelines

Use of Workflows

1. the center modeling paradigm
  1. a center designs the workflow to run specific models, the results executing the workflow are specialized data products that are then published by a community of users
  2. ex. Iowa Flood Forecasting models
2. the long tail paradigm
  1. Individual investigators run their own workflows. The investigators prepare the data themselves, integrate it with data from shared sources, etc...
  2. Ex. Tome Harmon Aquatic River Science, Sensor Network models
    1. Process system and analyses models
  3. comment, constraint with this is expertise, a lot of long tail scientists might say its over their head and that they don't have time for this

- 4.
3. the big data computing
  1. Very large datasets with computational intensive codes are routinely run, requiring high-end computing resources and scalable infrastructure
  2. Comment: If EC can provide a way to make this easier, people will be more inclined to do workflows for paradigm
  3. Comment: model of computation sometimes compatible and sometimes not, ways to come up with standardized way, so that it can work in multiple systems
  4. Comment: some people prefer to write their own since using a workflow system is often programming language specific
    1. Trying to find commonalities
  5. Workflow systems are experience based
  6. Planning problem, there can be an order but
  7. Question: Kepler
    1. People use commercial workflow systems
    2. Proliferation of workflows systems come because there are so many types of workflows
  8. Ex. Earthscope:
    1. Comprehensive understanding of active continental deformation from days to hundred to millions of years
    2. Need this type of infrastructure
4. the whiteboard paradigm
  1. A group of investigators in a collaboration, interdisciplinary investigation, drafts, a sketch of what the workflows look like
5. the metadata data rich paradigm
  1. want to keep track of metadata of all the data and produced by their analysis
    1. "the vitamins of data"
    2. geochemical data

### Discussion

1. These workflows are not mutually exclusive
2. Software marketplace aspect is seen as more of a feature that could apply to any of the paradigms
  1. Use case in the earth science collaboratory
  2. Ability to share those workflows
  3. Particularly interdisciplinary scientists:
    1. If processes can be put into instantiated workflow
    2. Quality filtering workflows
  4. software
    1. shared APIs for people to deposit workflows, if we could get all of those to work together
    2. if we could get the software to work together that would be helpful
  5. uniform way of extracting semantics
    1. to understand how workflow was processing data
    2. hard because software does things differently, might not be intuitive to people, example, how loops are created might seem simple in software, but really isn't
  6. would not make it just sensor data, but anything that's repetitive
    1. if you're collecting data once a month or once a year its important to record how
    2. its about automated management
  7. Are you doing anything with end users

### Workflow Vignettes

- collecting workflow vignettes from the community

- current and potential uses of workflows, highlights benefits,
- expose success stories

NSF guidance for roadmaps

- Roadmap
- Communication to community
- Requirements for group to be successful
- Potential solutions
- Process to get that requirement
- Timeline
- Management

Roadmap:

- A workflow working group with a number of task forces
- Communication, status & requirements, engagement, prototyping, assessment, community interaction
- Set of professionals paid to listen to scientists, can
  - Center that is formed with full time people (Synthesis center)

Risk:

- Not to capture user requirement
- Groups doing workflow technology forced to compete with one another
- A lot of it is a social issue, people need to want to collaborate

**Next Steps:**

- Actively revising roadmap, getting feedback
- Continue to collect workflow vignettes
- Continue to collect responses to workflow questionnaire
- <http://earthcube.ning.com/> [9]
- AGU fall meeting session on EarthCube workflows in December
- Fall/spring workshops

**Comments/Question:**

Specific ideas on how to engage people invited to workshops

- participants not wanting to be inundated with questions
- if I'm a scientist, do I have a brokering problem? I know I have data problems
  - everything is a data problem in some way

Europeans have been very involved with eScience, have they done anything with workflows?

- no one has done anything with workflows
- see more eScience for social science and humanities
- has not been this kind of concerted effort to create workflows for science

**Session Leads:**

**Name:** [Chris Mattmann](#) [10]

**Organization(s):** [Jet Propulsion Lab](#) [11]

**Name:** [Paul Ramirez](#) [12]

**Organization(s):** [Jet Propulsion Lab](#) [11]

**Name:** [Yolanda Gil](#) [13]

**Organization(s):** [USC/ISL](#) [14]

### Notes takers:

**Name:** [Angela P. Murillo](#) [15]

**Email:** [amurillo@email.unc.edu](mailto:amurillo@email.unc.edu) [16]

**Creative Common License:** Creative Commons Attribution 3.0 License

**Teaser:** Developing a strategic plan/vision, and series of vignettes for the ways that workflows play into the ever growing #EarthCube community

**Keywords:** [EarthCube](#) [17]

**Source URL:** <https://commons.esipfed.org/node/485>

### Links

[1] <https://commons.esipfed.org/node/485>

[2] <https://commons.esipfed.org/event/summer-meeting-2012>

[3] <https://commons.esipfed.org/session-type/breakout>

[4] <https://commons.esipfed.org/taxonomy/term/260>

[5] <https://commons.esipfed.org/collaboration-area/climate-education-working-group>

[6] <https://commons.esipfed.org/collaboration-area/earth-science-collaboratory>

[7] <https://commons.esipfed.org/collaboration-area/education>

[8] <https://commons.esipfed.org/collaboration-area/information-technology-and-interoperability>

[9] <http://earthcube.ning.com/>

[10] <https://commons.esipfed.org/node/330>

[11] <https://commons.esipfed.org/taxonomy/term/197>

[12] <https://commons.esipfed.org/node/455>

[13] <https://commons.esipfed.org/node/484>

[14] <https://commons.esipfed.org/taxonomy/term/319>

[15] <https://commons.esipfed.org/node/633>

[16] <mailto:amurillo@email.unc.edu>

[17] <https://commons.esipfed.org/taxonomy/term/314>