AGU 2013 Town Hall:

Building a Sediment Experimentalist Network (SEN): focus on the SEN Knowledge Base

Wednesday, December 11, 2013, 12:30 pm

Conveners: Wonsuck Kim (University of Texas, Austin), Leslie Hsu (Columbia University), Brandon McElroy (University of Wyoming), Raleigh Martin (UCLA)

Notes taken by McElroy and report prepared by Hsu.

The Town Hall "Building a Sediment Experimentalist Network (SEN)" was convened to further promote the SEN, inform the community about the planned efforts, and solicit feedback on the SEN Knowledge Base. As an introduction, the three major components of SEN were presented: (1) SEN-KB: Knowledge Base (a data catalog and wiki for discovery and sharing of data and methods), (2) SEN-ED: Education and Standards (consensus building and outreach about data & metadata management), and SEN-EC: Experimental Collaboratories (studying Earth surface grand challenges through shared and in-parallel experiments through webinar systems and physical visits).

The SEN team showed a presentation of the Knowledge Base, a new online platform consisting of a Data Catalog, which describes and links to experimental data sets, and a Wiki, which describes experimental methods and equipment. The data catalog includes descriptive metadata for datasets such as the creator of the data, location, time, links to other resources, keywords, rights, and degree of processing: raw, processed, derived. The user-interface for submission was also previewed.

The SEN Team acknowledged that use and success of the Knowledge Base will depend on community buy-in and perceived benefits. The Knowledge Base is planned to be an interactive place for sharing methods and stimulating ideas. Use of the knowledge base will ideally satisfy most requirements for data management plans in proposals.

The Knowledge Base is a "next generation" version of the SEN fusion table, which in turn came from the original NCED (National Center for Earth-surface Dynamics) data repository. At this time the complete solution for data storage, archiving, and interoperability with broader systems does not readily exist for the SEN community, nor has the degree of moderation been determined, but these projects are steps toward that goal.

Issues and concerns raised by the audience during the question and answer session included: inadequate data storage, data quality assurance, need for rich of metadata,

the importance (or not) of storing raw data, cost-benefit for filling out lengthy metadata descriptions, the need for light moderation of the knowledge base, and history and version control of the Knowledge Base content.

Q & A from the audience

1. **Data storage:**
    a. Q: I would be interested and use this platform, but what if we have several Tb of data and no way to store and serve it? Researchers at smaller institutions in particular have less support for storage space.
    b. A: SEN does not have a solution for this yet, but we are looking to partner with and leverage other projects or programs that may supply this.
2. **Data Quality:**
    a. Q: How will we address data quality issues of what is submitted to the knowledge base? How will people know to trust data?
    b. A: this is a big problem across earth sciences – we might see how informatics folks are thinking about this. But we are not looking to necessarily exclude data right now.
    c. Comment: Peer review of datasets should be considered. I would be willing to review datasets of my peers. Comment: This is a huge topic elsewhere and we should keep our eye out to follow this trend.
3. **Richness of metadata:**
    a. Comment: I've had the experience of storing data with the CZO system. Because CZO covers lots of different types, I've had the issue not having enough metadata options. The breadth of metadata fields needs to be very thorough, as much as possible.
4. **Importance of Raw data:**
    a. Comment: We are all using a wide range of equipment. If I give you raw data you are likely not to be able to use it without my explicit info on how to process it. So raw data may not be important to store in SEN, but rather some level of processing that is "usable."
5. **Completeness of metadata**:
    a. Comment: The completeness of metadata documentation relates both to quality of data and also how much time is available to spend on documentation. We should allow users to first provide a minimum amount and then provide all the data including the raw. Thus the knowledge base will be communication channel.
6. **Degree of Moderation:**
    a. Comment: To avoid subprime data in the system, a moderator would be good to weed out bad contributions. However it shouldn't be a police force, we should minimize the time commitment for editors.
7. **Version control and history**:

   a. Comment: I think that aggressive version control is important, so if changes are made, the old version still exists. It would also be good to notify users of significant updates to their entries.

8. **Scope of SEN**
   a. Q: What is the scope of the data that SEN will accept, does it include field data?
   b. A: Right now, we need to stick with laboratory experiments as a test case, to demonstrate what we can do, before it can be expanded to other things. We can expand later to field data if our small community can make this work. Right now there are no limits except involving sediment and a laboratory.