

[Developing the Big Earth Data Initiative \(BEDI\) Common Framework \(BCF\) in the U.S. Group on Earth Observations \(USGEO\) Data Management Working Group \[1\]](#)

Submitted by erinmr on Sun, 2014-06-29 15:24 Tuesday, July 8, 2014 - 16:00 to 17:30

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

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

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

Collaboration Area: [Data Preservation](#) [5]

Abstract/Agenda:

The purpose of the Big Earth Data Initiative (BEDI) is to improve the discoverability, accessibility, and usability of Federal data and information products derived from civil Earth observations. This activity is an OSTP initiative which will be coordinated through the US Group on Earth Observations (USGEO) Subcommittee of the National Science and Technology Council (NSTC) Committee on Environment Natural Resources and Sustainability (CENRS).

The USGEO Data Management Working Group (DMWG), co-chaired by NASA, NOAA, and USGS, has been analyzing existing agency policies, standards, services, and tools and considering them for inclusion in the recommendations for a BEDI Common Framework (BCF). This session will provide an opportunity for participating agencies and other interested parties to share recommendations for the BCF, plans for implementing the BCF in agencies and data centers, and other ideas to make BEDI a successful venture.

Notes:

Developing the Big Earth Data Initiative (BEDI) Common Framework (BCF) in the U.S. Group on Earth Observations (USGEO) Data Management Working Group

Notes:

Introduction by Jeff

Recent initiatives for public access based on a white house memo. They had requirements for open data policies as well, things that did not often have details on implementing them. There is a proposed initiative from BEDI on Big Earth Data Initiative. There is also the climate data initiative which includes outreach. These activities relate to work done by ESIP.

Background on US Geo - with high level representation from all related agencies and three working groups.

Earth Observation assessment - EOA 2, which will take place every three years, next due in 2015. They deal with data sets needed to make assessments in relation to these areas. What products and services are needed, and what would happen if they were lost?

There was a document made available last year and another will be released soon for civil earth observations.

Big data earth data initiatives - NASA, NOAA, USGS - common approaches to making data accessible, discoverable, usable. Gave some examples of high impact sets. And they would like to create a common framework. Someone asked how long the group would be working? They all put in proposals this year but only a few got funds, and they are trying again next year. That this would be a three to five year project. It is not yet determined until they get the budget. And trying to do some of these things anyway.

BEDI common framework - any user tools should be able to connect to any EO data source. On slides, specific user tools, and a diagram of what the framework would do to provide access to the data from the various groups.

NOAA activities that are already relevant to this effort. Not everything is archived, working on that little by little. Working on a NOAA data catalog. Have a few committees and are working with other agencies to create DOI's. See slides for more information. NOAA will create a place for hosting data for smaller NOAA divisions that do not have their own storage. Will be using ESRI but would like to make an open source option as well.

Curt

NASA got a bit of funds to work on this project. Curt is reporting on these activities. Catalog and data discovery. They have a project working on DOIs. They are also working on a common model for DOI landing pages. Not to make them identical but require certain content and recommend other aspects to make them consistent. Integrating schema.org tags - those recommended by search engine providers that can extract more structured information from webpages. Make it easier for people to find the pages based on key words. Also populating data catalogs.

Working with OPeNDAP standard for data. Will fund enhancements which will benefit users and providers. Make it easier to access and easier to use. There has been criticism that NASA is too focused on research communities, and what can they do to make their research accessible by the broader community? Also implementing JSON which is easier for web developers to use.

GIBS - would like to extend the coverage to cover more like high impact data. Prioritization. They can already see mobile apps using this data, and would like others to take advantage of the software to leverage this data and make it more public.

CEOS WGISS Integrated Catalog (CWIS) funded by NASA and others, would like to see more data sets part of this quick catalog.

Participating in open standards development. This includes developing standards for NASA and beyond to other agencies.

EDG - Environmental Dataset Gateway from EPA. Gave some similar examples from this project. EPA released a new version of their EPA metadata Editor. It is encouraged to be used by others to help you make better metadata for data sets based on the ISO standards, they are suggesting it to many of their smaller programs.

Sky

BEDI in the USGS perspective. What they are doing with no new funding, based on the open data directives.

Cartoon of what they envision of a future thing they would like to get to, lots of catalogs and indexes where locations of materials are listed, and they want to aggregate them into higher level catalogs. They want to include state government records and other groups outside of the federal government. This will be integrated into some larger catalogs. Some past efforts were successful, but they need to determine how they materials get up there and become part of catalogs like GEOSS and other data brokers catalogs. These drive other variations (see slides). Instead of a one stop shop, a variety of options.

Example of something they are working at the USGS today. Public data under the open directory, but they are an agency under bureau. They want to create a data management dashboard which would allow people to log in and create metadata. They have a DOI tool, and online metadata editor they would like to plug in to this. Behind all of that, they have a lot of messy things. Scattered in many programs throughout the USGS. INDEXING data in to the new science data catalog. This is CSDGM data into a single system.

They are using JSON to share the DOIs outward. It has a nominal data catalog and some backdoor access still.

Dashboard - examples on slides. They are giving feedback about broken links etc. And leveraging NOAA tools to look at how complete the records are. The outward interface is shown in the slides. Showed comparison searches in google and the dashboard. and other resources. Showing the differences where the results do not exactly match up. Google is doing a pretty good job, not sure about other catalogs.

Are they barking up the right tree? Are they improving discoverability, usability etc.? And if not, why not? Are we spending too much time focusing on ... see slides for list. Is it inertia, execution or something else? They have good foundational pieces, relates back to the digital government strategy from a few years ago.

How this relates to ESIP - Use cases and interoperability experiments etc. One example is from Rich Signell of the USGS. Testing open standards to real world problems. Looking across several catalogs and looking for data that would work, querying them and determining in workflows how to tie them together.

Questions?

Question - When going from catalog to catalog, which steps increase value of the information? Sky - to be honest, the further we got from the originating area the less reliable they became. Less context of the people in the organizations listed and key words or thesauri, or with facets to drill into data.

Jeff said some of the catalogs are general purpose. Which creates some dilution. But also specifically saw a dilution in what the open data policy was telling us to do. With the basic requirements. And that would make them dumb down the data they already had. Now trying to connect catalog to catalog instead of delusion as you go up the chain.

Curt said as an advocate for aggregators, there is some value in the aggregation itself. Like look at google - there are people who will go to google and not USGS or NOAA to do searches. Data.gov is to have a little bit of metadata for everything in the federal government.

Steve - we are finally at the point where metadata is flowing between these catalogs. Is a good thing. Now we are refining how we define the standards. Sky said it is just an implementation problem. Should put effort into many areas, but make sure full slate of information goes there.

Rama asked what the there was in the full suite of metadata goes there? Sky - that we don't dumb down our data in interdepartmental catalogs.

Question - what work have you done to, what type of commitments that the full suite is going in to the registry. Sky spoke about this, talked about a bit of fuzziness in the process, but focusing on highest level of metadata as it goes out. Quality over quantity.

Jeff said NOAA is very committed to seeing that their data is represented in GEOS. Traditionally they have done this and it was labor intensive. And required to put it in multiple systems. He would like to see this done easier. And if you tell a local catalog, it flows up to GEOS and you don't have to do it again. He would like to see it another way, if you don't edit or share that is listed and puts the burden on them.

Denise - loved Sky's slide of questions. How can we go about measuring the impact of this work? Sky - there are a lot of ways to go about this. Interoperability projects, based on a large enough scale scenario, like disaster response, which requires data pulled from many repositories and feed them into one system and make them usable. And that you can count on the quality based on the provenance.

Jeff - as we start with some of these DOI data sets, we can track usage of the data sets. It is hard to measure the impact, you can kind of tell if people are using your data but if we are using our own metadata to answer questions from above, it was automatically generated, it would save labor and time. But hard to measure that impact.

Question - liked what Sky said about interoperability experiment. Wondered what impact BEDI will make on these types of issues. Some are documentation, project pages, services, and understanding what kind of information is in the different catalogs and the quality, would be important to know. So that when you exercise it in a use case, you can find out if you are getting the right information and other questions like that. When it is all mixed up, it can be hard to pick through.

Sky - that is where usability and accessibility comes in. If you look at some of the intergovernmental catalogs, you run a search, get a lot of listings and some facets to use to narrow those down. If you look at the results, you can see a series of buttons that take you somewhere else, or output a csv, or they point to a service. But how do we look at the catalog and metadata and know where we have data represented even though it is not in the facets.

Questioner - with these use cases you would have an idea of what the different types of users might need.

Curt - we have been looking across these different activities - why do people need this data (use cases), prioritizing the data to inform those use cases, and are there characteristics of the data that lets them be used in these use cases. We can provide that information back to a data center or agencies. If you enable y you will have use case z. We can encourage implementation of this data. Given various scenarios as opposed to a dictation from on high.

Question - prioritization of use cases? how will you do that? Curt - Jeff pointed out a few examples, and with the climate data initiative we are talking to end users, so if you have some ideas, bring them up. There is not one mechanism.

Question - maybe we can be us, as we represent a lot of different groups. BEDI and ESIP - have more sessions in the future, bring them up on the cluster calls, etc.

Jeff - the open call said we were supposed to get feedback from users, but they don't know the most effective way to do that. To decouple the feedback from web servers. This might be subject matter experts only and not commercial etc.

Rama - in ESIP, it is generally data providers as opposed to end users. It is not a big user community. so if you have money to target users communities to provide information for use case, that would be useful. Curt - NASA and NOAA are doing stakeholder sessions.

Sky - lots of activities along these lines at Earthcube. What are they doing, what do they need? And they have issues workshop reports which the EarthCube building blocks are supposed to be based out of. It is mostly academic focused but others are included. ESIP is a great nexus point through the clusters and working groups to bring this together and get to specifics. We need to figure out if we are data science and provider types, we need to work on ways of opening out doors more broadly. And get meaningful involvement?

Question - what they say is constrained by what they know. So they might not know what to tell you.

Someone answered it has to do with the problem you are trying to solve.

Curt - where do we find these good use cases?

Question - is the data management problem fixed? Curt - knowing what to do and doing the right thing. We have been lagging behind though on doing the right thing. There is a lot of history behind metadata in this community, the new incarnation of this has included a lot more areas than before.

Earth Observation has brought in other smaller groups that might not have metadata. Questioner - do we know enough on the first step to say we are solid or sustainable? Sky - it is not a research problem as much as an an engineering or social engineering question. We have underfunded data management over the years. We have not invested as significantly as we should have. We are doing

data preservation and rescue work on things that are only a few years old.

Question - There is a gap between data providing and users. Building catalogs has useful areas but getting a hold of area expertise as to why I would use one data set over another. Or model over another. That takes a long time for them to explain, so how do we capture this use or limitations and share that? That could be a focus in the next 20 years. A model output is a model input for another group. We are not qualified to address some of those questions by experts. Who build a model for a specific result, and citizens or businesses have basic questions that are less robust. So it is great that people are interacting, but we need to focus on a data base of what we are generating as agencies and serving it to people, citizens and corporations. Maybe some working groups can be established along these lines? With resources available to create these smaller groups. Things that can be used to create a knowledge base. There has been a lot of progress in catalogs. But what is the next step? Chris Lynnes is talking about this tomorrow.

Sky - the idea of a knowledge base - what is working and what is not? Will not get funding to build a synthesis institute, but we need to do that incrementally that we can use to evaluate the system as a whole. Recent example where people had to scramble to bring data together for a specific trip to produce some impact maps, then brought data together and decided could not do modeling properly. They identified 5-6 critical things that they don't have the right data, not in the right format, or the model is not good enough and needs further research. What is the clearing house for that? With a collective response?

Comment - cautionary tale from someone in the audience. Too many products, need to bring users together to determine which are of value. They said all are good, and had ideas for new products. Positive story, but also they will provide new things for you to do. Which is good, but maybe not enough resources to provide these things.

Jeff - comment, we have the data management working group but also an assessment group. This group is more, given a product was created, we should make it available. In some standard fashion. Access and discoverability level and not why are we doing this.

Question - big business would already have a why question before creating things.

Other comment from the crowd - they way science is conducted, traditionally the research scientists, the data management concerns, they don't want to be bothered by it. So the intersection is thin and needs to grow.

Curt - getting back to the modeling question - pointed to the US-GEO sub committee, and there is another which focuses on modeling. Leaving it to them by this group so not to step on their toes.

Question - working on the belmont forum - just got around to talking to use cases, with the belmont challenge, to move forward to pick scenarios that would maybe be another good source.

Question - google public data offering - a lot of it is business level 5 products. We produce 1,2,3. This is by location and time. I want to see data by location and time - they can convert it in this system that makes it easier for end users. It does not have the details, but they do not want to see them anyways. Something they can consume and understand. Where they can see the sea ice being consumed etc. Which then leads them to go look for more.

Ethan - use case areas, have use cases been derived for societal benefit areas? Curt - they have been looking at the questions to the data sets to see how they might inform the questions.

Sky - these things will matter, it is a three year cycle, and the first one in 2012 needs work, but it will be a tool for other higher groups to figure out budgets etc. Identifying the systems and data files was work too.

Ethan - competing prioritizations, and needs to be agreed upon by a number of agencies. Like defining use cases... Curt -- NASA has been mapping products against needs areas. Ethan - maybe use cases are a different animal than the objectives need. Comment - use cases are a higher level of granularity.

Comment - the assessment could be tied to gether with the use cases. And the objectives, when doing the assessment do they have use cases to do this, specific ones in mind? They would be good use cases if they are.

Comment - The larger use organization did something a few years ago. That conversation is happening. Curt - we haven't done that yet. Commenter, there is a lot of overlap and based on (??)

Commenter - the usgs has the money to send people to these meetings and there is a reconfigured geo based on this, the societal benefits board is trying to work on better definitions of what is being produced. So maybe for the US part (of GEO) how does it all work, for ESIP at least, if some energy was used to put in to how it is looked up on the website and how can I help.

Comment - COPEUS working group is working on this and is meeting friday morning to discuss if people want to get involved.

Session Leads:

Name: [Curt Tilmes](#) [6]
Organization(s): [NASA Goddard Space Flight Center](#) [7], [USGCRP](#) [8]
Email: Curt.Tilmes@nasa.gov [9]

Name: [Jeff deLaBeaujardiere](#) [10]
Organization(s): [NOAA](#) [11]
Email: jeff.delabeaujardiere@noaa.gov [12]

Name: [Sky Bristol](#) [13]
Organization(s): [U.S. Geological Survey](#) [14]

Notes takers:

Name: [Sarah Ramdeen](#) [15]
Organization(s): [School of Information and Library Science UNC-CH](#) [16]
Email: ramdeen@email.unc.edu [17]

Creative Common License: Creative Commons Attribution 3.0 License

Teaser: The purpose of BEDI is to improve the discoverability, accessibility, and usability of Federal data and information.

Accepted:

Source URL: <https://commons.esipfed.org/node/2563>

Links

- [1] <https://commons.esipfed.org/node/2563>
- [2] <https://commons.esipfed.org/2014SummerMeeting>
- [3] <https://commons.esipfed.org/session-type/breakout>
- [4] <https://commons.esipfed.org/taxonomy/term/260>
- [5] <https://commons.esipfed.org/collaboration-area/data-preservation>
- [6] <https://commons.esipfed.org/node/344>
- [7] <https://commons.esipfed.org/taxonomy/term/246>
- [8] <https://commons.esipfed.org/taxonomy/term/266>
- [9] <mailto:Curt.Tilmes@nasa.gov>
- [10] <https://commons.esipfed.org/node/1403>
- [11] <https://commons.esipfed.org/taxonomy/term/242>

[12] <mailto:jeff.delabeaujardiere@noaa.gov>

[13] <https://commons.esipfed.org/node/323>

[14] <https://commons.esipfed.org/taxonomy/term/397>

[15] <https://commons.esipfed.org/node/557>

[16] <https://commons.esipfed.org/taxonomy/term/373>

[17] <mailto:ramdeen@email.unc.edu>