Earth Cube - Data Discovery Mining and Access [1]

Submitted by superadmin on Fri, 2012-06-29 19:19   Thursday, July 19, 2012 - 08:30 to 10:00
Event: Summer Meeting 2012 [2]
Session Type: Breakout [3]
Media/Video: session recording (streaming) [4]
session recording (download) [5]
Expertise Level: Intermediate [6]
Identifier: doi:10.7269/P3X63JT1
Collaboration Area: Discovery [7]
Semantic Web [8]
Abstract/Agenda:

The National Science Foundation, through its EarthCube initiative, is supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the Geosciences. This breakout will focus on Data Discovery Mining and Access related issues within EarthCube.

Notes:
(notes from Kelly Monteleone)

EarthCube – Data Discovery, mining and access – Rahul Ramachandran, Chaitan Baru, Tanu Malik

Rahul = discovery

Chaitan = mining

Tanu = access

- Agenda
    - EarthCube and Dataspace overview
    - Distilling the dataspace concept
    - Mapping between dataspace and earth science collaboratory
- EarthCube vision
    - Data knowledge management (keywords)
- An exaflood of data – can see data and then the tail is the derived data
- Data landscape
    - Paper by Heidron 2010 – the long tail of science – the total amount received by the investigator is 350K (80%)
    - Larger projects have well defined data lifecycles
    - The long take – unservered portion – have ad hoc processes
    - Do not have the IT capability to share and mange their data – $1 mill that haven't share their data
- EarthCube Data Activity
    - 11 expressions of interest
    - 60 data of workshops
    - Had webex and phone conversations about discovery, mining, and access – later combined
    - It was well attendant – invited state of the art project to talk about their work
    - Lots of great ideas
- Yes, we can
    - Outcome – lets put together a conception vision for EarthCube from the data perspective
    - Called the DataSpace – Conceptual infrastructure for distrusted, discoverable data access enabling "value added"

- Data-centric process – not built in a linear process – this is a process to get community engagement, to get process, to get to implementation team, so people can see how development is taking place
- Datapace
    - Integrates both big head and long tail of geosciences
        - Requirements, technology and governance
    - Already have existing services – want to leverage
    - Will be new things to be built – and adding things together
- What is DataSpace
    - Open to interpretation
    - Low-barrier of entry – people can join and start sharing
        - Q (Ken) – have you talked about this expressly – should you focus on higher barrier
            - Maybe EarthCube would be better served to focus on long tail
            - How can you allow them to participate
            - Maybe have cloud storage capability – with minimal metadata requirements
            - 1) did talk about details – not solutions but as problems
            - Not a binary, but a continuance
    - Uniform access
    - Need to be compliment with some set of capability
- Questions…
    - Are you an archive
    - Have not be decided
        - Not sure if it will be mandated to be an archive
        - Need to be discussed
    - Entry – implies people come in and do something (Ted Habermann)
        - Low barrier of entry – FGDC has low barrier – few fields – after you run a system like that for a while – have low quality – only get what is required
        - Leads to low quality, poorly documented data
        - There is no evidence that that leads to usable data and scientific reproducibility
        - Makes people feel good, but not good for science
    - People concerned about garbage on web
        - Better vs worse content
        - Metadata is incomplete – community concusses
        - This will sort its self out
    - May have shared the data… but may want to do more
    - It is daunting to spend a lot of time for scientists to input data – have simple form and then provide experts to help expand it … also means that the resource exist… know what is out there because some are working in isolation
    - Process of improvement needs to be explicit and funded
    - Chris MacMerman – how do you come up with a framework to make these decision
        - How can goverence develop a committee to define these terms – where are pit-falls – so a framework can be in place
    - Have not worked out details yet
        - Chris – have not enforced information to different groups… what want to support group
    - Need blended approach – need strategic info from governace
        - Have looked at Apache, Climate and forecast – see how not a monarchy
    - If there is such a space – funded/supported – another way to improve quality – have a human available to call on to ramp things up fast to hot have to face large issue
    - Structure already exists in NOAA and …. How to make that structure work
        - That is the intent
    - Everyone has an artifact in mind with EarthCube – looks like a datacenter (including discussion here) is that correct impression?
        - Not "a" datacenter – it is a distributed space – each place can have space and participate
        - DataSpcae (Earthcube) is a set of enties supporting a specific set of

- capabilities
  - Siri – email sent out with high level architecture view – will be in workshop report
    - It is a set of capabilities, not repository or fix set of requirements
    - Allows communities to expose their data as a repository
    - Infrastructure help expose to others
  - Have different "what is" because of the different groups
    - Different perspective – based on where group started – cross domain, data,
    - If can build this in your head… tell them
  - Dataspace – kept it unstructured – until they have more feedback of core capabilities
  - There have not been a lot of interaction between NSF, NOAA and NASA – datanet and other programs have funded other – how going to interact both within/between agencies
    - Had groups presenting – NASA – had some of NSF big projects
    - Have to include lessons learned from big projects
  - Past efforts – 40 weather cube… smaller version of earth cube?... hard to do this – better to not repeat – part of next-gen air transportation (Beth Haufer)
  - Chris (NOAA - – was in 40 weather cube – FAA and national weather service – involved in both programs – NOAA has been closely involved in earthCube – lead architect is very supported
    - Principle from 40 weather cube are in alignment
    - But they were much smaller – between National weather service and FAA and possibly with euro-control… much less broad than earthcube
    - Project still exist – getting RFPs out and getting funding for RFP, work with OGC and basing data exchange on their standards
  - Debra McGinnis – in dataspace – does that include scheme, vocab and ontologies or is that a separate space
    - There is a sementic group
    - If you look at the dataspace as usespace – there is role of semantic web technical, but not all/ only in dataspace – also in other parts (Chris NASA)
    - Need to get vocabulary – identify if vocab and ontology are part of dataspace
    - Not sure if inventory is going on right now – Denny did survey of semantic technology – large excel spreadsheet – if have tool for consideration fill in excel… propose that group create this spreadsheet again.. need to interoperate with other centers
      - Stated with catalogue of these tools, acess
  - Anne – at concept award meeting last week – working on last issue
  - Chaitan – NSF talks about EarthCube at scope of internet
    - Might be a good thing – create something like internet with more structure
    - Also – not enforcing standards – doesn't work – you are using standards, ex. Ascii
    - Using sql, which has standards – entry vs. full level
    - Dataspace – not a single data center – it is like the internet – if put somewhere – there are standards or prerequisite – may have notion (like cloud)
    - Metadata – there is attitude from archives about having clean data, don't want messy data, FGDC – people intimidated and don't put data up
      - Scientist know what they are doing – data will evolve over time – as used & reference – know if good
      - Metadata vs evolution – don't be afraid, need some minimum standards
  - Chris (Nasa) – separate data provide vs data consumer
    - Often both roles
    - If become involved in dataspace – get feedback mechanism – if improve metadata – then I can work with my data and other data… not a fire and forget for data provides
    - Want provides to come in and stay in dataspace… see improvement over time
  - Ken (NOAA) – need to be a member to comment and review – part of community to comment or review
    - Not sure NDGC if of same mind – continual improvement
    - As long as there minimum standards – that is where a dissagrement is

- - Whole dataspace/cube would accelerated faster if things started better off
    - Don't see it getting used, and commented on and improved
  - Chaitan
    - Lets trust these processes where uses defines quality
    - Because people give up because of standards
    - What is the new angle of attack we can try…
  - Ted Habermann – new angle of attack is important – but what doesn't work is collection of poorly document work
    - It is not just the data – it is data + documentation
    - Group places value on data – Jen yesterday – everything you need to reproduce the science – includes data and data
    - Give well documented datasets – will work
    - Creating another collection of poor documented data sets will not help anyone
  - Question is what is the minimum…. What is maximum (Ted)
  - One of the assumption – assume that people want to share their data – many researchers have no desire to share data – mandate doesn't mater – they will find a way not
    - Need to facilitate mindset change – need people wanting to come in the first place
    - Then data is the it will be shared effectively and used well
  - Low-barrier for entry – does not just mean metadata
    - Services to help ….
  - Dave (USGS) – low barrier of entry has nothing to do with quality – it is an ideal or goal for system – great thing at top – change mindset –
  - Steve young (EPA) – ebay – as datasharing platform – still require metadata… important part – it is based on ratings
    - For dataproviders, will welcome feedback – the social feedback mechanism would be an important aspect of moving people up in ranking
  - Not willing to share – you get a citation by including your data – gives an incentive
  - Quo – when we share data – we don't have to – we can share to small group of collaborative first before the world – talking about immediately sharing to world
    - Talked about data co-op that lets you deal with a small group first
    - Is there a set of requirements for governance policy (Debora) – NIH requires sharing of data & have policy in place to require who sees it or time of when
  - Debora Smith (remote sensing – producer) – want to share with people that will use it well – primary users that are funded to – is scientific users – general users are the ones that misuse/don't understand
    - Change will come when the funding agency requires metadata or repository
    - Don't have the training to learn all the metadata formats, etc. – because they have too much work with producing and working with –
    - In earthcube, want be both provider and user – learn more if able to use what handing off – incentives need to come from the top and include funding
  - Chaitan – have been looking into levels of access
    - When put data up in dataspace (exposing to services) – can be completely private – can share with students, next level = group, then world
    - That implies different – how data is shared among experts is different when others – what to get interpretation of work when to expert
    - Lidar – 80% want dem – 20% want the raw data and metadata
  - Ken Casey (NODC) – wondering if have time to get to agenda – will be hear about automated systems
    - Details are not in slides
  - Options on dataspace concept – what is it, will it work
  - Ken – don't know enough about automated system to have this conversation – like concept but not sure if it will get data to the right level quickly enough for someone to use
    - In documentation – get data in (cloud approach) – metadata – who and when put it – then search and other
    - If metadata search – then query use ask user for more information
    - Chaitan – he wrote this part – not what to cover in this conversation/session

- In visual phase – examples are not necessarily methods
- Slides – process and iteration
- Chaitan – talk about revised diagram -
  - Angle of attach is BBMA group approach – how would we incorporate ideas from other groups
  - Still work with long tail and big head
  - Have levels of compliance (full, intermediate, entry) – what they are is still discussed
    - Notion of going from incomplete to complete
  - Cube in middle = dataspace
    - Curation, discovery, acces, mining
    - If something enters this space – there should some information (DOI or other)
    - Metadata – long discussion – will take anything for metadata – as soon as you can provide whatever
    - There are extent data archive and systems out there – they hold the bulk of accessible data
      - Need to broker with existing databases – need to be compatible… they are not compatible amongst themselves
    -
  - Interoperate group – talk about compliance
    - Notion of different domain – what levels of compliance at the domain, but really talking about individual PI
    - Have to worry about how ready are people
    - Talked about need for people services – if tangible accepect of EarthCube that will be it… that entity of people (either long tail or big head) to help comply
- (Chris Jones) Compliance levels – interesting term – turns science community off – think about value levels – based on what services get base on what put in
  - Many scientist avoid large database, but use figshare because it is easy to entry
  - Participating at all is a good thing – moving up value chain
  - Chaitan – participation level
  - Other term – "readiness level" – but like value part
- Reference Architecture (this will be on the website by Friday)
  - Different communities – show how communites can share data
  - Earthcube allow diff communities to interact – ability for services to apply to different clouds – applies to readiness
  - The long tail and collaborative environment – represented different environments – if you through in the width of the pipes could equate the ability of federated system to participate
  - Agree to a common figure, insert concepts and then have a single way to overall accept earthcube
  - The ability to broker between different protocol, but can't go to the web because of requires, so might have to be installed locally – but maintained on web
  - To show value proposition in infrastructure – target the low hanging fruit – show connection with large repositories or small long tail participant with agreed to standards – broker these different elements – achieve a given scenario or benefit – can also be the curation services – need to leverage what is already built – do not want to reinvent metadata tools or other
  - Earthcube infastrcture that is on the rest of the clouds – does not require the separate domain infrastructure to do anything

(notes from Sarah Ramdeen)

Data cube session – Data Discovery, mining and access

Speaker - Rahul

Agenda – EarthCube and Dataspace overview (new workflow, should cover 15mins – gain feedback to include in the final NSF report), Distilling the Dataspace concept, Mapping between Dataspace and

Earth Science Collaboratory

EarthCube Vision – development of community guided cyberinsfrustructure – the key is data, information and knowledge.

There is raw and primary data as well as derived data which they are trying to manage. The current landscape – the long tail of science, (a paper from Heidorn 2010) is the smaller scale funded projects as opposed to the few large scale projects who have well defined data lifecycles to manage and share the data. The long tail has a limited budget and ad hoc processes – on the fringes because they do not have the IT capabilities to manage their projects.

EarthCube was formed because of 11 specific forms of intersrest from NSF and grouped in this community – they had 60 days of workshops on Discover, Access and Mining and by the end they merged it all to form a roadmap. It was a lot of time and and effort - invited state of the art projects to speak which generated many more ideas.

"Yes, we can!" Let's put together a perspective from the data on how to manage this – data centric with value added. So this would be a platform you could build off of, and that it is build with community input, engagement and push to get the implementation team has an open process so people can see how development is taking place.

Initial vision of the dataspace – intergrating the big head and the long tail. Requirements from our community, the right technology, and finally the governance. The key feature is there are a lot of existing services you do not want to throw away, but leverage these and create new options.

What is DataSpace? Provides a low barrier for entry to sharing data and uniform access, enable people to join as long as they meet minim compliant capabilities.

Question – low barrier of entry – easy to say, but ambitious goals. Did they talk about if these are at odds with each other? Or should we focus on higher barriers and only bring in stuff that will really make a difference? Answer- that is a valid point, one of the things this community felt was that this should focus on the long tail which has been neglected for a long time, how do you allow them to participate? Maybe cloud storage capabilities, so that if you are a lone scholar you can drop your data in to the storage with minimal metadata. We tried to widen this at this time. Additional comment – one we did talk about issues with solutions not problems. With domain scientists who were part of that long tail and brought us back to reality. It is not binary but more of a slope, of readiness levels. Part of what role these larger groups can play to bring the smaller projects up the slope. Providing data management tools for this long tail.

Follow-up question – more curious about phrases – "low barrier of entry" was this really addressed, or was it just tacked on? Additional comment – preconceptual stage, brainstorming on what EarthCube might be. Asked are you asking EarthCube to be a long term archive? Not sure that is an actual goal? Rahul – that is a good question – there are places where you can upload data to be

managed in NSF, not sure if this will be a manadated thing, btu these are important questions we need to have a discussion on. Ted – the word entry implies that people join and then do something. The low barrier – all the evidence I have seen, there might be a few requirements, but after you have run a system for a while, you end up with low quality, because all you get is what is required. So this idea of low barrier of entry as a good thing will lead to low quality, and it does not suggest it allows for reuse, does not support science across the community. Additional comment – I see this and I think about the web, and how it is a low barrier to create content, but there developed an eco-system for sorting better content from worst content. And there is a community consensus as to what is quality and what is not.

Rahul – share the data, but also have incentives for doing better work. Additional comment – it is hard to get people to come back to things. It is very daunting to spend time with this, so having minimal forms, to get an inventory of what is out there, and later specialists can do the more in-depth part of the process, you can follow up once you know a resource even exists! Often times you don't know what it exists on that long tail, a very minimal introduction. Ted – that process needs to be explicit.

Chris – how do you come up with a framework for making decisions? How can the governance come up with a frame work to avoid pitfalls that people have mentioned?

Rahul – we thought we would interact with you all, and you would provide the larger governance, bound our scope. But the details, we have never gone to that level. Chris – that is one of questions. We did not want to enforce to specific groups, but have a framework that will help the communities. Instead of being a top down approach. Have looked at different communities where there is not a monarchy dictating how things should be done.

Ann – if there is such a space, assuming there is some sort of financial support, another way to improve quality, would be to have a human, live expertise that can help you ramp things up fast instead of looking at huge barriers you have to meet.

Ted – the structure at Noaa and Nasa already exists, and NSF should look at that structure instead of inventing another structure for data upload. How to make that structure.

Rahul – leverage with DataSpace.

Jim – It seems like everyone has an artifact in mind for DataSpace, a place where you put stuff, a data center. Is that a correct impression? Because he can think about EarthCube without thinking of a place to store stuff.

Rahul – you can take part without having to, having your own server space. But there is a discussion on what is the difference with EarthCube and DataSpace, DataSpace is more data focused.

Commenter – Concept PIs met last week and made a high level architecture of EarthCube and it has

been mailed out to people. It will be in the workshop report. Earth cube is the set of capabilities, not a repository, not a fix set of standards, it allows each community to expose there data and services in a way that best suits their communities, and to foster discovery with standards and models.

Additional commenter – Different groups have different visions and that will be presented today. But what you will hear are different prospective from where each group started. It is sort of something that everyone is trying to touch and get a feel for it.

Rahul, what should be the core and start from there.

Commenter – I was wondering, how much people have thought about NSF prior activities. Like Datanet program that have funded infrastructure. How do you interact with these different agencies?

Rahul – we had presentations from some of the big NSF groups, maybe not all. But we have to go back and take a lesson learned from these projects, what you can learn from it.

Beth H. – there have been some efforts in the past, and it might not be a bad idea to check in with those folks and find out what went wrong. Not repeating mistakes would 4Dweathercube is one example, and the DOD has been doing some similar kind of things.

Commenter – (Colorado state working with NOAA) 4Dweathercube has similar principles, but the scope of it is much smaller. It is for the exchange of data between the weather service and the FAA. The project still exists, but needs RFPs and funding for those RFPs. Basing this on OGC standards.

Deborah – in the data space, does that include skemas ontologies etc, or is that a separate space? But it is needed. Beth asked if there is a semantic group.

Chris L, NASA – the DataSpace has a clear role to underpin the functionality, but all the needs of the semantic web of EarthCube are not needed for DataSpace

Deborah – hopefully we are not creating new vocabularies, unless really needed. There have also been discussions on what language we use to describe things. So whether ontologies are part of the conversation. Also is an inventory going on? Some of you, Michael Deny – inventory of sematic technologies, asked for people to fill in a list of tools about 10 years ago. Asked that this group does something similar. It might be too much to do the inventory, but lets send out something to a mailing list and ask people to contribute.

Rahul – we started a catalog.

Ann – at the concept award meeting, they are working on exactly that.

Chintar – (Remote attendee) wanted to comment on some of the things – the discussion has been fantastic and can be used for our own roadmap. Those of us who have attended the meetings are still figuring out what exactly EarthCube is, however the vision is quite large. Like the scope and feel of the internet. This might be a good thing, given the opportunity to create something with more structure but also it tells you something else, but not enforcing standards does not work. Every time you use a computer you are using standards. There will be standards. We discussed borrowing standards from SQL – it is a standard, three levels. Entry, intermediate and full blown. We are considering something like that, as Chris mentioned entering the data space then being pulled along to something more full blown. DataSpace is not a single repository, it is more like the internet and there might be notations of assistances, like DOI's things that you need to have as prerequisites, but it is not a single place. People who are on data archives have ideas about clean data and metadata, but we need to be careful where people do not get intimidated and not put stuff up. Put stuff up and the quality will evolve over time. The community will say that it is a good data set or not. Metadata set is left to evolution. Others have used it and it has good metadata and it will get reused. Can't build a system with no standards.

Chris – attack the assumption that we separate data providers from data consumers. In the long tail, some of those people are both, providers and users. And get them in the DataSpace – that you get feedback mechanism going, where if I make my data better in this respect, then I can work with my data as well as other datasets. This is not a fire and forget, we want data providers to engage in the data space.

Ken – this sounds like this is an angies list – where you have to be a member to comment and review, only members of the community can evaluate quality. Going to the phone comments, we definitely in the NOAA data centers have an idea of continual improvement with the metadata, if there is a disagreement, it is at the level of what is the minimum level of standards. The whole thing would be more useful, accelerate faster if the stuff started off better off. And if you really start with limited stuff, how will people be able to find it? Using tools to extract data, that are robust, that would be useful, but these tools need to be developed. And if things are going to just go in to the system with just time and name, that is too limiting.

Chintar – the moment you say standards, what is standards becomes the question. Let's facilitate sharing of data and begin down the path where trust, and usage create the quality. Data just sitting on computers and not know how to preserve. So what is the angle of attack?

Ted – I agree that a new angle of attack is important, and I would suggest that the angle of poorly documented data sets does not work. The value, of a community of shared data, that can use it effectively, the data plus the document is important. The value in this group seems to focus on the data, but it is really the documentation. If EarthCube wants to take a leadership position – give us well documented data sets and we will help you create them so they can be trusted and used. Another collection of poorly documented datasets is not going to help anyone.

Rahul – the question is what is the minimum? Ted - actually the maximum! Additional commenter –

there are people here assuming that people want to share their data. Even with a mandate, there are people who will come up with a reason not sharing data. SO want to facilitate a mindset change. Cant build it and assume people will come, you have to want them to come. Ted - Where data is shared and used well.

Chris – not a minimum set of metadata, but includes services that help manage metadata extract metadata and write metadata.

Dave – low barrier of entry does not have to do with quality but goal with the system you are building. Disagrees with Ted. And says this encourages changing this mindset.

Steve Young – analogy of Ebay – there is a low barrier, but minimum metadata, with lots of bad stuff being sold, but ratings of sellers which can allow judgments. This social aspect with ready feedback, some of those data providers would welcome feedback about what the problems are. That social feedback would help move people up the ramp. Get better at what they are doing, and if they do not get better, we might not want them in the shared space.

Rahul – not wanting to share – you are getting an incentive, a data credit/citation.

Quo – when we share data, we do not have to share to the world all at once, can have levels – sharing with collaborators first. If I have a collection of data that I can share with collaborators first. Not immediately. That might not be the case.

Rahul – you could have your own rules, the infrastructure should support that capability.

Deborah – is there a set of policies for those plans? IF not make one, and models – the NEH requires sharing data, but also has policies on who you decide to share with and the time limits of when the data will be shared. We need to have something like that.

Deborah S. – we're one of the data producers – we do want to share with people who want to use it well, and we have different kinds of users – primary users, scientists who will reuse. The general user tends to misuse data, and we are funded to help that kind of user. We are given limited money to do the documentation and help people use these things. The change will happen when the agencies say you have to do things, because they will have to fund the work. It is not that they do not want to, but they are producing the data and do not have the time to do the work. Want us to be data providers and users, we would learn a lot more about what we are handing off if we could learn from using. And we need other data, and if we could see how it fit, it would improve that way. There has to be money to help us do it.

Chintar – As Rahul said, we are looking in to levels of access – you can put data up and make it completely private. The next level up could be selected access and finally world. Cloud data shared

among experts vs sub disciplines are two different conversations. Different levels of expertise is needed for different levels of access to the data. Some people are experts and want to manipulate and others just want to use it as a direct product. Also this goes to provenance – people just want raw data but the 80% are happy with the data set as long as they can get the background of how it was created. And if they do have a question, they can dig in farther and figure out how it was derived.

Ken – data mining and data discovery – will we hear about those automated systems?

Rahul – do not have the details, but it is open. What we would like is opinions on DataSpace – is it worth perusing and what is the perspective – would be valuable be for mining.

Ken – I do not know enough about these automated systems, so start from a different approach, how do we get the data so it can be found and reused. Were you going to go over these tools so I can form an opinion on these tools. Interesting concepts but we close enough to create these tools? Something that will pull out metadata and prompt people when it is not enough.

Chintar – still in development stages, and we can talk about it offline, we are very much in the scoping and visioning stages, and some of these things we are using as examples might not be the exact systems we want to implement. What we build will have to be compatible to these things.

Rahul – instead of going over the slides, does Chintar want to go over the diagrams for DataSpace? The device diagram. This will be updated in the Roadmap.

Chintar – progression and our own thinking about it. There are lots of activity in this group. Still focusing on the long tail and the big head. And funding is a good way to distinguish these groups, and level of compliance. (speaking about the diagram) the cube in the middle will be DataSpace. It includes curation of the data – (Management, discovery, access etc.) and then it will include DOI's and other identifiers, and then Metadata. There are a bunch of data archives that hold the bulk of the accessible data, and some of our concepts represent these groups. Huge well-funded operations. There are existing things, and these might not be compatible with this and they are not compatible amongst themselves, so brokering will be important. There is also a discussion of interoperability. And compliance – how ready am I to comply? What level are these different domains? Interoperability either at the long tail or the big head is what will be important.

Matt J. – compliance levels, interesting term. That can turn off science communities – lacks value. Talk about value levels and what services you may provide if scientists met certain levels. There are a lot of scientists who avoid large groups because they see barriers, but do join these smaller groups that let them participate without lots of requirements. So we have to get away from the concept of compliance.

Chintar – good points.

Chris – I like that, readiness levels but value levels are more positive.

Rahul – reference diagram.

Commenter – *described the diagram,* and talked about the different aspects that allows the community to interact and how EarthCube fits in those services. In those different clouds based on readiness. The long tail has other resources that can be connected in to the EarthCube system, the width of the pipes, these different federated organizations to interact depending on their readiness. A high level diagram. This should be published this Friday on the EarthCube website. This is also posted on the link site as a photo. It is a reference architecture, but more of a conceptual architecture. High performance computing environments because of speed of access to the systems, there might be a set of protocols since some of these would not be able to work over the web and might need to be locally stored. But maintained like the web itself for all of the participants. SO they can access it through the infrastructure. Exposes the data for discovery. To show the value of the proposition, it is obvious to target the low hanging fruit, but do not have to reinvent tools, preservation services, there are already projects funded to do some of these things.

Rahul – there is infrastructure, but does not require separate domain organizations to do or install anything. We are still having our weekly webex, if people could join us and provide different perspectives. Feedback would be really useful. Will send an email to tell everyone where to look.

Actions:
email will be sent to participants with more information to continue the discussion

Session Leads:

Name: Rahul Ramachandran [9]
Organization(s): University of Alabama in Huntsville [10]
Email: rahul.ramachandran@uah.edu [11]

Notes takers:

Name: Kelly Monteleone [12]
Organization(s): University of New Mexico [13]
Email: krbm@unm.edu [14]

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: This breakout will focus on Data Discovery Mining and Access related issues within #EarthCube.
Keywords: EarthCube [18]

  Source URL: https://commons.esipfed.org/node/461

Links
[1] https://commons.esipfed.org/node/461
[2] https://commons.esipfed.org/event/summer-meeting-2012
[3] https://commons.esipfed.org/session-type/breakout
[4] https://esipfed.webex.com/esipfed/ldr.php
[5] https://esipfed.webex.com/esipfed/lsr.php

[6] https://commons.esipfed.org/taxonomy/term/261
[7] https://commons.esipfed.org/collaboration-area/discovery
[8] https://commons.esipfed.org/collaboration-area/semantic-web
[9] https://commons.esipfed.org/node/207
[10] https://commons.esipfed.org/taxonomy/term/186
[11] mailto:rahul.ramachandran@uah.edu
[12] https://commons.esipfed.org/node/430
[13] https://commons.esipfed.org/taxonomy/term/297
[14] mailto:krbm@unm.edu
[15] https://commons.esipfed.org/node/557
[16] https://commons.esipfed.org/taxonomy/term/373
[17] mailto:ramdeen@email.unc.edu
[18] https://commons.esipfed.org/taxonomy/term/314