

CDI Monthly Meeting 20160113

CDI Monthly Meeting - January 13, 2016

The Community for Data Integration (CDI) meetings are held the 2nd Wednesday of each month from 11:00 a.m. to 12:30 p.m. Eastern Time.

Audio:

USGS/DOI Dial In Number: (703) 648-4848 (for USGS and DOI offices)

Toll Free Dial In Number: (855) 547-8255 (for other offices and telecommute locations)

Conference Code: 47919# (same for both numbers)

Webex Recording

Webex recordings are available to CDI Members. Please login to view the recording. If you would like to become a member of CDI, please email cdi@usgs.gov.

Agenda (in Eastern time)

11:00a Welcome - Tim Quinn - Chief, Office of Enterprise Information

Opening Slides: Slides are available to CDI Members. Please login to view the slides. If you would like to become a member of CDI, please email cdi@usgs.gov.

11:10a CDI Land Cover Trends Photo Project - Chris Soulard, U.S. Geological Survey

Presentation: Slides are available to CDI Members. Please login to download the slides. If you would like to become a member of CDI, please email cdi@usgs.gov.

Abstract:

Chris Soulard started with the U.S. Geological Survey in 2002 and has been with the Western Geographic Science Center in Menlo Park since 2005. Most of his career and 30+ publications have focused on multi-temporal change analyses using Landsat imagery and aerial photography, most notably the USGS Land Cover Trends Project, which inspired the 2015 CDI project aimed at getting a decade's worth of Land Cover Trends field photos loaded online. Today Chris will discuss the original field photo collection and how a small team has converted slide film negatives into a digital, geo-tagged format available to the public online via the [Land Cover Trends Field Photo Map](#) and Earth Explorer.

Questions and Answers

Roland Viger: The land cover scheme that you are classifying into, you said that it is similar to NLCD but a little different? Did I get that right?

Chris Soulard: Yes, NLCD has the Anderson Level II classification scheme. We have a modified Anderson Level I. So they are fairly analogous to one another. They will have a cultivated crop class and pasture class. We boil those down into one agricultural class for the Land Cover Trends Project.

Roland: So would you say that the correspondence of classes, even if it is nested and generalized sometimes is pretty straight ahead for you guys for going back and forth?

Chris: Yes, we actually use NLCD to start our whole classification effort. We have a few definitional tweaks, but yes, they essentially nest. With the exception being their automated algorithm sometimes makes a different classification call compared to our manual classification.

Roland: Yeah, I am mostly interested in the classification scheme rather than how you make the assignments, which is a whole different ball of wax. I guess the reason I am fishing around is that I am aware of several different things going on where folks like NAWQA (National Water Quality Assessment) folks, as well as NLCD and Gap and Landfire are all kind of messing around with different land cover classifications. I know that Gap, NLCD, and Landfire all have an MOU to try and figure out, even if they are using different schemes, how to make them relate to one another. So say some new user can move from one data set to another and communication information back and forth. So your scheme caught my attention there.

Chris: Yes, to crosswalk our classes to NLCD or Landfire, because Landfire is largely using NLCD, ours is very easy to crosswalk to NLCD. There's a whole body of literature on that crosswalk between classifications. We term it harmonization for our intents and purposes, but yes crosswalking, ontology matching, are all different ways of looking at it. Actually, that ties into a large CDI issue of having common keywords within a community. At the meeting last year, there was a great play about ordering food. Over the course of the meal there was a question of the different definitions and how a machine may interpret those.

Roland: I would love to capture that in some of our wiki pages. So that you guys as individual programs do what you do and take care of your needs, but maybe facilitating recording those in a centralized place so that someone who is choosing between different data products can determine which ones right for me and what does it mean compared to this other thing that I happen to know. That might be nice. It is good to hear that you guys have published on the harmonization and stuff.

Sky Bristol: I was taking a look at some of the photos that you had there through the Earth Explorer and I was interested to find out if you had thought about incorporating some of the value-added metadata as EXIF tags in the actual photo data themselves, in the jpegs, in terms of the location information and everything else that you are picking up through the processing.

Chris: I don't know if I fully understand your question, so I'll try to answer it, but please feel free to reword it if I don't address it properly. The EXIF tags are kind of built from scratch on our end. To some degree, when we moved to digital photos, we got some more EXIF information that came along with the digital collection. Prior to that, when we were scanning, the EXIF tags were essentially blank, so we are building those out from scratch. Our viewer and Earth Explorer will allow you to search those EXIF tags using the keywords, but the coordinate values are what display the images geographical context. The only other EXIF tag that we have recently sourced is...originally the plan was to boil down the collection date to a month and year instead of the specific date that it was collected. Not only could we not go back and find the exact date that it was collected, but we also realized from a PII perspective that by just going to the month that it bought us some flexibility for any privacy concerns. So if we were doing new collects, digital collects, we will have much more EXIF information that comes free and ready with those collections that we could display, similar to the adopt a pixel project with respect to the orientation of the photo, that we previously didn't have.

Sky: Yeah, I was thinking more about injecting information into the EXIF tags for your distribution version of the photos themselves. So that as those things get downloaded and we pick them up for something else that we can pull the information out of the files as opposed to having to go to a separate metadata source for that.

Chris: Oh you mean add something like the NLCD class that corresponds to the point or a landfire disturbance or something like that?

Sky: Right, I mean essentially what you are able to interpret, if you could send those along with the files themselves. That seems like it could be tremendously useful.

Chris: That's a great idea.

Rex Sanders: I think one of the values of this database is that all of those photos are in the public domain and for certain people like journalists who are desperate to find interesting photos of things and places on a short deadline, this could be a goldmine.

Chris: Thanks, Rex. Yes, when we put out the press release, we tried to identify journalists that typically write about photography. We got some traction from that middle ground...there are a few GPS publications that picked up our press release. I think that the people who are trying to write a piece in Sunset on the Sierra Nevada, they can go to our collection and just get 500+ photos on the Sierra Nevada. That was exactly what our goal was...to reach average users, press, to research communities, we really wanted to run the gauntlet there.

11:40a CDI Geographic Searches Project - Rex Sanders, U.S. Geological Survey

Presentation: Slides are available to CDI Members. Please login to download the slides. If you would like to become a member of CDI, please email cdi@usgs.gov.

Abstract:

*Today Rex will give us an update on the Geographic Searches for USGS Publications Project, which CDI sponsored in FY 2014. Now anyone can search for **new** USGS publications, and a growing collection of **older** publications, by drawing map polygons on the Pubs Warehouse web site.*

Questions and Answers

Andrea Toran: This is Andrea Toran from the Woods Hole Coastal & Marine Science Center. We have over one hundred footprints that we have created for some of our older publications and they are just sitting on a harddrive at this point in time. And they are ready to go, so it would be great if we had a specific contact person, I don't know if it is Jim or Kathy, so we could come up with a plan to get some of our footprints posted.

Jim Krefit: Those are the ones that are up in ScienceBase, now? The ones that are up in ScienceBase will be there tonight. We had some hiccups and personnel issues. Various people who were supposed to be available got pulled into other projects. But those should be up there as of today.

Andrea: We have others that we are just sitting on at Woods Hole. I don't know what makes sense for you as far as how to batch them.

Jim: If you can just continue pushing those up to the ScienceBase folders we can get those up. That will be a nightly build.

Chris Garrity: This is Chris Garrity in Reston, Virginia. I have a couple of questions. Is this just Leaflet draw out of the box or did you guys extend it and if so, do you have the code available for people to contribute to or use on their own?

Jim: Yes, this is leaflet draw. It is pretty much out of the box. We did do a little customization to generate WKT to use to push up to the search view since the search view actually uses WKT. We We also incorporated a tool called Wicket. I am digging back into last summer, I'm sorry. Everything is open source and it is up on GitHub. If you send me an email I can send you that. It is just at https://github.com/USGS-CIDA/PubsWarehouse_UI.

Chris G.: How many of those footprints are relatively detailed. I know you said 30,000 have footprints but are those just big boxes or are many of those kind of detailed footprints.

Jim: I would say roughly, at this point, what would you say Kathy...3,000 or so are probably more detailed?

Kathy Wesenberg: I think there are more than that, Jim.

Jim: Yeah, that's right because the National Wildlife Health Center has done a lot of that and for historic pubs too, so yeah maybe it's more than that.

Kathy: Everytime we touch a pub, we add a footprint.

Rex: Well, that's a good point. If you are going through your older pubs at a science center and you see that one of those has a big bounding box on it and you'd rather have a polygon that makes it more specific, that's a part of the workflow that we haven't quite worked out yet.

Chris G.: What if I just have a point? Does it handle points or does it just handle polygon intersections?

Jim: I theory I think you should be able to add whatever you want to add to the geospatial index. It's really just sending off a GeoJSON object that can have pretty much anything in it. I think the footprinter tool is just primarily focused on polygons, but we should be able to accommodate any of the standard geospatial features.

Rex: That actually brings up a good point. Woods Hole Science Center is actually doing this. For most of their publications, they are uploading ESRI Shapefiles. If you have a workflow that is much more used to dealing with shapefiles, rather than going on to a website and pointing and clicking, we are able to take those into our system.

Roland: My understanding is that what goes into PubWarehouse in the first place is stuff that goes through SPN. Things like open files for data sets is probably the most common container there. What I am wondering about is...there are workflows being built up for publishing datasets via ScienceBase, but they don't go through SPN and don't show up in PubsWarehouse. I'm guessing there have been discussions about that and reasons for why things have gone that way. I'm wondering if you can talk to that a little.

Jim: Others on the call...other CSAS folks can probably chime in here too. Basically, we're trying to figure out exactly how the data release and publication interaction is going to be happening. It's a discussion that is happening across OSQI and CSASL. That is an ongoing discussion. Exactly how that's all going to be working out is still being hammered out, but basically, we're working out a system where if a data release is linked to a publication or a publication is linked to a data release, we'll have that semantic linkage built into both systems eventually.

Alan Allwardt: There is a category in the PubsWarehouse called USGS data website, which does not go through SPN and they are in there.

Jim: Yeah, at this point...those were sort of a stop gap and we haven't really been adding those lately. And we're working out a way for the Science Data Catalog and other USGS data focus tools to take advantage of those things.

Roland: If a data set gets a DOI as part of its non-SPN data release process, that would be a great place to start rather than a website, which can be a bunch of data sets on one page.

Alan Allwardt: That is the process that we have been following for data releases that go through ScienceBase. We don't approve those for dissemination until they have a DOI. And correct me if I'm wrong, Jim, isn't it the case that you don't necessarily get stuff from SPN. You get stuff from IPDS.

Jim: Right, so SPN is for official USGS series, but we also catalog books, and articles and all of the other things that have gone through IPDS.

Alan: And data releases that have gone through IPDS?

Jim: Well, I'm hedging on that a little bit. We were pulling data releases and we actually have them in the backend, but we haven't actually been releasing new ones lately because we're trying to figure out, working with ScienceBase and the Science Data Catalog, where the long term home for these pages are going to be.

Kathy: Beyond getting things from IPDS, we only put in things that have Bureau approval. So you're not going to see abstracts and some conference papers in there.

Roland: Well, that's interesting in that a lot of ScienceBase data releases only get science center level approval.

Alan: But that's delegated Bureau approval. The Science Center Director is giving delegated Bureau approval.

12:10p New Communication Strategy for CDI - Leslie Hsu

You might have noticed the announcement for the monthly meeting looked different this month. Instead of sending a direct email reminder, we shared the Confluence page where we have the Monthly Meeting agenda. Why are we trying this new strategy? Many of us have experienced an overload of emails and it is easy to lose this information in our inbox. We would also like to promote the CDI wiki as a way to find information that has come through your inbox previously. So this email that you received was pushed from the Confluence wiki.

Hopefully, most of you have come to the wiki before. On the [homepage](#), there are announcements, such as the sign up for the Software and Data Carpentry workshops that were mentioned earlier on this call. If you want to find that link again, it will be on the homepage. So what we are doing now is using the wiki to push announcements out to the community. If you look on the far left hand side of the wiki, at the very bottom, you will see a link to the [CDI Community Forum](#). Anyone, whether you are part of CDI or not, can view this forum and see the discussions and the announcements that have been taking place here. CDI members can also create new topics and share them with the community or make comments on existing announcements and discussions. Members can choose different levels of participation. You can "watch" the forum to automatically receive updates to any of the topics or you can simply "watch" individual topics or wait for them to be pushed to your email inbox. If anyone has any comments or questions about this new method please let us know at cdi@usgs.gov.

12:15p Working Group Reports

- **Communications** - JC Nelson

New working group. Communication working group has come out of what used to be the Science Data Coordinators Network. SDCN has undergone some changes and we thought it would be worthwhile to bring this under the CDI and open things up. Participation used to be restricted to delegated individuals; however, we want to open it up to any volunteers. We are going to focus on communication from CDI to Science Centers and from Science Centers back to CDI. We are going to set up a spot on the wiki space. Membership is open to everybody. If you have ideas or an interest in Communication let JC (jcnelson@usgs.gov) know. Until we get the wiki going feel free to email JC Nelson or Marcia McNiff (mmcniff@usgs.gov) with suggestions or questions.

- **Citizen Science** - Dave Govoni

Sophia has been doing most of the coordination for this working group. Sophia has been attending meetings regularly with OSTP. Some talk going on among Federal entities of supporting a citizen science day in the spring of this year. Some coordination with the 100-year anniversary of the parks service. The thought is to sponsor or support activities at refuges or parks in conjunction with federal entities or with local groups: (e.g. BioBlitzes). Also using that event to advertise ongoing citizen science and crowdsourcing activities.

- **Data Management** - Michelle Chang

A lot of really great things came up at Monday's Data Management Working Group meeting:

1. John Faundeen announced the International Data Week, which will take place in September 2016. He is looking for suggestions for CDI session proposals for the conference.
2. Lisa Zolly is looking for feedback about use of the CSDGM Shoreline Profile and Remote Sensing Extension from USGS employees to help guide the development of the Online Metadata Editor.
3. JC Nelson put together a draft Data Sharing Policy for the USGS and is looking for feedback.
4. Steve Tessler is planning to develop out the Process and Analyze sections of the Data Management Website. If you or anyone you know of would like to contribute to this work, get in touch with Steve.
5. Ruth Duerr will be presenting next week at the Data Management Webinar Series on data citation.

For more information on any of these topics, visit the [Data Management Working Group meeting agenda](#) or contact Michelle Chang (mchang@usgs.gov), Viv Hutchison (vhutchison@usgs.gov), or Heather Henkel (hhenkel@usgs.gov).

- **Earth-Science Themes** - Roland Viger

Earth Science Themes working group is playing around on the wiki. Anyone is welcome to come play too! We are about exploring questions such as, "What are the best methods for answering XYZ question?" or "What is the best data set to use for XYZ?" We have a number of adhoc focus groups (Elevation, Water, Soils, Land Cover, Oceans, Integration). Some are just stubs for now and others are more active. We have a lot of participation from people outside of the USGS, which is great! Roland would like to get some people involved in leading some of these focus group efforts. He has been doing this solo for a while and could use some help. Let him know if you are interested in getting more involved (rviger@usgs.gov)!

- **Semantic Web** - Fran Lightsom

Next Semantic Web meeting is Thursday, January 14 at 2pm ET. SWWG will continue working on their GeoSPARQL endpoint. They will also have a discussion about looking into what's going on with a new Research Data Alliance working group.

- **Tech Stack** - Rich Signell

Next Tech Stack meeting is Thursday, January 21 at 2pm ET. John Jediny from data.gov will talk about geonode, CKAN and pycsw.

- **Connected Devices** - Tim Kern and Lance Everette

Tim has scheduled a call for January 28 at 12pm ET. Send Tim Kern an email to be added to the email list. At the meeting they will discuss the process for getting mobile apps released through the iTunes store.

12:30p Adjourn

Attendees

Madison Langseth (Host, me)



Chris Soulard

Jim Kreft

CHris

John Bumgarner

Chris Rusanowski

John Faundeen

Chris Skinner

Josh Young

clara brown

Karen Adkins

Daniel Wieferich

Kathy Wesenberg

David Govoni

Kelly

Denver

Kevin McNinch

Ed Olexa

Leslie Hsu

Emily Fort

Marcia McNiff

Enrika

Mark Wiltermuth

Fran Lightsom

Megan Dethloff

Gita Urban-Mathieux

Mikki Johnson

Great Lakes Science Center

Natalie Latysh

Jayme

Paul Wiese

Jeff Runde

Raad

Jen Bayer

Rex Sanders

Jessica Hopple

Rich Signell

Rob Dollison

Roland Viger

Sophie Hou

Stephanie Galvan



Steve Aulenbach

Steve Tessler

Tim Kern

Tom Kalvelage

Tom Shoberg