

Deb Paul: (8/7/2014 12:36) Google Doc for Notes:https://docs.google.com/document/d/1W0MRrCvksi-DUWVz4_DAJx3VN72RnonpDazIepBwnIq/edit

Deb Paul: (12:51) Wiki:https://www.idigbio.org/wiki/index.php/Data_Management_Interest_Group

David King: (12:58) Sounds good here too

Sandra Brantley: (12:58) Yep, fine. Sandra

MK: (12:58) Hi Deb

Laura Abraczinskas: (12:58) Sounds fine, thanks

Christina Byrd: (12:58) Good

Janet Bala: (12:59) Good. what do I need to have to speak?

Alex Thompson: (12:59) sorry

Alex Thompson: (12:59) that was me

Elizabeth Sellers (USGS Reston, VA): (13:10) That's OK. For now. I'll just mainly listen.

Sandra Brantley: (13:12) The LTER programs have data managers with similar problems - they may be a good resource, particularly for metadata.

Sandra Brantley: (13:12) Yes, I have an LTER office here and can contact them.

YPM Susan: (13:12) please send link to google doc! Thanks!

YPM Susan: (13:13) got it!

Kevin Love: (13:13) https://docs.google.com/document/d/1W0MRrCvksi-DUWVz4_DAJx3VN72RnonpDazIepBwnIq/edit

Chuck Miller: (13:13) Technical side. Doesn't this require an abstraction layer between the various different endpoint data and the exchanged data?

Chuck Miller: (13:13) No working mike

Chuck Miller: (13:14) N squared problem.

Chuck Miller: (13:14) How do you deduplicate the UUIDs between different sources coining their own UUIDs?

Nina Sandlin: (13:15) Btw, Google Doc notes "Wow, this file is really popular! Some tools might be unavailable until the crowd clears.."

Brian Westra: (13:15) In some cases, taxonomic info may not be well-established, and will be modified later when a species is fully identified. So the provider may want to update record(s) later on too.

Joanna McCaffrey: (13:15) There is also the security issue of allowing outsiders access to your collections management application.

Nina Sandlin: (13:16) No worries, I don't need to type right now, just didn't want others to think it was "just them"

Jason Holmberg: (13:16) I noticed "Duplicate Detection listed in the Google Doc. Is there a standard around using UUIDs to support interchange? We just implemented those in Wildbook as our software interacts with other remote systems that may push or pull data.

Jason Holmberg: (13:16) Portland, OR

Jason Holmberg: (13:16) Wild Me non-profit

Jason Holmberg: (13:17) <http://www.wildme.org/wildbook>

Brian Westra: (13:17) How important is it to have detailed tracking of who made what change (provider, aggregator) and when? Would make the database more complex, but would probably be nice to know.

Chuck Miller: (13:17) How do you deduplicate the UUIDs between different sources coining their own UUIDs?

Brian Westra: (13:18) I saw an older discussion that was not resolved, about UUID for the specimen vs. for the record.

Alex Thompson: (13:19) Chuck: UUIDs are virtually guaranteed to be unique

Alex Thompson: (13:19) based purely on probability

Chuck Miller: (13:19) But UUID referring to what? It's the what that is being duplicated.

Brian Westra: (13:20) ARKs are a nice idea.

Chuck Miller: (13:20) It takes a resolving service to enable multiple sources using UUIDs to record the duplications.

Anne Thessen: (13:21) Its done

Deb Paul: (13:21) Google Doc for Notes:https://docs.google.com/document/d/1W0MRrCvksi-DUWVz4_DAJx3VN72RnonpDazIepBwnIq/edit

Elizabeth Sellers (USGS Reston, VA): (13:22) 'Information Models' also synonymous with 'standards'? e.g. DfW, ISO, FIPS etc..

Chuck Miller: (13:22) Debbie, Are you logging this chat here?

Gail Kampmeier: (13:22) Could someone make a shorter version of the Google Doc URL? I cannot click on it from my device

David Bloom: (13:23) <http://goo.gl/S7nzs3>

Gail Kampmeier: (13:23) Thanks!

Elizabeth Sellers (USGS Reston, VA): (13:23) Looks like Yes. Thanks.

Jason Holmberg: (13:23) not all models are standards

Elizabeth Sellers (USGS Reston, VA): (13:24) Indeed

Stan Blum: (13:25) Information models are more structured than term lists like DarwinCore. Many different models could be implemented from DarwinCore.

Brian Westra: (13:27) Can you provide notes on filtered push (sp?)

Alex Thompson: (13:27) <http://wiki.filteredpush.org/wiki/>

Brian Westra: (13:28) Thanks!

Larry: (13:29) I am working with the National Park Service and i have had botanists annotate the records using our online database. They have administrative rights and we now have the ability to track the annotations online but that means the actual specimen sheets do not get annotated unless someone goes to the location and annotates them.

Reed Beaman: (13:30) Not me.

Alex Thompson: (13:30) - Alex

Laura Brosskelle: (13:32) Larry, what program allows you to do that?

Chuck Miller: (13:32) I can't edit the Google Doc. Need to include TDWG in the data model list = Darwin Core, Semantic Darwin Core, Audubon Core, Plinian Core,

David Bloom: (13:33) <https://github.com/>

Nina Sandlin: (13:33) www.github.com

Brian Westra: (13:33) The vertnet/Github process sounds like a good way to provide notifications of issues/updates, etc. More about this online?

Laura Russell: (13:34) <http://www.vertnet.org/resources/issuetrackingguide.html>

Laura Russell: (13:35) <http://www.vertnet.org/resources/developers.html>

John Wiczorek: (13:35) <https://github.com/mvz-vertnet/mvz-bird/issues>

John Wiczorek: (13:35) An example for Birds

John Wiczorek: (13:35) Includes usage reports as well.

Laura Russell: (13:36) example of the mvz-birds usage report in html preview, http://htmlpreview.github.io/?https://github.com/mvz-vertnet/mvz-bird/blob/master/reports/MVZ-mvz_bird_2014_07_02.html

Chuck Miller: (13:37) Paul Morris, Harvard MCZ

Kevin Love: (13:38) <https://www.idigbio.org/biblioff?search=Filtered%20Push>

Larry: (13:39) Laura, currently we have our own database that we created in the libraries with the help of our IT folks. We are hoping to move to Symbiota and we will need that capability built into it.

Christina Byrd: (13:39) how does github connect to your site? do you need a programmers knowledge to link them?

Larry: (13:39) yes

Chuck Miller: (13:39) Will Specify 7/8 be in the cloud?

Jason Holmberg: (13:39) Filtered Push mp4 video is giving me a missing record page on IDigBio

Jason Holmberg: (13:40) <https://www.idigbio.org/videos/idigbio-summit-FilteredPush-Macklin-30NOV2011.mp4...missing..>

Christina Byrd: (13:41) oh..never mind. the answer is in the doc

Alex Thompson: (13:42) <http://symbiota.org/tiki/tiki-index.php>

Sandra Brantley: (13:42) Yes, many in the SW arthropod group use Symbiota both as institutional databases and as a portal for group databases, checklists and images.

Larry: (13:42) Most of the park service herbariums are actually part of the museum so a curator is managing the collections. Most of them do not have a science background so it does complicate things.

Chuck Miller: (13:43) Didn't KU get another Specify enhancement grant from NSF?

Elizabeth Sellers (USGS Reston, VA): (13:43) Depends on your definition of "best"

Elizabeth Sellers (USGS Reston, VA): (13:43) It's almost always 'buyer beware' regardless of data source

Sandra Brantley: (13:43) With arthropods, maybe not, if we don't have the expertise in house to make an ID. In that case putting a record/image

Brian Westra: (13:44) Sorry, can you repeat the question?

Sandra Brantley: (13:44) ...record/image out there may get us help and a better record.

Mariel Campbell: (13:44) We have the same issue with parasites - some of our data are linked to host data at other collections which may or may not be more accurate

Mare Nazaire: (13:45) It might be achievable for some institutions that have the capacity to make their data of good and accurate quality.

John Wiczorek: (13:45) This is an overwhelming number of people and interests. What will be the scope of this Interest Group?

Stan Blum: (13:45) One reason the provider/owner could/should be an authoritative source for data is that the provider can get annotations from many "users".

Chuck Miller: (13:45) Best is a difficult term. In this context does "best" mean "nothing better available so far"?

Brian Westra: (13:45) It may be a collaboration; field collection and initial data entry by one group, but then getting species identification by another institution. Is this uncommon?

Alex Thompson: (13:46) (Dan Stoner) georeferenced data by 3rd party also is "better" potentially than original source.

Arturo H. Arico: (13:46) sound gone

Mare Nazaire: (13:46) But for others who lack the resources, staff or time, this could be more difficult to achieve.

Peter Obyski: (13:46) Source data is likely to be the most precise (relative to the original collection information), but not necessarily the most accurate. So it depends on what "best" means.

Sandra Brantley: (13:46) Yes, Brian, this kind of collaboration is common.

Jean: (13:46) Regardless of how the data is improved I think the museum should always have that data and be able to provide it.

Reed Beaman: (13:46) Can you post the powerpoint Deb?

David Bloom: (13:47) YES

Liath - Univ. of TX: (13:47) @Chuck Miller, I hear tell that Specify 7 will be in the cloud

David Bloom: (13:47) correct, Liath

Brian Westra: (13:48) Could create work teams to work on specific issues and share with the larger group. This was very informative.

Ron Eng: (13:48) Source data should be considered to be the most precise. However, we should also be sure to view it with a "critical eye."

Chuck Miller: (13:48) There has been/is a lot of work going on with the collection data management question in Europe. I see Arturo here. How much interaction between this group and the EU groups?

Alex Thompson: (13:48) Might need to spin off one or more working groups to accomplish tasks

David Bloom: (13:49) is this a working group or an interest group

David Bloom: (13:49) different foci

Sarah Timm: (13:49) Under the question to clarify the scope of this group, could you comment on which collections you are focusing on? I realize museums are being discussed but what about university departments? I feel they are often forgotten and yet there are many important specimens in those collections and these issues do apply there as well.

Alex Thompson: (13:49) Within the scope of idigbio
Sandra Brantley: (13:49) Yes, a great idea - the collection managers are in the middle of taxonomy, curation, and data - we can use help from many sources.
Arturo H. Ariño: (13:49) Chuck: I'm not aware of specific inter-continental collaborations. I believe most everyone is on his/her own.
Nina Sandlin: (13:50) More on "best," sometime even simple things like counting the number of specimens in a loan, or decoding seemingly illegible handwriting on a label
Nina Sandlin: (13:51) I meant, which could happen in a transaction like a loan, but would not necessarily be done in databasing by the institution of provenance
John Wieczorek: (13:51) Also, TDWG has a charter almost ready for submission for a Biodiversity Data Quality Interest Group.
Laura Russell: (13:51) I believe initial deliveries of Specify 7 will be mixed. The will be a web-based tool, but it will not have all functionality and using the desktop tool (6.5) will be required. Features will continue to be added to the web-based tool. I've heard Andy mention that Specify may be able to begin providing hosting capabilities for data with Specify 7.
David King: (13:52) Inter-continental collaboration is developing through the Research Data Alliance: <https://rd-alliance.org/>
Brian Westra: (13:52) CoDATA, ISB/Biocuration?
John Wieczorek: (13:52) The Charter is meant to go to the Executive Committee on the 11th.
Chuck Miller: (13:52) Indeed, what does management mean?
Arturo H. Ariño: (13:52) DQ is perhaps a little more focused. Could be thought of as a subset of DM? Allan Koch is convening within TDWG I believe.
Mariel Campbell: (13:53) Our parasite collection has been using the Arctos online database to link parasite vouchers to host vouchers in other collections - we are working to be able to create relationships with institutions outside of Arctos with host vouchers to our specimens; this is a priority for us
Sarah Timm: (13:53) Thanks Deb, I also wonder how many people here are aware of Earth Cube. They are having some similar discussions about actions to take to solve some of these problems. See link here to most current All Hands Meeting summary: <http://workspace.earthcube.org/all-hands-meeting/news-announcements/2014-all-hands-meeting-overview>
Brian Westra: (13:54) DataONE, like LTER
Laura Russell: (13:54) If you have interest in joining the core group for the TDWG data quality group, you can follow this topic on the GBIF Data Quality Group...the charter is coming soon, <http://community.gbif.org/gg/forum/topic/43303/core-member-invitation-tdwg-bdgiu/>
Alex Thompson: (13:54) @Mariel: Katja Selmann and the Tri-Trophic Interactions TCN have also been working on association data
Elizabeth Sellers (USGS Reston, VA): (13:54) SO many issues... SO little time. :o)
Elizabeth Sellers (USGS Reston, VA): (13:55) Deb - is there any specific goal/piece of information you're looking to get out of this group to inform/help iDigBio?
Reed Beaman: (13:55) With respect to Jason's comment about sensors, there is a project called DataTurbine that assimilates streaming data from sensors and the idea of humans as sensors has come up in some of those discussions. Data Turbine is based at UC San Diego, I believe.
Brian Westra: (13:56) I'm just starting to dig into how to use Specify and manage more data files besides the specimen (sequence data, GC-MS) and relationships between the specimens (insect collected from a flower)
John Wieczorek: (13:56) Only that it is too big this way.
John Wieczorek: (13:56) I don't have a particular agenda though.
Chuck Miller: (13:57) Focus = organization, structuring of a topic. What are the components of "data management"?
Brian Westra: (13:57) Will you share the chat too? Thanks!
Laura Brenskelle: (13:57) In my mind, data management also includes a "longevity" aspect. how can we keep all of this data so that it is accessible to people for a long time after it is made? which is especially challenging if we have digital-born data in various new formats.
Mariel Campbell: (13:57) Thanks, Alex!
Sarah Timm: (13:57) I'd like to mention that I am working with some people to put a proposal forth to the Geological Society of America (GSA) for their board meeting this month with the focus on action and support from a "top down" approach so we can have workshops and procedures etc to start getting specimens preserved that haven't been before ie- university materials. More of a priority for geological collections management. The link to the google doc for that which anyone is welcome to edit and sign is here: <https://docs.google.com/document/d/19y6JK2Gzvc8CD4R0YioGnvwCGTRgJcDCuXzISD-eu0/edit>
Jennie Kluse: (13:57) is there a system to FIND and then reconcile duplicate specimen records in different institutions with different histories of edits and annotations?
David Bloom: (13:58) Gotta run, meeting at 11amPT. Thanks folks.
Brian Westra: (13:58) Some use data curation and data management interchangeably.
Laura Russell: (13:58) Sorry, must sign-off for another call.
Chuck Miller: (13:58) Gotta go too.
Joanna McCreffrey: (13:58) @Mariel Katja's work is here: https://docs.google.com/spreadsheets/cc?key=0AvHeAV_gLfnadEhVjRKSUDxUmK3HX3vWmtSGVnNLE&usp=drive_web&gid=0
Chuck Miller: (13:59) Great topic.
John Wieczorek: (13:59) Eager to see so much enthusiastic participation.
Sarah Timm: (13:59) Agreed!
Rich Rabeler: (13:59) Jennie: Filtered Push, Scatter-Gather-Reconcile, and a built-in app in Symbiota attempt to do that.
Jennie Kluse: (13:59) Great! That's what I'm looking for!
Arturo H. Ariño: (13:59) Jennie: Without some GUID/LSID in place it's somewhat flimsy, depending on some match metrics that may at best give a probability function rather than reasonable certainty.
Ron Eng: (14:00) Thanks to everyone. So much to learn!
Elizabeth Sellers (USGS Reston, VA): (14:00) I got temporarily booted out of it when it got busy... it happens
Sarah Timm: (14:01) Can't get in google doc yet, says it's too full!
Gail Kampmeier: (14:01) Chrome handles it better than some other browsers
Elizabeth Sellers (USGS Reston, VA): (14:01) Titanpad is also a useful alternative... (sometimes) for group doc/note-taking
Elizabeth Sellers (USGS Reston, VA): (14:02) Yes - Etherpad is similar
Alex Thompson: (14:02) Leaving cameras off does help with scaling a meeting to this size
Alex Thompson: (14:02) we probably couldn't have more than 10-15 people with cameras on before the meeting was unusable
Deb Paul: (14:03) https://www.idigbio.org/wiki/index.php/Data_Management_Interest_Group
Mariel Campbell: (14:03) Thanks, Joanna!
Laura Brenskelle: (14:03) monthly
Alex Thompson: (14:03) Chat history all the way back is preserved by the way, its just cut off in display
John Torgeresen - Michigan: (14:03) Every other month is great.
Mariel Campbell: (14:03) every other if fine
Liath - Univ. of TX: (14:03) every other month sounds good
Elizabeth Sellers (USGS Reston, VA): (14:03) Vote: Monthly
Sarah Timm: (14:04) I like monthly
Roger Burkhalter: (14:04) Every other month
Elizabeth Sellers (USGS Reston, VA): (14:04) or every other month
Mariel Campbell: (14:04) But monthly works too!
Christina Byrd: (14:04) Monthly
Brian Westra: (14:04) Monthly
Janet Bala: (14:04) Every other
Larry: (14:04) every other month
Roger Burkhalter: (14:04) Survey monkey
Brian Westra: (14:04) Either is fine w/me
Tom Trombone: (14:04) Every other month
Jennie Kluse: (14:05) Thanks Deb!
Sandra Brantley: (14:05) Thanks Deb and everyone!
Brian Westra: (14:05) Thanks!
Elizabeth Sellers (USGS Reston, VA): (14:05) Challenges: One of the biggest challenges we have encountered with both observation-based and specimen-based data is 'cleaning' it to improve indexing for search and retrieval - without reinterpreting or replacing original data.
Roger Burkhalter: (14:05) Thanks
Arturo H. Ariño: (14:05) thanks
Liath - Univ. of TX: (14:05) Thanks
Elizabeth Sellers (USGS Reston, VA): (14:06) Exactly. We don't want to 'change the story' told by the data.
Laura Abraczinskas: (14:06) Thanks everyone
Mariel Campbell: (14:06) Thanks!
Elizabeth Sellers (USGS Reston, VA): (14:07) And the only other comment I have is that I think this community has made great strides in overcoming the 'black hole' concept - data goes in - goes through magic processing cloud - comes out the other side looking different
Janet Bala: (14:07) Will the chat summary show up in the google doc?
Dorothy Allard: (14:07) Thanks, look forward to the next one
Elizabeth Sellers (USGS Reston, VA): (14:07) Thanks!
Deb Paul: (14:08) Bye Everyone!