

Minutes of Internal Advisory Committee Meeting

Date/Time: Wednesday, 1 May 2019 2:00–3:00 PM Eastern

Attendees: Allie Blackwell (iDigBio); Amy Kasameyer (Pteridophyte TCN); Andy Miller (MycoPortal); Anna Marie Countie (HUH); Ashley Dineen (EPICC); Cindy Skema (MAM); David C Blackburn (oVert); Diego Barroso (MiCC); Ed Gilbert (ASU BioKic); Jenn Yost (CAP TCN); Katie Pearson (CA Phenology); Laurie Anderson (Cretaceous World); Matthew Pace (Endless Forms); Michael Denslow (SERNEC); Michaela Schull (Harvard Herbaria); Nico Franz (ASU); Phil Tonee (UNM); Phillip; Renato Figueiredo (ACIS); Rick Levy (Denver Botanic Gardens); Rick Williams (RMBL); Ryan Allen (SoRo); Talia Karim (FIC); Jonathan Kennedy (Harvard Herbaria); Shari Ellis (iDigBio); David Jennings (iDigBio); Molly Phillips (iDigBio); Gil Nelson (iDigBio)

Discussion Items:

Serving and Storing Images

- This IAC meeting discussed how TCNs are serving images (e.g., jpgs) during their grant and how their participating institutions are (or anticipate) serving images for specimens outside the scope of the TCN, after the grant ends
- It seems that the TCNs have a mixed–bag approach based on what the institution had in place prior to already joining a TCN. See below:

Cyverse

- There were quite a few institutions that were using Cyverse to serve images or have looked into it as an option.
- SERNEC is using Cyverse to serve images. They have a sub–contract with them and pay Cyverse for storage. They harvest the URLs from those jpgs. They actually use them for RAWS as well, but don't serve them to the portal. They have an appliance through Symbiota – it connects to and populates the records with the URL and serves the images
- Cyverse will give out a certain amount of storage for free (but its a low amount of space).
- They require you to have a sustainability plan (backup plan) – If Cyverse loses funding and doesn't exist anymore, they want you to know what you are going to do with all the images
- Most people believe that Cyverse is a much better deal than any other option.
- Cost: varies but estimate \$256/TB/year

Amazon

- Harvard Herbaria is getting a similar deal from Amazon S3, that is comparable to Cyverse if you just look at storage. The prices are (storage only, not online delivery):
 - \$250/TB/year for S3 Standard
 - \$150/TB/year for S3 Intermittent Access

- \$50/TB/year S3 Glacier
- Harvard Herbaria uses S3 extensively – but not exclusively. They have a research computing facility within the faculty of earth and science. It keeps copy of all their files as well within the research computing storage. They charge them 50 dollars per TB (equivalent to Amazon Glacier).

Hardware

- MAM bought hardware as part of their funding from NSF. They've been using all that space for the Academy of Natural Sciences of Philadelphia
- That goes for the image serving and archive storage. There is no space on the server to put new collections
- The price included long term maintenance, IT support, and upgrades
- This is good as long as the hardware is functioning – they don't know when it's going to stop.

Specify

- A few institutions were using Specify attachment server. It's a public facing web server.
- When you upload attachments – they save the copies to the attachment server. It renames your file and finds a URL. Then you can export those URLs when you do your data export to upload it to IPT. It creates a persistent URL.
- Specify is just providing a software attachment tool for you – but you would still have to find hardware
- It's in ingestion tool
- It handles all of the metadata creation for those images and files as well.
- It creates a relationship between the specimen objects and image objects and also patches together a URL that you can export your images into IPT or ingestion in iDigBio

Internet Archive

- Internet Archive would be good for small institutions – large institutions may not be able to absorb it
- Generally, Internet Archive was receptive to the idea. They don't charge and people can make their own accounts independently. Haven't been able to reach person that we have initially interacted with at internet archive.

iDigBio news

- The Digital Data Conference will be held from June 10-12. We encourage all to attend. Registration link: <https://www.eventbrite.com/e/3rd-annual-digital-data-conference-methods-protocols-and-analytical-tools-for-specimen-based-tickets-54760252389>