

Meeting Information

Meeting Date/Time: April 25, 2014 / 10:00-11:00 AM USA Eastern Daylight Time

External Advisory Board Members: Vince Smith, Stan Blum, Stinger Guala, Donald Hobern, Karen Powers (absent)

iDigBio PIs: Larry Page, Bruce MacFadden, Greg Riccardi

iDigBio Staff: David Jennings, Renato Figueiredo, Shari Ellis, Cathy Bester

Executive Summary

The External Advisory Board (EAB) members met with the iDigBio PIs and project staff on April 25, 2014, via an Adobe Connect virtual conference. Presentations were given by David Jennings and Greg Riccardi to review iDigBio's activities and progress during the past year. Following the presentations, the EAB provided advice to iDigBio on several aspects of the project, including taxonomy and serving data to GBIF. The meeting concluded with a discussion and plan to rotate the membership of the EAB, including a plan for them to meet more frequently in the future.

iDigBio Presentations

The following presentations that were given to the EAB by iDigBio personnel during the meeting are provided as attachments:

Topic	Presenter
iDigBio External Advisory Board Meeting	David Jennings
Data Workshop Honolulu March 2014	Greg Riccardi

Meeting Minutes

Role of the EAB

The EAB inquired how they can better align their activities with those of iDigBio and vice versa. iDigBio has been active for 3 years with a focus on digitization activities and cyberinfrastructure development and is ready to begin working on E&O and research use of the data along with international alignment of activities. iDigBio welcomes advice, guidance, and participation from the EAB. NSF is encouraging the EAB to meet independently during each year, in addition to this annual meeting with iDigBio, including informal discussions at other professional meetings.

Request for Advice and/or Suggestions

Greg Riccardi stated that iDigBio recognizes the need for a global taxonomy structure, which has been planned from the beginning. Stinger Guala informed the group that BISON is currently working on a hierarchical taxonomic structure down to family level for all taxa on earth. This may be useful for iDigBio and will be provided to other data aggregators. iDigBio should consider what requirements are needed for taxonomic hierarchy – GMA/GBIF architecture may be better for us. Greg also noted that discovery is much different than the data analyses services that Rich Pyle has – are these reality and can they be employed and shared? Data analyses require consensus taxonomy more so than data discovery.

Donald Hobern noted that it would be valuable for iDigBio to carefully consider the dual role in organizing geospatial data and the higher value role as a metadata catalog for materials (images, audio, etc). Donald suggested that the iDigBio Portal could specialize for these roles, the second of which is currently poorly served by other data portals; it is important that iDigBio provide these useful services to data users. Donald also stated that this second role of providing media other than just geospatial data will act as a catalyst for making data more discoverable and groupable, which could include a toolbox for collections and taxonomic use of the data. The National Museum of Natural History (London) is taking a similar strategy, using a toolbox/community-centric approach that has significant value. In regards to EOL, they have a taxon-centric view; there is potential for other ways to present data through iDigBio. Larry Page remarked that this is an excellent way to view the value of iDigBio and will continue this dialog with Donald and the EAB.

Stinger Guala inquired about the status of iDigBio data service to GBIF. iDigBio (ACIS) is ready to send data to GBIF but needs to know the exact format of the data – will data be harvested via web links, RSS feed, or IPT? Donald Hobern stated that the most preferred method from GBIF's standpoint is publicly through web services, which can be done via zip files that are retrievable and can be automated at regular intervals or when prompted. The data do not all need to go into an IPT. Tim Robertson should be put in contact with Alex at ACIS to work out the details to enable progress on this front. [Subsequently, Larry talked to Jose Fortes at ACIS, and they agreed that Reed Beaman of iDigBio would work with Tim Robinson.]

Rotation of EAB Membership

The NSF site visit team recommended that the members of the EAB be rotated and the diversity increased. iDigBio asked which of the EAB members would like to drop off and for the EAB to recommend replacement board members. Larry Page suggested that is important for Donald Hobern or Stinger Guala to remain on the EAB due to iDigBio's relationship with GBIF. Donald indicated a willingness to remain as a member of the EAB but could also rotate off if need be. Once the new EAB membership is established, a chair will be elected by the members. The EAB will be expected to meet once or twice a year, independent of iDigBio, with reports from these meetings submitted to iDigBio.

Final Comments

Stan Blum and others on EAB pointed out that one valuable service and role of iDigBio is that of a digitization clearinghouse. iDigBio's ongoing digitization workshops are highly valuable to the community. iDigBio agreed and noted that they will continue providing this service in the future.

Vince Smith also noted that data providers will need access to usage statistics of data provided to data aggregators such as iDigBio. Data are being utilized at the aggregator data portals in increasing amounts while being accessed in lesser numbers through the data providers' individual portals. Providing this information is difficult but is on the iDigBio road map.

Finally, the EAB congratulated iDigBio on a job well done.

Addendum

Based on activities to date and the recommendation from NSF, iDigBio suggests that Vince Smith and Donald Hobern remain on the EAB and that others rotate off (with a big thank you from iDigBio!). This suggestion is made in large part because Vince and Donald are located outside the U.S., whereas Stan Blum, Karen Powers, and Stinger Guala are located in the U.S. and can more easily meet with and interact with iDigBio. iDigBio sponsors 1-2 workshops/month on average, providing ample opportunities for meetings, and we welcome any of you to visit iDigBio at any time. Secondly, NSF expects iDigBio to become more engaged internationally, and Vince and Donald can perhaps help with that through targeted recommendations from EAB.

If agreeable with all, we ask you to suggest potential new members for the EAB. NSF has requested a better balance of gender and underrepresented minorities. Potential members (suggested by Larry Page) are Bonnie Styles, Director of the Illinois State Museum; Hank Bart (an African-American), Director of the Tulane Museum of Natural History; and Michael Mares (Hispanic), Director of the Sam Noble Oklahoma Museum of Natural History. If you would like additional recommendations from us, or are on the board and wish to stay on, please let us know.

David Jennings has asked Karen Powers to organize and lead this discussion among present board members and make their recommendations known to us. Please note that NSF must approve all appointments to the EAB.

Thank you again for your support and involvement with iDigBio!

Agenda for the 2014 iDigBio External Advisory Board Meeting

- Date:** April 25, 2014
- Time:** 10:00 - 11:00 AM Eastern Daylight Time
- Invitees:** **EAB Members:** Donald Hobern, Karen Powers, Stinger Guala, Stan Blum, Vince Smith
Required iDigBio Attendees: Larry Page, David Jennings
Optional iDigBio Attendees: Pam Soltis, José Fortes, Greg Riccardi, Bruce MacFadden, Joanna McCaffrey, Shari Ellis, Cathy Bester
- Location:** iDigBio Conference Room (Building 105, Room 310)
- Remote Access:**
1. Go to <http://idigbio.adobeconnect.com/idigbioeab/>
 2. Choose **Enter as Guest**
 3. Type your first and last name in the box provided
 4. Click **Enter Room**

Please review iDigBio's web conferencing wiki page prior to the meeting to improve your experience with Adobe Connect (https://www.idigbio.org/wiki/index.php/Web_Conferencing). If you are unable to connect via computer, the audio-only portion of the seminar will be available via telephone conference:

- *United States: 866-284-8871*
- *Other Countries: check for your access number at <http://www.meetingone.com/us/customer/a1international.cfm>*

If you experience difficulties connecting remotely, please contact David Jennings at 352.273.1906 or Kevin Love at 352.294.1924

10:00 AM Call to Order

- Review of iDigBio's progress in Fiscal Year 3
- Specific advice from the EAB to iDigBio
- Rotation of EAB membership

11:00 AM Adjourn

iDigBio External Advisory Board Meeting

David Jennings
iDigBio Project Manager

April 25, 2014
10-11am Eastern Daylight Time



This material is based upon work supported by the National Science Foundation under Cooperative Agreement EF-1115210. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



Agenda

- Review of iDigBio's progress in Fiscal Year 3
- Specific advice from the EAB to iDigBio
- Rotation of EAB membership

Role of External Advisory Board

“External Advisory Board (EAB) whose membership will be subject to the approval of NSF’s cognizant program official and will meet at least once a year; provide written and verbal advice to the national resource on its activities, including progress and integration of digitization projects, research, education and outreach activities among all funded institutions; advise the national resource leadership on strategic directions and management policies.”

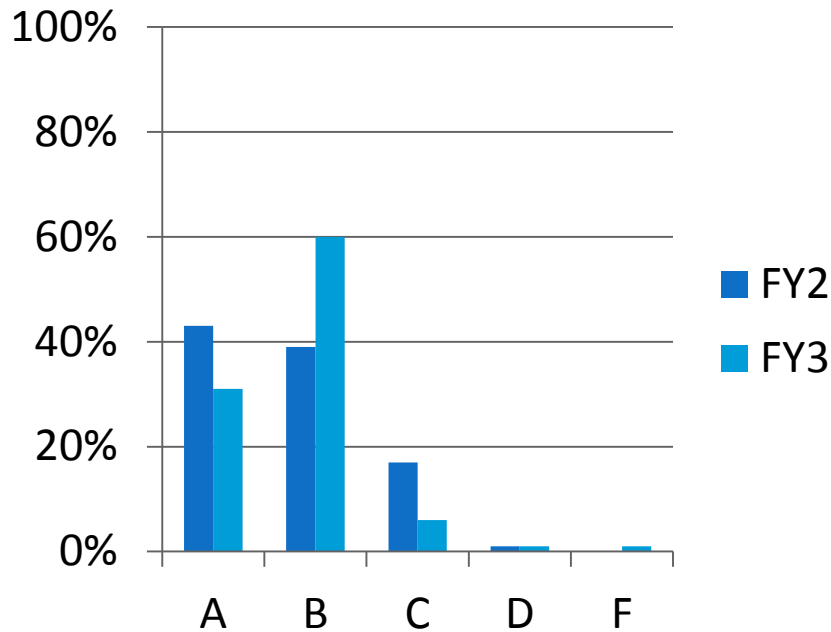
Key Accomplishments & Community Feedback

- Workshops and Training
- Data and Media Ingestion
- Website and Portal Upgrade
- Collaboration / TCN Support

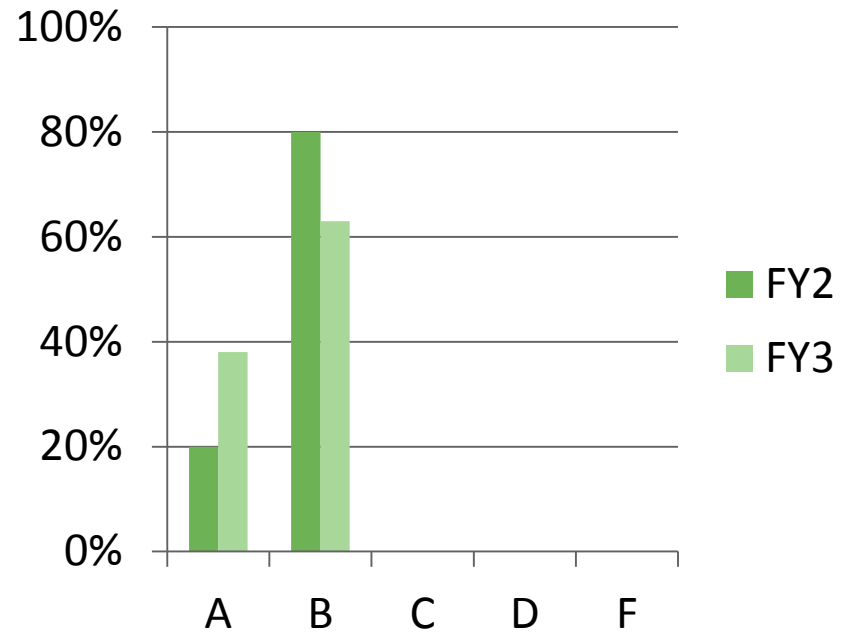
Community Feedback – Overall

What grade would you give iDigBio in moving forward national digitization efforts?

External (non-iDigBio)

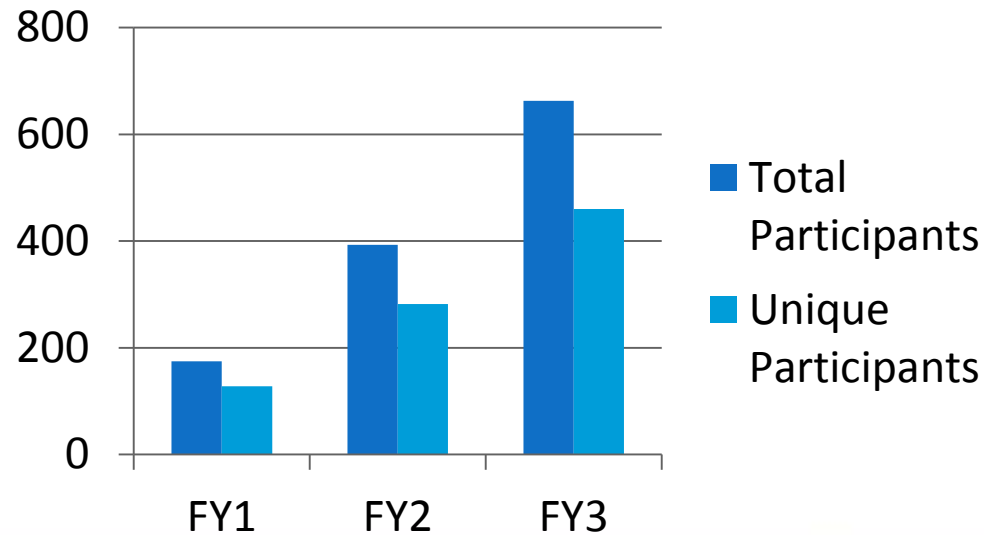


Internal (iDigBio)

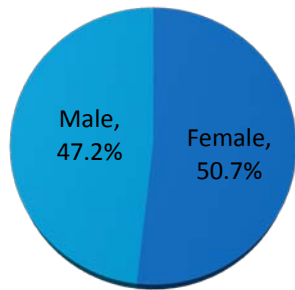


Workshops and Training

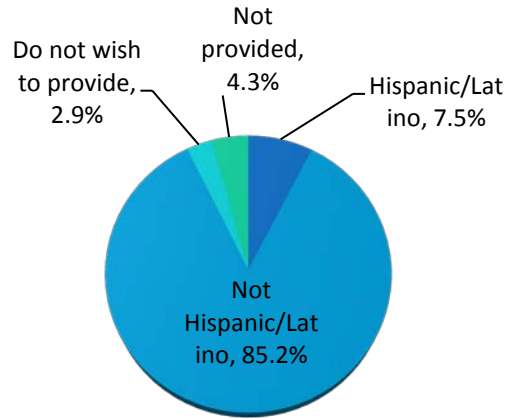
	<i>Total Number of Workshops, Summits, and Symposia</i>	<i>Total Number of Participants</i>	<i>Number of Unique Participants</i>	<i>Number of Unique Institutions</i>
Fiscal Year 1	5	175	128	65
Fiscal Year 2	12	393	282	160
Fiscal Year 3	19	663	460	211
OVERALL	36	1231	682	305



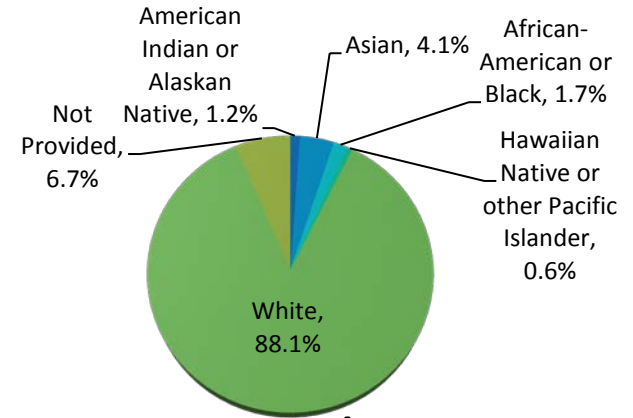
Workshop Demographics



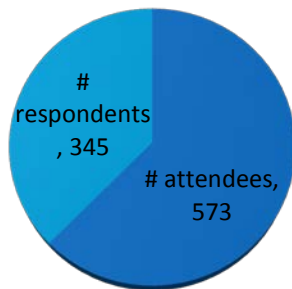
Gender



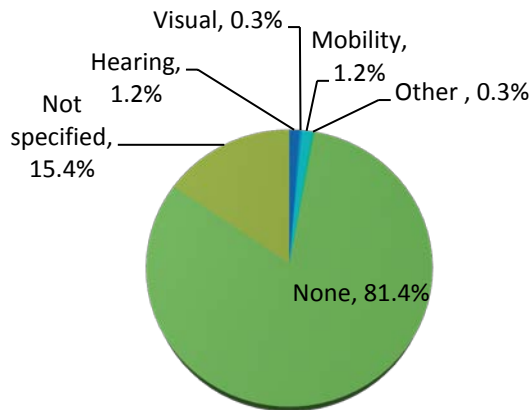
Cultural Background



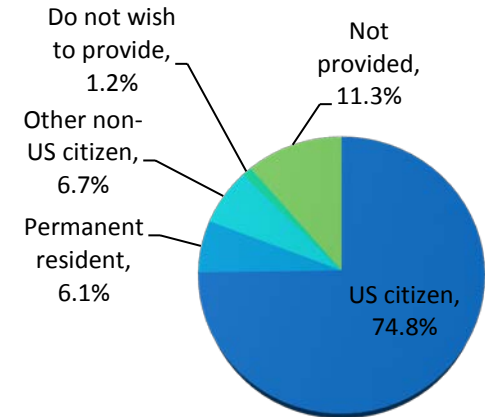
Ethnicity/Race



Average Response Rate = 60%



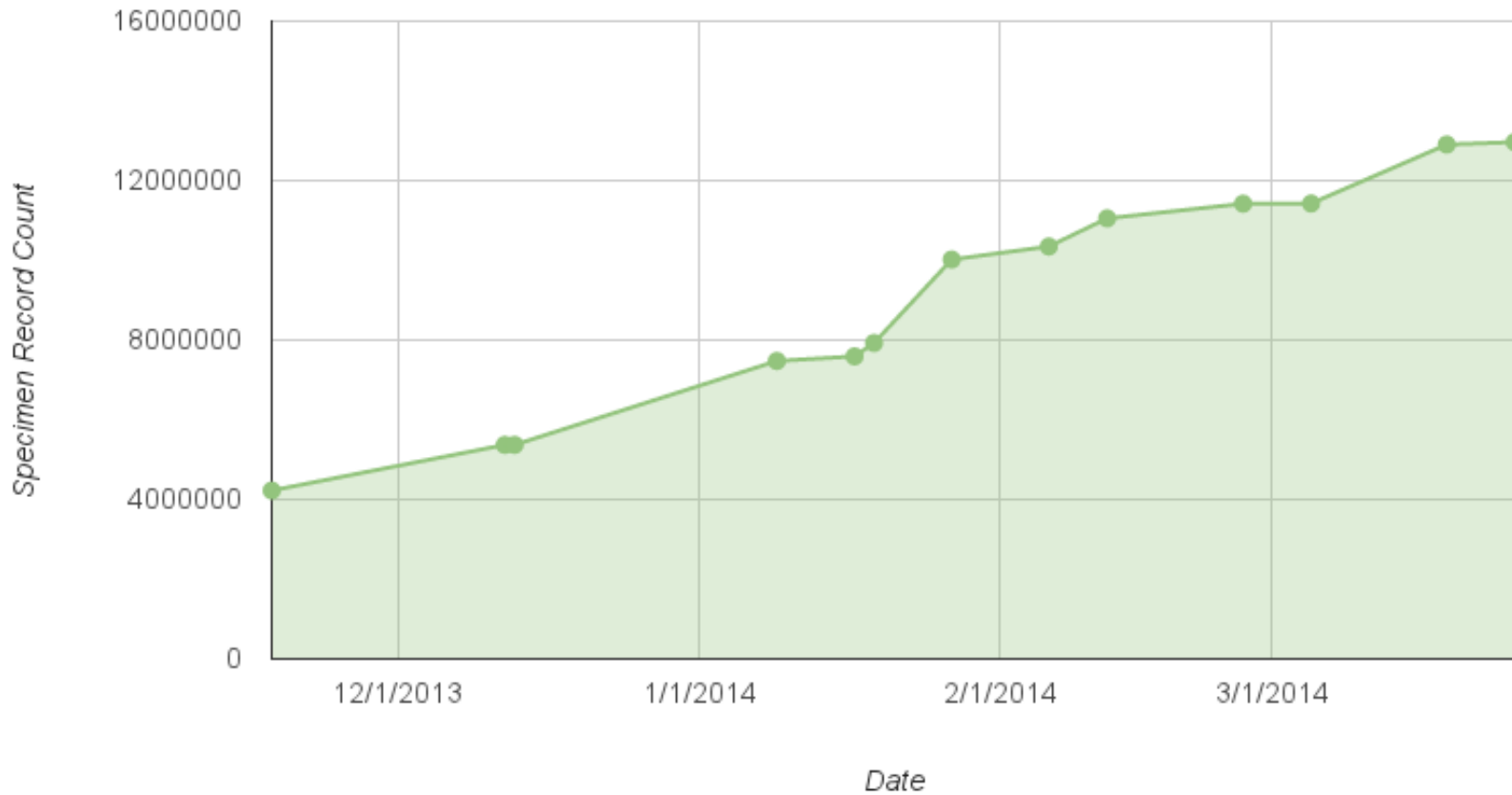
Disability



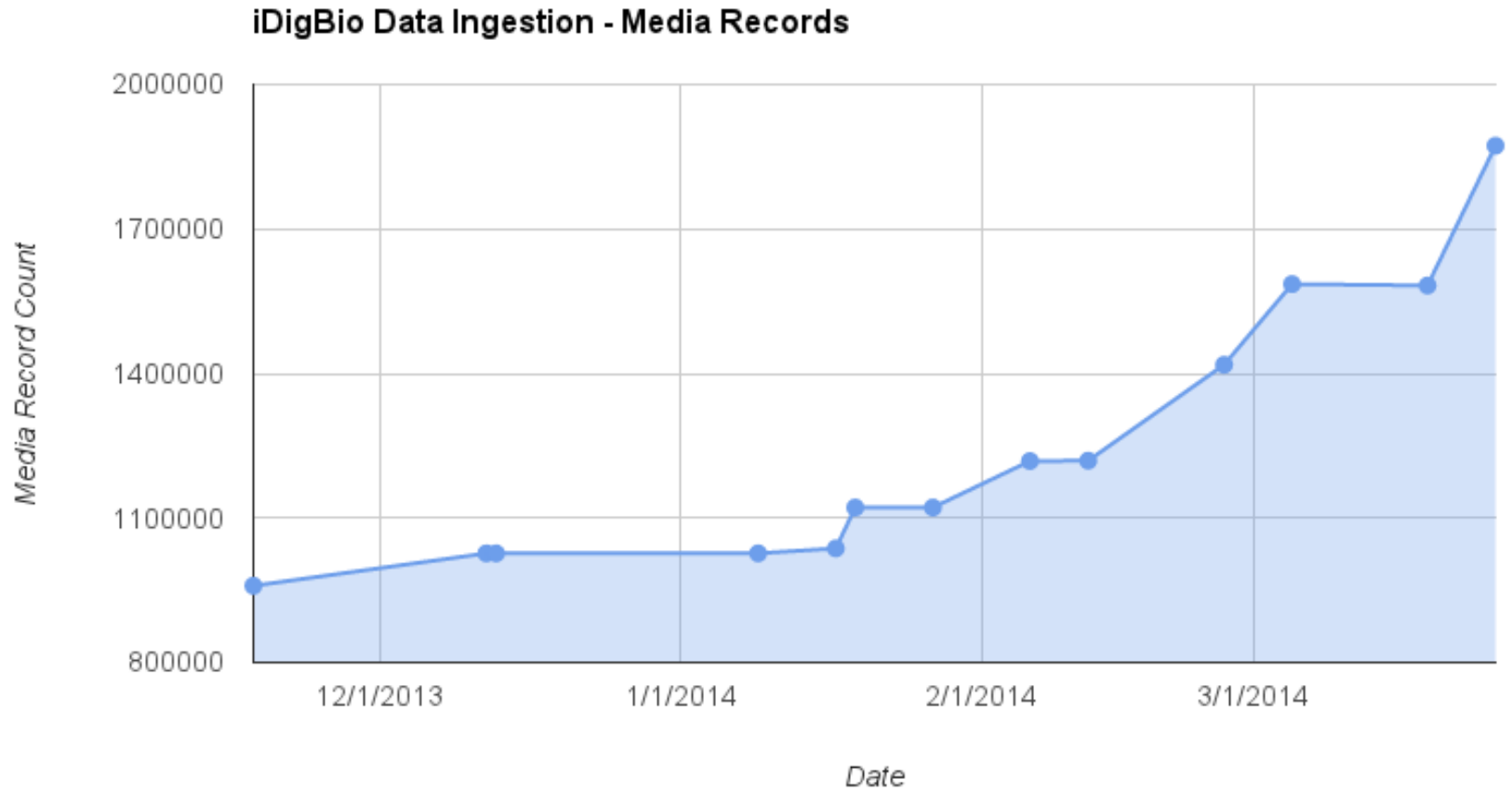
Citizenship

Data Ingestion: 13,268,750 Specimen Records

iDigBio Data Ingestion - Specimen Records

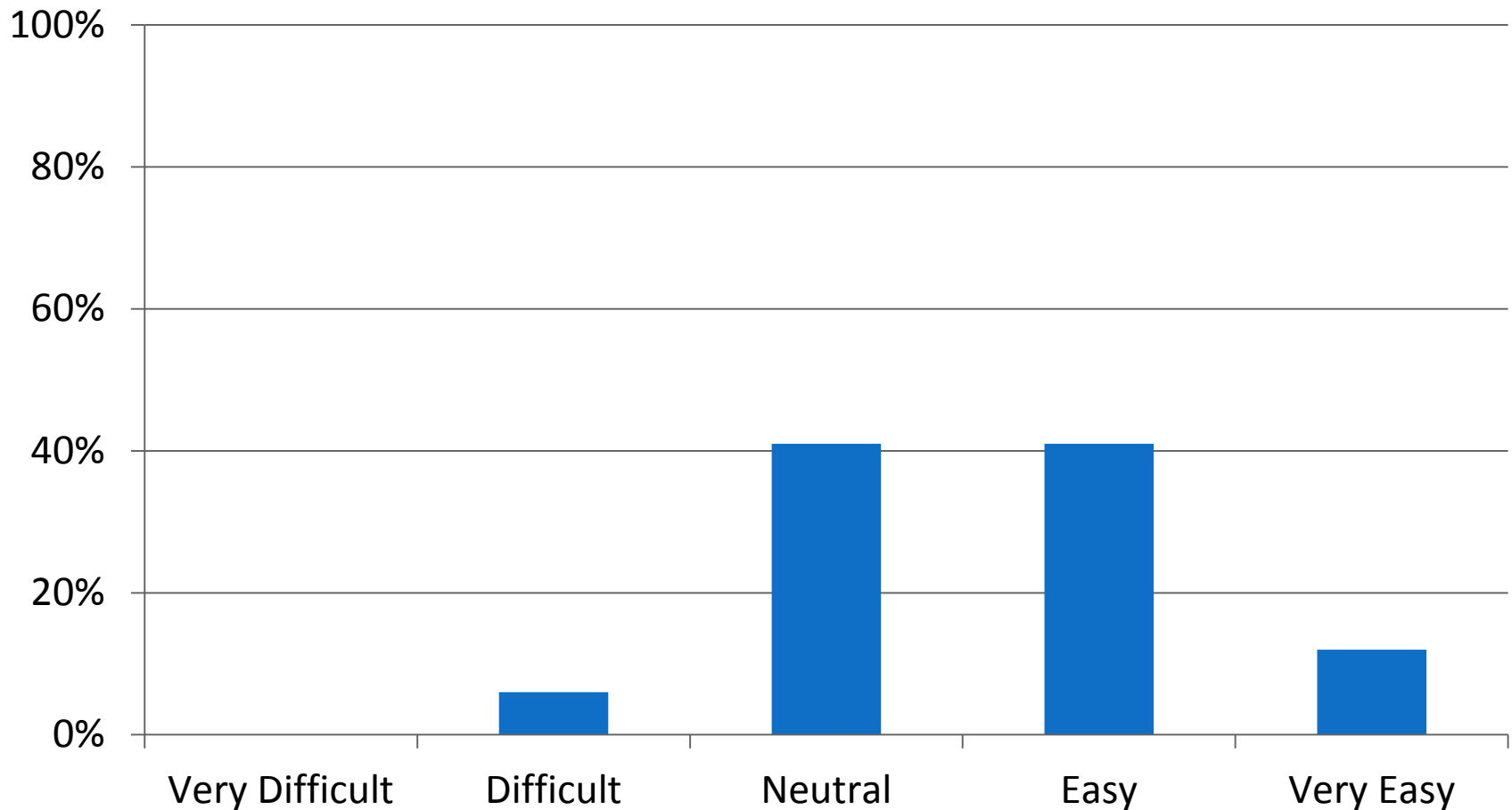


Data Ingestion: 1,985,011 Media Records



Community Feedback – Data Ingestion

Level of difficulty in submitting data to iDigBio

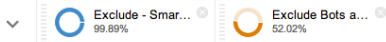


Portal Traffic

Audience Overview

Jul 1, 2013 - Apr 18, 2014

Email Export Add to Dashboard Shortcut

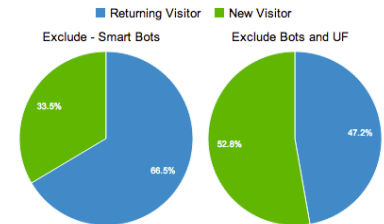
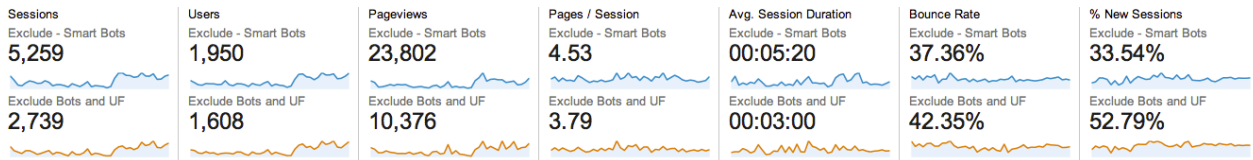
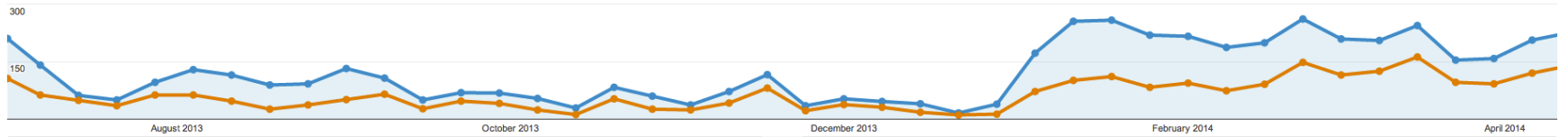


Overview

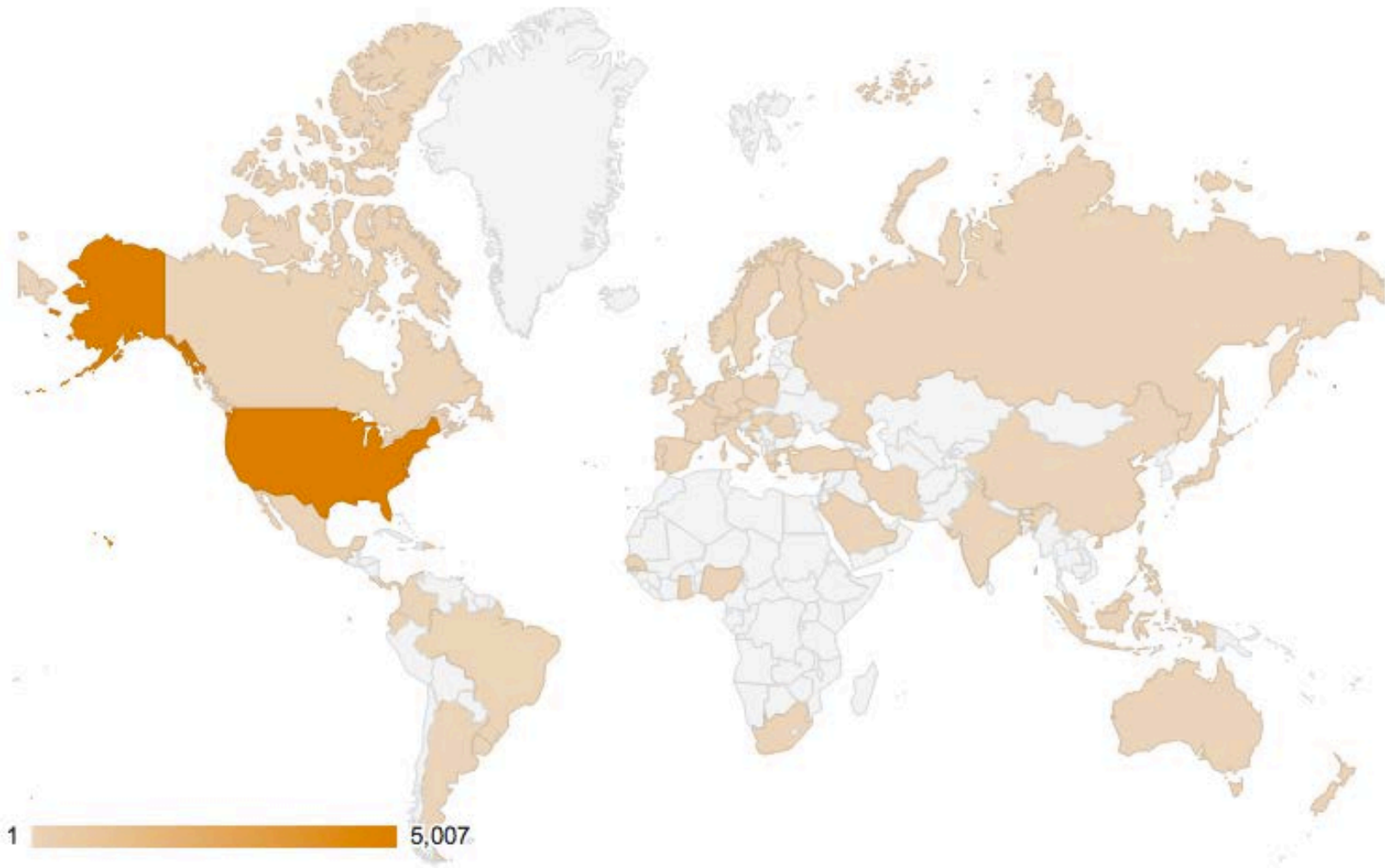
Sessions vs. Select a metric

Hourly Day Week Month

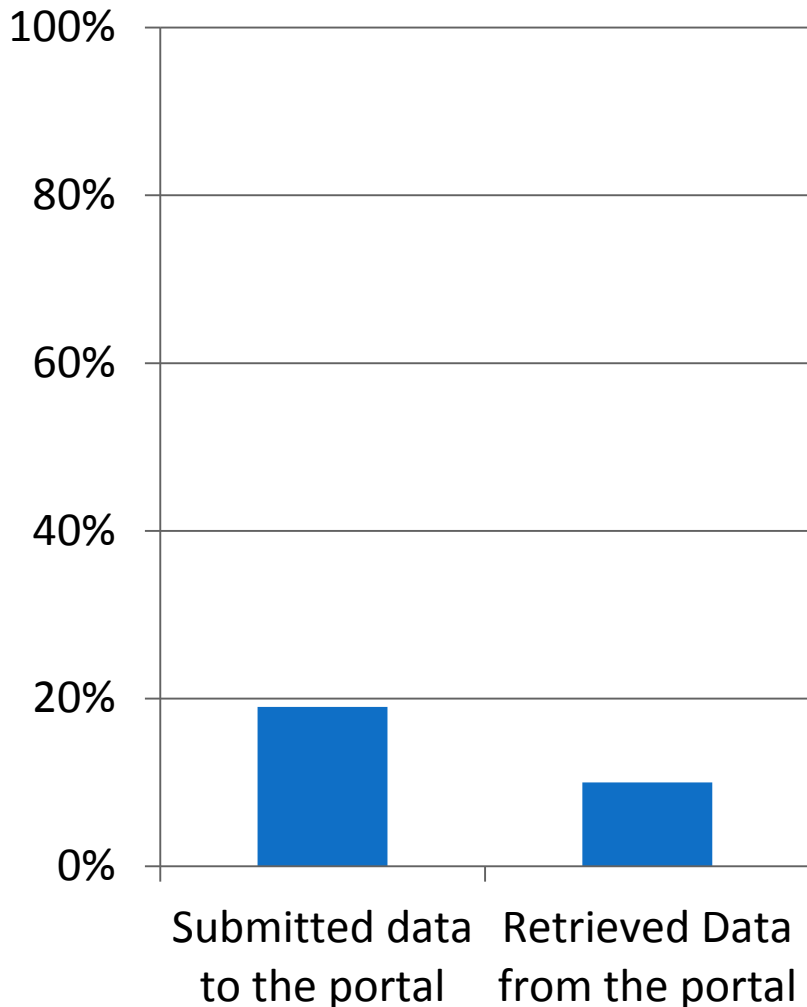
Sessions (Exclude - Smart Bots)
 Sessions (Exclude Bots and UF)



Portal Reach



Community Feedback – Portal



“The data portal seems to be going well. The team has done well in dealing what turns out to be a complex problem.”

Website Traffic

Audience Overview

Jul 1, 2013 - Apr 18, 2014

Email Export Add to Dashboard Shortcut



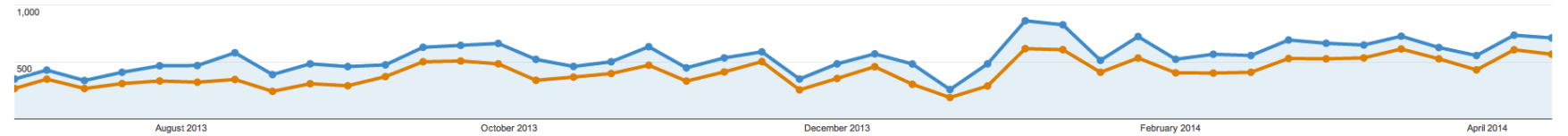
Excluded - Smart Bots 97.84% Excluded Bots and UF 73.49%

Overview

Sessions vs. Select a metric

Hourly Day Week Month

Sessions (Exclude - Smart Bots) Sessions (Exclude Bots and UF)



Sessions

Exclude - Smart Bots
23,012

Exclude Bots and UF
17,286

Users

Exclude - Smart Bots
10,689

Exclude Bots and UF
10,090

Pageviews

Exclude - Smart Bots
77,084

Exclude Bots and UF
48,686

Pages / Session

Exclude - Smart Bots
3.35

Exclude Bots and UF
2.82

Avg. Session Duration

Exclude - Smart Bots
00:04:34

Exclude Bots and UF
00:03:13

Bounce Rate

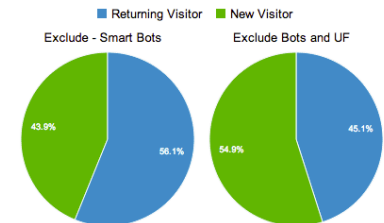
Exclude - Smart Bots
52.14%

Exclude Bots and UF
56.06%

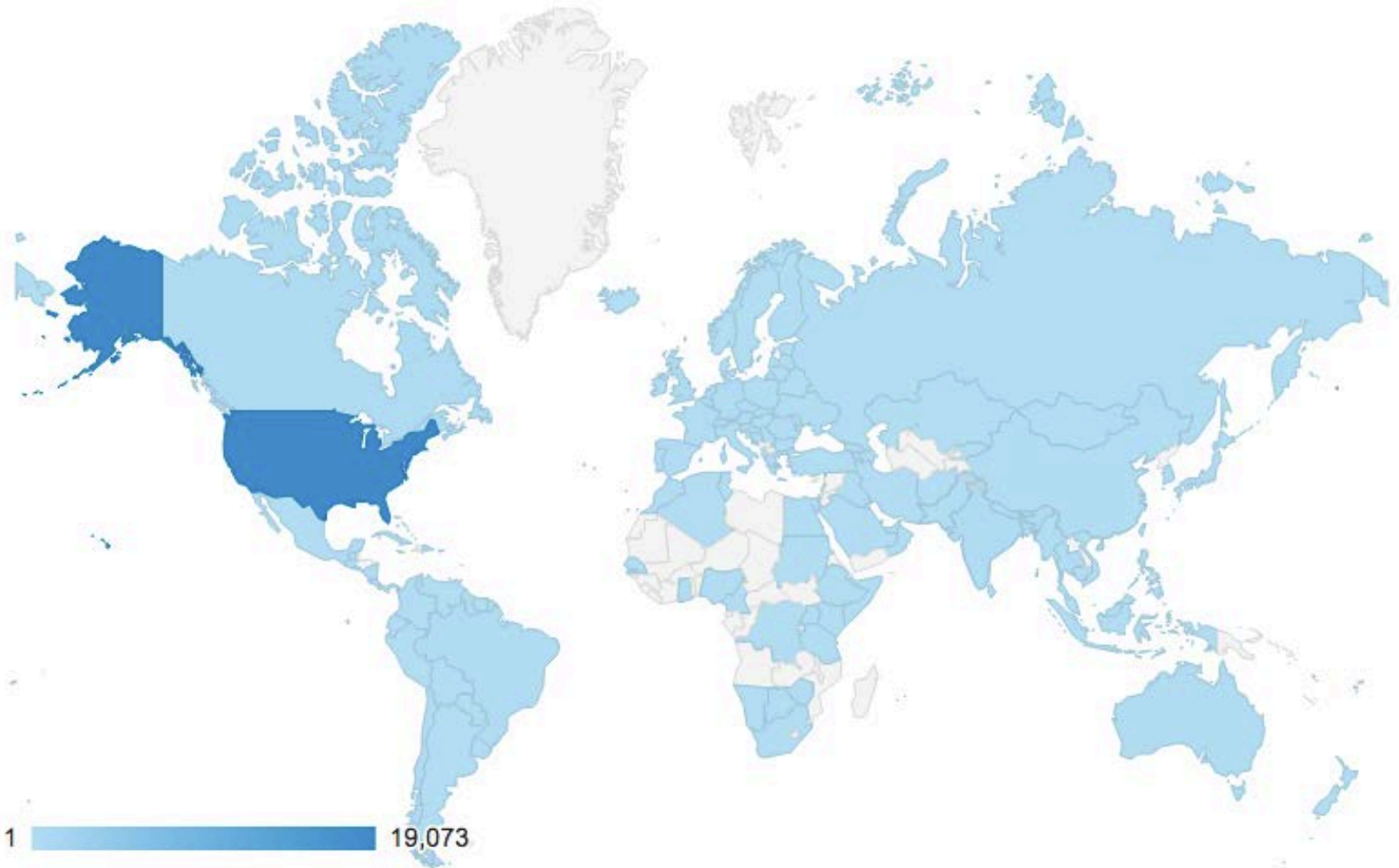
% New Sessions

Exclude - Smart Bots
43.86%

Exclude Bots and UF
54.93%

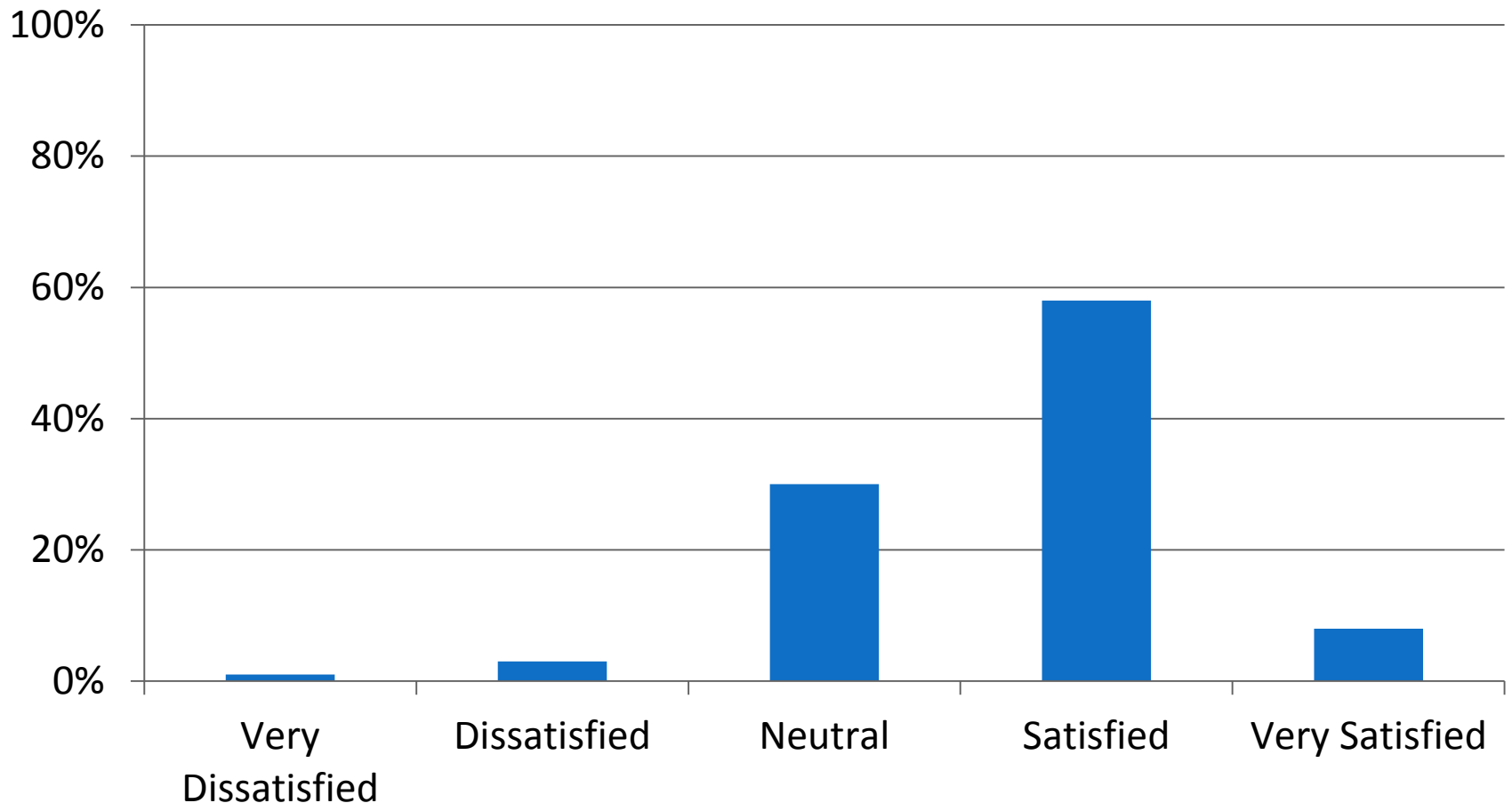


Website Reach



Community Feedback – Website

Overall satisfaction with the iDigBio website

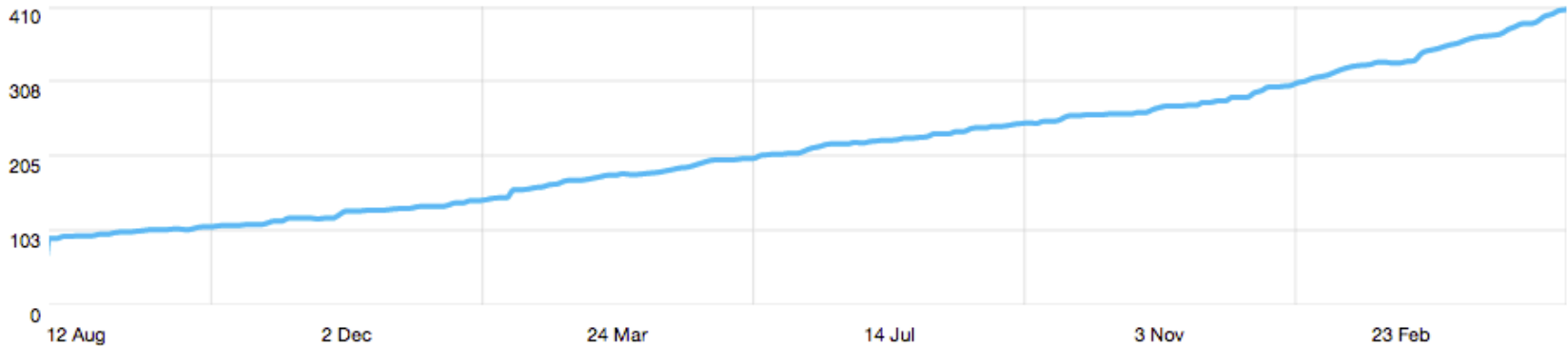


Collaboration – Adobe Connect

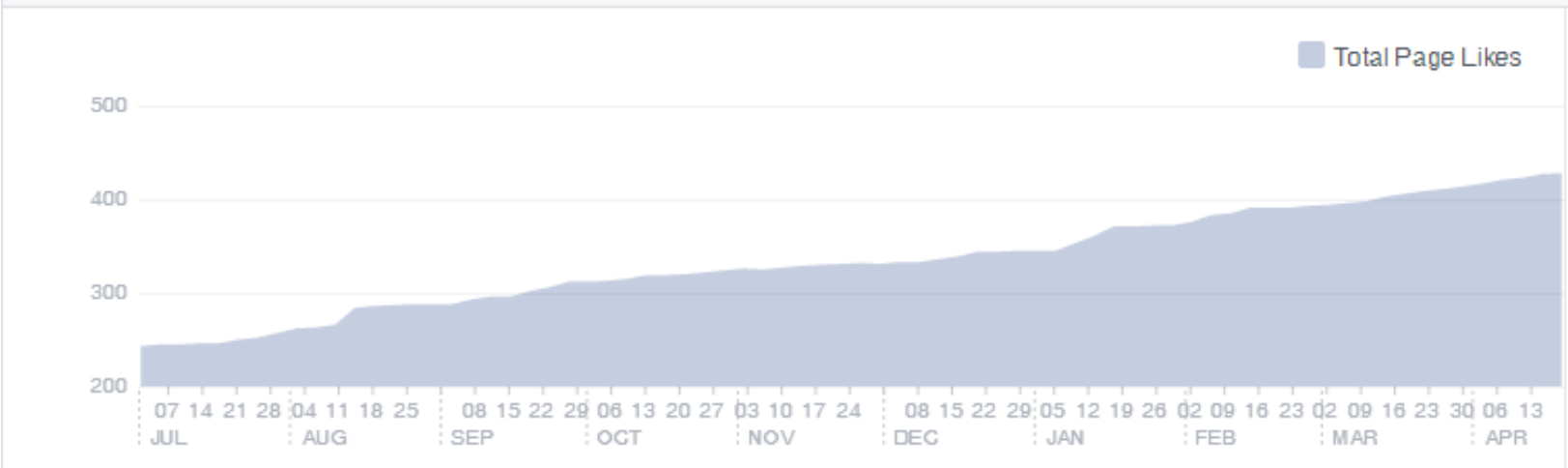
	FY1	FY2	FY3
Total Users	40	65	86
Total Hosts	12	26	35
Distinct Meeting Rooms	37	62	65
Total Meeting Hours	768	4429	5932
Total Host Hours	304	2182	2268
Peak Concurrent Users	22	44	44
System Storage Consumption (GB)	0.8	2.6	5.2
Meeting Storage Consumption (GB)	6.8	8.1	484.3

Social Media – Twitter/Facebook

405 followers as of 4/20/2014 (days shown in Pacific time)



Total Page Likes as of Today: 433



Social Media – Vimeo

This year at a glance



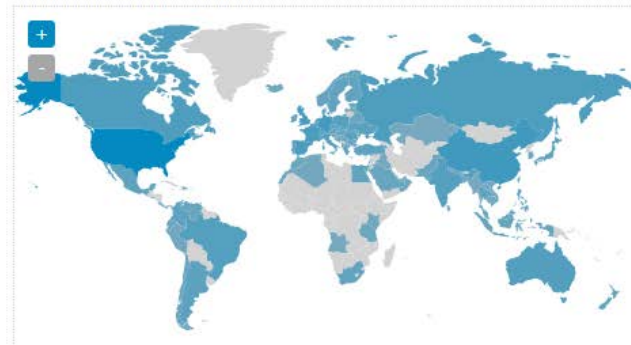
Yearly Overview

Date	Plays	Loads	Likes	Comments
Apr 2014	152	569	0	0
Mar 2014	339	923	0	0
Feb 2014	342	917	1	0
Jan 2014	285	1,065	0	0
Dec 2013	200	312	0	0
Nov 2013	191	365	0	0
Oct 2013	162	1,134	0	0
Sep 2013	106	1,361	0	0
Aug 2013	118	1,446	0	0
Jul 2013	111	1,295	0	0
Jun 2013	136	1,258	0	0
May 2013	106	1,250	0	0
Apr 2013	32	318	0	0
Totals	2,280	12.2K	1	0

+ View your full total stats

Geographical Locations

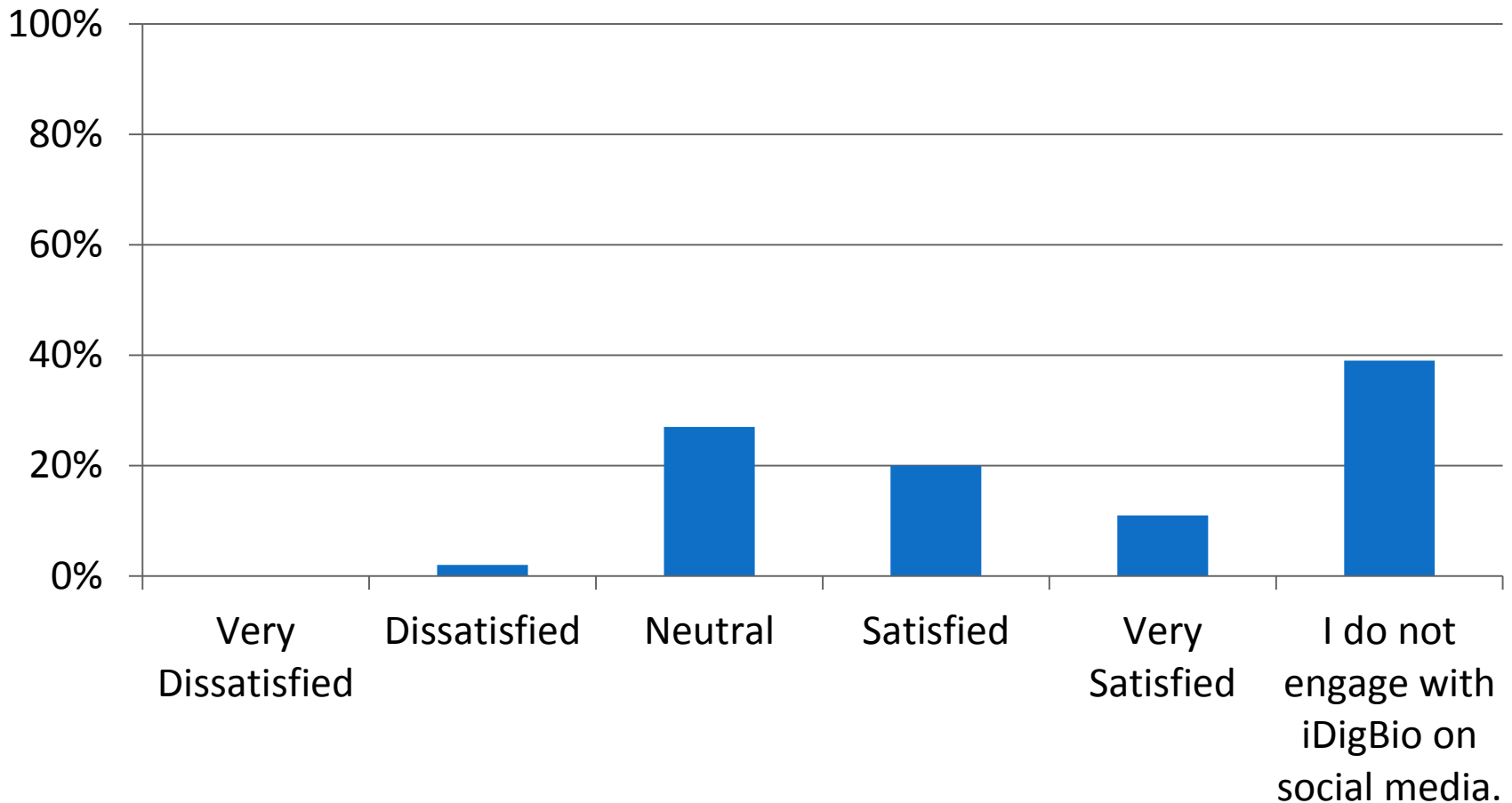
Drag, zoom and pan around.



+ View your full geo stats

Community Feedback – Social Media

Overall satisfaction with iDigBio's use of social media



Data Workshop

Honolulu March 2014

Greg Riccardi
Florida State University
iDigBio
griccardi@fsu.edu



This material is based upon work supported by the National Science Foundation under Cooperative Agreement EF-1115210. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



Goals of workshop

- Draft a requirements document for aggregators that describes information and services that are crucial to the success of biodiversity informatics.
- Writing group sessions
 - Discussion of issues for plenary
- Plenary session
 - Discussing issues from the perspective of
 - Providers
 - Users

Draft list of issues for discussion

- Full record-level information discovery and delivery
- Metadata harvesting protocols
- GUID per record with persistence
- Attribution metadata with all data records
- Media information ala Audubon Core
- Bi-directional portal
- Feedback from data users to providers (e.g. data quality)
- Usage analytics
- Attribution to providers from analysis
- Annotation management
- Active repository technology (incremental updates)

Plenary Discussion Summary

- Providers need
 - Attribution for data use
 - Help with managing taxonomy
 - Feedback from determination and data cleaning
 - The effort required to process feedback will be considerable
 - Tools are needed to help providers with feedback
 - Details about determinations and geolocations
 - Help with identifiers
 - Registries for people and localities
- Users need
 - Good global information discovery services
 - Assessments of data quality, per record or dataset
 - Data cleaning services
 - Feedback for data cleaning so that improvements are made by providers
 - Tools to find related data, e.g. sequences
 - Tools to aid in integration of data from multiple sources

Information Integrity and Attribution

- Provenance tracking
 - Keep track of source of information
 - Ensure information is not changed
 - Control versions
- Attribution
 - Keep track of delivery of information
 - Make attribution information available to providers
 - Provide mechanism for users to report
 - Publications
 - Derived data
 - Evaluation and corrections

Identifier and identifier services

- Encourage identifiers for objects
 - Require stable identifiers
 - Provider must commit to consistent use of identifiers
 - Strongly suggest that providers maintain GUIDs
 - Aggregator should add GUIDs as necessary
 - E.g. Don't export a record without occurrence ID
- Identifier services
 - Return metadata document upon request
 - Any known identifier can be used to fetch document
 - Discover and maintain relationships among identifiers
 - E.g. If a provider changes the identifier, the aggregator must record that the old and new identifiers are equivalent

Search and Discovery

- Search by common properties
- Discover across object types
 - E.g. Find image by scientific name or geography of specimen
- Provide for download in common formats
- Provide APIs for search and download

Taxonomic services

- Assumption
 - Provider sends scientific name and possibly higher classification
- Externalize taxonomic names and classifications
 - Participate in shared services
- Allow discovery beyond name string
 - Synonyms
 - Common names
 - Higher taxa

Dealing with extended schema

- Assumption:
 - Providers will have important information that is not Darwin Core
- Properties
 - Keep track of properties and evaluate new data sets for new properties
 - Allow both literal- and resource-valued (relationship)
- Transformations and normal forms
 - Maintain information content when changing formats
 - Transform or coalesce properties according to community standards
- extending schemas
 - traits, measurements, interactions
- property similarity with respect to discovery

Thanks to all participants

- Writing group
 - Greg Riccardi (iDigBio)
 - Reed Beaman (iDigBio)
 - Donald Hobern (GBIF)
 - Rich Pyle (GNA, Bishop)
 - Robert Whitton (GNA, Bishop)
 - Paul Flemons (Biodiversity Info Manager, Au)
 - James Macklin (Biodiversity Info Manager, Can)
- Plenary Group
 - Joanna McCaffrey (iDigBio)
 - Deb Paul (iDigBio)
 - Neil Evenhuis (GNA, Bishop)
 - Shelley James (Bishop, Macro Algae)
 - Michael Thomas (Biodiversity Informatics Manager, U. Hawaii)
 - Chris Neefus (Macro Algae)
 - Matt Goodale (data management/IT supervisor, NTBG)
 - Tom Schils (Biodiversity Informatics Manager, Botany, phycology, UOG)
 - Aubrey Moore (Biodiversity Informatics Manager, Entomology, UOG)
 - Ryan Caesar (IT Manager/programmer, Entomology, U. Hawaii)
- Others?