Enabling the TCNs and Collaborators
Breakout Group #1: Geo-Referencing

Facilitator Name: David Bloom

Scribe Name: Jesse Grosso

Time Allotted: 150 minutes

Group Participant List: Zack Murrell, Edward Gilbert, Jeffrey Holland, Randall "Toby" Schuh, Nelson Rios, Austin Mast, Andréa Matsunaga, Bruce MacFadden, Reed Beaman, Matthew Collins

Objectives:
Discuss and produce a report to summarize geo-referencing needs within the ADBC community. Focus on opportunities to leverage existing tools/systems, standards, practices and techniques. Nominate a reporter to deliver a 15-minute summary report to the plenary session at the conclusion of your session.

Deliverables:

1. Define and order <u>at least five</u> critical challenges faced by the TCNs related to geo-referencing of specimen data (#1 is the most critical challenge).

Rank Order	Challenges Related to Geo-Referencing of Specimen Data
1	**High Prioroity** -Resources are needed to train groups of people to geo-reference (instructional videos, online modules, workshops, etc); Make sure new collections/people know where to go to find the knowledge - basic GIS course, what programs do you start with? Is there conflicting info? Lack of information in regards to Geo-referencing; aggregation of training required.
2	**High Priority** - Advocating Best Practices. Collectors need to have good field practices. A best practices document for geo-referencing is needed.

3	People overriding other people's geo-referenced data therefore replacing initial data.
	
	Remedies:
	    - Treat changes as annotations
	    - Keep a history of all changes
	    - Justification to override geo-referenced data is needed (i.e., supporting validation for the recommended geo-reference revision)
	    - Data manager sets one annotation as the current geo-reference

4	Include Stratigraphic data, particularly for paleo specimen data.

5	Train people to use GPS to acquire data for reference, projection, Lat./Longitudinal points, spheroid data (need to include spatial reference system). Record more than just the GPS data. Information systems need to have a place to record spatial reference system.
6	Concerns and issues of masking data; sensitive data Remedies: • Get everyone who is involved with "sensitive" data to agree to keep it masked (Heritage program, US Fish & Wildlife Service, etc.) • GBIF did workshop on sensitive data on challenges, etc. to do it properly; there is a document available with results on workshop • Need to make this and other documents "more visible" • Social science side of sensitive data
7	Need to provide geo-reference data where appropriate; features beyond lat./longitudinal data, ex: polygon used to define a lake
8	Interpretation of legacy localities - data that is on a collection label prior to GPS (prior to mid 1990s); older data

2. Identify and order <u>up to five</u> existing practices and techniques that can be leveraged for geo-referencing (#1 is the most preferred practice/technique). If more than five, focus on the five that are currently the most viable, commonplace, and applicable to the needs of the TCNs and collaborators, while keeping a list of all references to existing practices.

Rank Order	Geo-Referencing Practices and Techniques
1	Collaborative geo-referencing
2	Dividing work up by area of expertise within domains (geographical) • Could benefit from a collaborator tool for researchers who are geo-referencing. • Extend existing tools to make them more useful for the cloud.
3	Geo-referencing by use of a map
4	Involving local communities in geo-referencing areas
5	Uncertainty Calculations/radius – this is a manual process supported by a published paper - Google "Uncertainty Radius"
6	Automated geo-referencing should always be verified by a human. Can the HUB host a solution that leverages existing geo-referencing data from other collections?
7	Utilize Google Maps

3. Identify and order <u>up to five</u> existing standards that can be leveraged for geo-referencing. If more than five, focus on the five that are currently the most viable, commonplace, and applicable to the needs of the TCNs and collaborators. Explain the choices.

Rank Order	Geo-Referencing Standards	Explanation of Selections
1	Best practices document from GBIF	start with this document because techniques are standard (even though it is outdated - does not include polygon data)
2	Darwin Core	Specifications to communicate geo-referenced data
3	GML - Geographic Mark-up Language	"over the top" and may be hard to get people to use it
4	European Petroleum Survey Group (EPSG) Codes	refer to codes
5	MaNIS/HerpNet/ORNIS Georeferencing Guidelines	available online at VertNet and possibly HerpNet
Other	Ability to represent data points: how to represent other data for example, county, township	polygons can be used at a minimum; polygon to be used when there in uncertainty in locality = uncertainty polygon
Other	- different features have different sizes - storage data - GMC - "Bulky" - implementation detail - GeoJason - good for web-based information	

4. Identify and order <u>up to five</u> existing tools/systems that can be leveraged for geo-referencing (#1 is the most preferred tool/system). If more than five are proposed, focus on the five that are currently the most viable and beneficial to the greatest number of stakeholders. Explain the choices. Link tools/systems to the practices/techniques (identified in Deliverable #2) and standards (identified in Deliverable #3) that each enables or supports.

Rank Order	Geo-Referencing Tools	Explanation of Selections	Linked Practices/ Techniques (Line Numbers)	Linked Standards (Line Numbers)
1	Geolocate	generic georeferencing systems that people use		
2	BioGeomancer	"		
3	Google Maps	"		
4	Specify	"		
5	Arctos	database for Natural History Data - Oracle; use geolocate as plug-in, google maps		
Other	HerpNet Site; listing of other resources	**GIS/Georef Resources** link on the website that lists many resources; listing of other resources		
Other	USGS Name; listing of other resources	sources for other tools		
Other	Atlas of Living Australia (ALA); listing of other resources (3-4 years old)	Summary of tools for digitization, resource list with evaluations and comments.		
Other	Life Mapper	a manual tool		
Other	Investigate what NEON is using			

5. Define specific gaps that exist within each of the identified top 5 tools/systems (e.g., functionality problems, scalability limitations, availability, licensing issues, cost, lack of standard usage, missing features).

Rank Order	Geo-Referencing Tool (list 1-5 from table above)	Gaps, Issues and Opportunities for Improvement
1	Geolocate; BioGeomancer	improve geo-referencing to be more intelligent in producing data
2		Marine data is difficult
3		support for other languages
4		lots of redundant geo-referencing being done to assure quality
5		data can be given in 2 different pieces ; format will be different; programmer will be needed
6		Geolocate contains polygons for National Parks, Forests, Lakes and Cities, but no counties yet.
7		Information is not always readily available (can be restricted by country)
8		Lack of built-in crowdsourcing tools or capabilities
9		Biogeomancer : not really collaborative, but "batch"
10	Arctos	Licensing issues (there may be costs associated with this program)
11	Specify	Has the same gaps as geolocate: • cannot plug-in to collaborative system • integrate collaborative tools

6. Identify the critical implementation date for HUB appliances that would enable/enhance geo-referencing based upon TCN project plans. Explain why this date is critical.

Critical Implementation Date	Explanation
ASAP	The Hub needs to start training workshops very soon because of the start dates of the TCNs of geo-referencing (InvertNet - April 2012; Tritrophic - Summer 2012; Lichen Bryophyte - in about a year)

7. Produce documentation related to the development/implementation of a geo-referencing appliance to serve the needs of the ADBC community.

Functional Requirements:	
Specific items the HUB needs to deliver to enable/enhance geo-referencing:	A work flow for geo-referencing. Workshops for training of general geo-referencing. Something needs to be built for "crowd sourcing, volunteerism, citizen science" - because nothing exists
Specific items the TCNs needs to deliver to enable/enhance geo-referencing:	

Input from Matt Collins:
Option 1: Desktop version of geolocate on virtual machine, excel spreadsheet, …
Option 2: Local installation geolocate web tool with bundled map layers and give back data
Option 3: Web Service that takes location information and gives back multiple answers

The HUB is needed to develop a work flow tool for geo-referencing. (ex: Kepler) - would need appropriate staff to do so.

8. Other notes, comments and details not captured elsewhere.

Comment made: Maybe the HUB can build onto "georeferencing.org" (owned by VertNet) and provide a site of central information.

Question asked: Can the HUB make GEOlocate a software appliance?

Question posed: Is there money to build tools?
Answer: iDigBio is not funded to build new tools. However, iDigBio can leverage resources to bring developers together for improvements to existing tools, conduct grant-writing workshops, and fund workshops and working groups to build technical requirements for other organizations that are funded to build tools.

Question posed: Can HUB address where people go to get these geo-referencing resources?
Answer: Yes, the HUB can host instructional materials, standards, documents, etc in the Wiki (particularly for items that involve ongoing discussion, input and development). Codified standards and finalized training materials can be posted on the primary www.idigbio.org website.

Scope of field in databases - there is not a huge difference in "field" within the continental U.S.

- Should there be a field that represents the maximum amount of variation at "x" latitude?
- Tools - uncertainty calculator (Geolocate, Biogeomancer) have ways to accommodate "unknown"
- Variation in World? - error in datum

Crowdsourcing:
- Question posed: Can iDigBio push across domains? Would fish people be comfortable with plant people's prior geo-referencing results? This would reduce duplication effort - if you go across domains you can decrease effort. Perhaps a geographical filter can be put in place: aquatic vs. terrestrial
- iDigBio drives crowdsourcing with a group of experts, but how do you organize the crowd? Stratified sampling within the expert community could be utilized to validate crowdsourcing results.
- Crowdsourcing involves people with no particular specialized training (K-12; highschool students; citizen scientists)
- The information is usually checked by someone with more expertise or validated with multiple checks on the same item.
- ADBC Community needs to have a workshop on volunteerism, crowd sourcing and citizen science and make it a high priority.
  - o FSU could possibly host a workshop in regards to this.
  - o Can have TCNs take the lead
  - o Suggested that contact be made with Tom Nash to see if he is interested in developing a workshop to bring in outside expertise

Collaborative Sourcing

- special training required
- data usually is not being checked

Suggestion: iDigBio facilitates workshop - expert on crowd sourcing in general within ADBC community (Astronomy, Google.....)

Final comments - HUB needs to start training workshops as soon as possible on the following topics:

- general geo-referencing: biogeomancer, geolocate
- start with paper maps
- need to communicate