

Chris Norris

Redaction of Locality Data

Point 1: This conversation, although it obviously has much broader implications, is primarily to do with iDigBio.

ADBC is an NSF-funded program – i.e. funds from ADBC cannot be used to digitize collections belonging to other Federal agencies.

However... although NSF funds iDigBio, it is intended to serve data from all U.S. biocollections.

We do not want to get into a position where collections choose not to share any of their data via iDigBio because of concerns over data security.

So establishing a common position would be beneficial.

Point 2: This is a conversation to do with data where we do not have pre-existing obligations with respect to data security.

In some cases, we are restricted by the terms of our collecting permits from releasing data without third party permission (agency or private landowner).

In other cases, our status as a Federal repository means that we do not actually “own” the data in question – so it is not our decision whether we release it or not.

Point 3: I’m concentrating on issues relating to locality data, but there are other, important categories of information where redaction is an issue.

It is possible that institutions may wish to restrict access to data that are the subject of ongoing research.

It is common practice to restrict access to information about donors, personal information relating to individuals (address, email, certain pieces of correspondence, etc), specimen loans, and internal security (physical location of objects within the institution).

These could potentially be of use to researchers, but are currently beyond the scope of our discussions.

Point 4: There are no “right” answers here, and no-one died and made me (or Gil, or iDigBio, SPNHC, SVP, etc) God.

At the end of the day, institutions and agencies have considerable leeway to decide what data they choose to share.

“10. What are NSF’s expectations regarding the release of data that include sensitive information (e.g., information about individuals or locations of endangered species)?

Such data must be maintained and released in accordance with appropriate standards for protecting privacy rights and maintaining the confidentiality of respondents. Within legal constraints, what constitutes reasonable data access will be determined by the community of interest through the process of peer review and program management.”

I believe that the less they choose to share, the less everyone benefits. So we need to discuss, as a community, what we’re comfortable with.

Where to start?

As I have the conch, I get to start! So, this is what I think...

Start from the position that

- (1) These are collections held in trust for the public
- (2) The purpose of digitization is to fully mobilize the collections data
- (3) Georeference information is one of the most important categories of data
- (4) We should not seek to predefine how these data will be used
 - a. To explain further – we want to find novel uses
 - b. Placing restrictions on access may create barriers
 - c. Especially likely for those unfamiliar with data
 - d. We may not be able to define what is or isn’t appropriate restriction
 - e. i.e. what people will “need” for their research
- (5) When we start seeing our role as restricting access to collections and collections data, I believe we have failed in our duty as collections professionals

Given this, why should we even consider redacting specimen data?

- (1) Certain resources are both valuable and static
- (2) Collections data can reveal where you are likely to find these resources
- (3) This can lead to illegal collecting of materials
- (4) When this happens, the material is no longer available directly or indirectly to the wider public
- (5) So it could be argued that if we do not keep the data secure, we are failing in our duty of public trust

So we have two conflicting obligations to the public

(1) To preserve specimens and data and make them accessible

Vs

(2) Protect natural resources from illegal collecting

I believe that you can make a strong case that (1) is our core responsibility. I think the case for (2) is much weaker.

You could make the case that (2) is the responsibility of State and Federal land managers, law enforcement, and – if they chose – private landowners.

Obviously we do not want to make their jobs more difficult. The question is how far we're prepared to compromise (1) – our core function – to do so.

I find it hard to answer this question in the absence of information. What I'd need to know in weighing this issue is:

- (1) What information are we talking about, and what do we mean by “redaction” – serving records with certain information removed, “fuzzing” information (e.g. point-radius, T/R, or to County only), or not serving the record at all.
- (2) To what extent is the information that we are seeking to redact available from other sources (publications, conference presentations, local knowledge, etc)?
- (3) How many cases of fossil poaching have there been where people have used collections data to locate a site?
- (4) For large vertebrates (my guess for the most vulnerable category of material) how useful are data from specimens that have already been collected?