

2017-10-17 Bioinformatics Call Notes

Bobbi Pierson, geneticist at Alaska Science Center, presenting on new data release standards and how her office is handling them. She has a Genetics Metadata Working Group making efforts to standardize and streamline the data release process for genetics data. Bobbi's contact info: bpierson@usgs.gov, mobile # 907-782-8189, office # 907-786-7107

- Bobbi conducted a 2017 baseline survey with 10 questions which received 10 responses, which she went over some of to reflect how scientists felt about data releases (areas of confusion, ease of use, redundancy, etc.)
- A survey found that the genetic data releases in ScienceBase were among the most accessed of the Alaska Science Center data releases, suggesting those data are being discovered and used.

- NCBI is considered a trusted repository for USGS genetic data but alone does not meet the FGDC or ISO metadata requirements that USGS must meet. Simply, NCBI submission does not meet our data standard on its own.
- The USGS REQUIRES either FGDC or ISO standards of metadata. ISO is less user friendly, so FGDC is being adopted even though it was originally for geospatial data rather than biological. Fields have been added to accommodate biological data.

- There is something called Metadata Wizard in beta testing on ScienceBase (version 2.0). It is being developed out of the Fort Collins Center and is another interface that creates metadata records. It was built in ArcGIS but can now handle non-geospatial formats.

- JC Nelson has been working on eDNA data releases and sees overlap with the genetics data releases. The eDNA group already have data dictionaries and templates. (Bobbi also has eDNA templates. They will talk offline and compare notes).
- The eDNA group is pushing to have NCBI as a repository, without the need for separate USGS data release, but require an FGDC metadata record so USGS data contribution to NCBI exists. The FGDC worksheet should reference the NCBI bioproject and accession numbers. Targeting end of this calendar year to have a working model.
- JC says we don't need a website data release if we have NCBI submission + FGDC...and he's working on explaining this to ALL the BAOs. JC sits on a committee with the BAOs and will bring up NGS datasets since on this call we had three different answers to what various BAOs are requiring for data releases.

- Note that XML is machine readable whereas PDF files are not.

- Note that there is a USGS requirement to release ALL data after a project is over, not just what is used in journal articles. Not currently being enforced because it is unclear where the data would be deposited.
- Bobbi emphasized the requirement for quality documentation of what has been released in journal articles.