

Institution Codes, Collection Codes, and Specimen Identifiers: linking voucher specimen to Tissues, DNA Samples, and Sequence Data.

Dr Daniel Mulcahy¹

¹*Museum für Naturkunde, Leibniz-Institute for Evolution and Biodiversity Science, Berlin, Germany*

ID2) Identifiers and labels in natural history collections: new technologies, challenges and opportunities for linking objects and data, Lecture Theatre 1, Appleton Tower, 11 Crichton Street, EH8 9LE, June 7, 2022, 4:00 PM - 5:30 PM

Nearly all disciplines of biology now have some form of molecular genetic analyses incorporated into areas of their research, from systematics, ecology, and behavior, to physiology and conservation. In order for science to be transparent and verifiable, the source and provenance of the genetic material used must be easily identifiable and traceable. Natural history collections are ever-increasingly facilitating the use of genetic components from collection objects, and in some cases increasing the number and types of collection objects under their care (i.e. tissue/DNA-only, e.g., blood, feather, skin- fin-clips, environmental samples, etc.). Most natural history collections are now making their holdings available online, either on their own platforms or via aggregate search engines like the Global Biodiversity Information Facility (GBIF) and the Global Genome Biodiversity Network (GGBN). Genetic resources are often sought out by researchers external to the holding natural history collection, and molecular sequence data are then generated and deposited in public repositories, such as GenBank and the Barcode of Life Database (BOLD). Making genomic material digitally discoverable to researchers, and linking published data to voucher specimens (and/or genomic samples) is an increasing challenge for natural history collections. As natural history collections digital management systems, digital identifiers (DOIs, URIs, etc.) are assigned to objects in collections, including multiple objects derived from the same individual organism (e.g., voucher specimen and genomic samples), which need to receive different digital identifiers to be uniquely identifiable in digital management systems. The National Center for Biotechnology Information (NCBI), which hosts GenBank, has created a BioCollections Database to curate metadata for natural history collections and link sequence data to voucher specimens. Institution and collection codes are linked to create a “structured voucher” annotation (following the Darwin Core Triplet) to standardize usage across interconnected databases (e.g., GenBank, European Nucleotide Archive, and the DNA Databank of Japan). In this discussion, these issues are presented and recommendations are given for natural history collections to make their genomic resources discoverable and accessible in aggregate databases (such as GGBN) and how to instruct borrowers of their genomic materials on proper publication and citation using institution codes, collection codes, and digital identifiers.

