

Building a Digital Extended Specimen One Association at a Time: What Does It Take to Extend OBI Herbarium Records with their Associated GenBank Sequences?

Jorrit H. Poelen Jenn Yost Katelin Pearson

2023-09-11

Abstract

Specimen from Natural History Collections are physical repositories of genetic information. Genetic sequences extracted from specimen are stored in genetic sequence databases like the openly accessible GenBank at NCBI, DNA DataBank of Japan, or the European Nucleotide Archive (ENA). While researchers and collection managers make efforts to associate (or link) Natural History Collection records with their derived genetic accession records, extra work is need to make these associations explicit. We describe how a collaboration between a biodiversity informatics expert and collection managers of the Hoover/OBI Herbarium at CalPoly, San Luis Obispo, CA was forged with the aim to extend OBI specimen records to include their associated GenBank records. In addition, we quantify the costs of creating these specimen extensions, and discuss the socio-economic capacity needed to repeat this digital specimen extension process for the hundreds of millions of specimen records available globally today.

Q1. How can we estimate the methods needed to support the discovery and maintenance of Specimen-GenBank records such that it can be sustained by those valueing the availability of Specimen-GenBank records links?

Q2. Do the current funding mechanisms allow for rapid development of ideas to address immediate needs in the biodiversity informatics community?