Institutional and Collaborative Work Perspectives on Specimen Databases

James Beach ‡

‡ University of Kansas, Lawrence, United States of America

Corresponding author: James Beach (beach@ku.edu)

Abstract

The U.S. National Science Foundation (NSF) funded a grand experiment on the U.S. biological collections community, although it may not have anticipated the significance of the results. For over 30 years, the NSF made recurring investments through competitive grants in software engineering and technical support for biological collections databases. The Specify Project (now the Specify Collections Consortium), and its predecessor the MUSE Project, which was first funded in 1987, represent a lineage of sustained NSF investment in biological collections database systems. Specify is largely scoped for institutional curatorial, collections management, and data publishing functions, and it is generally deployed collection-by-collection within research institutions reflecting traditional administrative and disciplinary boundaries. MUSE and Specify grew out of a need U.S. collections institutions had for a common data model, a source of ongoing technical support, and desktop applications for the activities associated with collections management, e.g., tracking loans, accessions, gifts, and for printing labels and reports.

In 2011, NSF announced its first "Thematic Collections Network" (TCN) awards from the "Advancing Digitization of Biodiversity Collections" (ADBC) program. TCN projects are also focused on the computerization of collections data but from a different perspective, that is the digitization of specimens organized around a particular taxonomic group and/or research theme. The NSF TCN projects use Symbiota database software. Symbiota is designed for collaborative digitization workgroups across institution and collection boundaries. Symbiota fits the model advanced by the ADBC Program's TCN awards

partially because it mirrors the way museum and herbarium scientists collaborate professionally—along taxonomic lines—in order to share common research interests and expertise, and to organize projects around their focal group of organisms. Thus after 10 years of successive rounds of TCN project funding, millions of specimen data records have been digitized for the first time in Symbiota databases, however, a considerable percentage do not exist in institutional collection management platforms.

Today in the U.S., we have a duality of approaches toward specimen data computerization, collaborative workgroup databases at one pole and institutional collection database management systems at the other (while the Arctos Project sits somewhere in the middle). Tens of millions of species occurrence records are now online from NSF grant funding for the TCNs and for Specify databases over the years, but the dynamics and constraints of the duality that exists today between these two perspectives for digitizing and publishing collections data, initially may not have been so obvious. The duality reflects the partially-distinct and partially-overlapping goals of their stakeholders—institutional collection owners on the one hand, and "thematic" research users of specimen data on the other.

Both methods of specimen data organization and the stakeholders associated with them are critical to the long-term engagement and sustainability not only of the data and their software platforms, but also of the biological collections themselves. This duality distinguishing institutional perspectives on collections data from thematic research perspectives has become apparent because of NSF's support of both types of computing, but their existence is due to more than grant funding patterns. They represent parallel computational perspectives that biodiversity data community architecture has yet to fully reconcile.

It will be critical to bridge these two worlds of specimen data practice and recognize the strengths and essential nature of both paradigms. This presentation will discuss the characteristics of these two modes of specimen digitization and why they should both be fundamental components of digital specimen architecture planning. Although NSF funding revealed and accentuated the differences between these two ways of processing species occurrence data, the duality is productive and permanent.

Keywords

Specify, Symbiota, architecture, data integration, interoperability, software

Presenting author

James Beach

Presented at

TDWG 2020