



Building a Unified Discovery System for Libraries and Museums

Hanna Bertoldi, Jeremy Friesen, and Victoria Perdomo's presentation at ARCS 2019 on Friday, November 8th was a comprehensive explanation of an initiative at Notre Dame University funded by a Mellon grant that seeks to build a platform through which multiple university collections can be viewed, browsed, and searched. This project, the end product of which is called MARBLE (Museums, Archives, Rare Books, and Libraries Exploration Platform), specifically focuses on collections held by the Hesburgh Libraries and the Snite Museum of Art, which over the course of their 100 year histories had become siloed within departments and schools, and disconnected from one another in any searchable way. The project seeks to address a common question posed to the institution by various students and researchers: "Why isn't there one spot to view these collections?" In 2017, Notre Dame applied for a Mellon grant with the intention of solving this problem, and were awarded the grant in December of that year. At the time of presentation, Notre Dame is about a year and a half into the three-year project.

Various grant-funded and existing staff are involved in and integral to this project, and the presentation describes these key team members accordingly. These include the Outreach Specialist, Abigail Shelton, the Collections Database Coordinator at the Snite Museum, Hanna Bertoldi, a Software engineer, Jeremy Friesen, along with a metadata team, student help, and others. Bertoldi noted that about half of those involved were from the Hesburgh Libraries and the other half were from the Snite. This was crucial in achieving the level of collaboration needed for this project to succeed.

Hanna Bertoldi explained the workflow and methodology from the museum database side of the project, and detailed what was accomplished on her end in the beginning stages of the grant, primarily writing a style guide, inventory and data cleanup. The style guide, created for the Snite's EmbARK collections database, was crucial to the beginning stages of database cleanup, and described the formatting and purpose of every field in the database; it is also an ever-evolving document and one that is amended and improved every single day. Bertoldi notes that existing information about best practices is scarce, and much of what makes up her style guide was guided by the work of Emily Nedell Tuck at the Museum of Fine Arts, Houston. In data cleanup and inventory, a lesson learned was that in order to clean up a digital space, one needs to also clean up a physical space. This meant an assessment and an organization project within the object files of the Snite Museum that involved refiling a large portion of deeds of gift and loan agreements that had been filed separately. The inventorying process, also integral to data cleanup, is ongoing and occurs every Friday with two staff members and a student working together in the collections vault to verify collection information.

Bertoldi also leads the metadata team, consisting of a metadata librarian and two software developers from Hesburgh Libraries. As part of their work, they have created an extensive metadata map (which Bertoldi displayed on screen) that maps fields to the different sources being used, rules, definitions, and then maps these fields to DublinCore and when necessary,



VRA Core, as well as schema.org to ensure that these objects will be searchable within Google. A key takeaway was that these fields all mapped well together even though they have different terms and are from different source systems; this was crucial to the collections being widely searchable across various formats.

Bertoldi also details what the project seeks to complete in year two of the grant, which includes solidifying a comprehensive workflow detailing data remediation that emphasizes the collaboration necessary between project stakeholders to complete data cleanup - it cannot be done alone. Year two goals also include expanding rights metadata along with keywording subject terms in order to make collections browsable.

The presentation then pivoted to a discussion of the technology needed to make MARBLE work, which was headed by software engineer Jeremy Friesen. Friesen described the metadata map created by the metadata team as “the one spreadsheet to rule them all,” emphasizing its importance in channeling information from many different source systems into one output. In order to illustrate this process, Friesen uses an analogy of a floor lamp with three legs, the legs representing different data sources. In this situation, our sources of data come from a library/rare book catalog, a museum collection, and archive collection, and are all very different; Friesen presents them as acting like international plugs that require adapters, and his work as creating these adapters for different data sources to merge into one common output. An example of this common output is the schema.org standards that are read by the internet easily and consistently. The middle post of the lamp is representative of the “manifest pipeline” where this merging takes place, and where IIIF (the International Image Interoperability Framework) is implemented, a process that formats data to allow rich, zoomed-in viewing of images in a system of tiles not unlike the way Google Maps operates. This data is then loaded into a search index, and wrapped in the “IIIF wrapper”. To continue the lamp analogy, the lightbulb is the space where the data is viewed and interacted with, and the filament is where the light comes from directly; this, like the original metadata should not be touched or altered, but the viewer is able to see and interact with this data (or “light”) through the wrapping of the lightbulb. Custom sites are built with a program called GatsbyJS, which allows for the standardized data to be accessible from a static site that includes a dynamic search function.

As a summary, Friesen pulled up a slide of what might be described as a Frankenstein lamp - another version of his analogy, but this time with several three-legged source stands, and several kinds of lampshades at the top. This slide illustrates that data sources and websites will likely be always changing as the needs of the university and its institutions evolve and as technology improves, but that middle post, the single “manifest pipeline,” should be altered as little as possible to allow for flexibility in source input and data output along with consistency across the system. This discussion concluded the presentation, after which a question was asked about how the presenters go about addressing subject headers and tags. The presenters response points to the mapping required to connect fields across library and museum environments; while recognized as “subject terms” in the library’s source system, this field is



referred to as “keywords” in EmbARK, and needed to be mapped together as such. Bertoldi added that most terms within EmbARK will be from Getty Thesauri, and Perdomo mentioned that this will be the focus of the next year or so of the grant’s work, and details plans to work with curatorial and faculty staff to adapt a set of words that work for all collections. Overall, the session was informative and detailed, and succinctly illustrated the data-level and technological work that goes into publishing collections online in a user-friendly and effective way.

Submitted by Madeleine Wieand, Collections Database Associate, New Orleans Museum of Art