Digital Cross Collection Search: using fedora disseminators to expose heterogeneous content in Blacklight

Hui Zhang

Wells Library, Indiana University Bloomington

While many collections of digital assets stored in Fedora benefit from specialized content models there is sometimes a need to provide an easy but powerful interface to search or browse all materials regardless of type. This presentation will describe the approach taken by Indiana University's Digital Library Program to create a cross collection search interface based around Fedora's disseminator framework using Solr and Blacklight.

Our discussion of the updates to our Fedora content models will begin with the process of selecting a set of metadata fields to serve as the "lowest common denominator" between the various objects to be included in this new cross collection search. From there, we'll show how utilizing a strategy of building disseminators to resolve complex determinations about objects simplifies customization of external tools and applications while making the repository more self-documented. We'll go over how we designed and implemented the focused Fedora Service Definition Objects needed to drive a cross-collection search and show how with very few changes, one can expose normalized access to a repository with even large amounts of dramatically varied content.

With the repository doing the work (through disseminators) of metadata and object normalization the indexing routine can be simple. From an item-centric point of view, the indexing algorithm will first find its collection with resource index lookup, then extract status for the collection and item using disseminators, finally generate and index the item's MODS record if the status indicates the item is ready for public view. Therefore, with normalized objects and newly developed Fedora service, it is not essential to hardcode certain metadata field (e.g., collection name, status) in the indexing routine for purposes such as quality control. After all, it makes the indexing routine requires little change over time, even as new collections with formerly unimagined content models are added.

Our presentation will conclude with a detailed discussion of the implications of our metadata choices on the customization of Blacklight. We are facing several new challenges in building a discovery interface for digital collections using Blacklight: the fields in Solr index are different to traditional library catalog (MARC records describing book items) because they are generated from MODS records used to describe multimedia items; certain fields in our index (e.g., collection name, source) are flexible and their values won't be determined until late in development; rendering thumbnails in the index view. Furthermore, we want item view to be its content-specific contextualized view developed at IU instead of using the rendering provided by Balcklight. In other words, Blacklight was only needed for discovery in this new system.