Presentation Proposal for Code4lib Conference 2014

# Leveraging XSD's for Reflective, Live Dataset Support in Institutional Repositories

Mark Sullivan, Library Information Technology, University of Florida

The University of Florida Libraries are currently adding support for active datasets into our METS-based institutional repository software, SobekCM. This ongoing project enables the library to be a partner in current, or long-running, data-driven projects around the university by providing tangible short-term and long-term benefits to the projects. The system assists project teams by storing and providing access to their data, while supporting online filtering and sorting of the data, custom queries, and adding and editing of the data by authorized users. We are also exploring simple data visualizations to allow users to perform basic graphical and geographic queries. Several different schemas were explored including DDI and EML, but ultimately the streamlined approach of using XSD's with some custom attributes was chosen, with all other data residing in the METS file portions. Currently the system is being developed using XSD's describing XML datasets, but this model should easily scale to support SQL datasets or large datasets supported by Hadoop or iRODS.

This work is being integrated in the open source [SobekCM Digital Content Management System](http://sobek.ufl.edu/)[[1]](#footnote-1) which is built on a pair-tree structure of METS resources with [rich metadata support](http://ufdc.ufl.edu/design/webcontent/sobekcm/SobekCM_Resource_Object.pdf) including DC, MODS, MARC, VRACore, DarwinCore, IEE-LOM, GML/KML, schema.org microdata, and many other standard schemas. The system has emphasized online, distributed creation and maintenance of resources including geo-placement and geographic searching of resources, building structure maps (table of contents) visually online, and a broad suite of curator tools.

This work is presented as a model which could be implemented in other systems as well. We will demonstrate current support and discuss our upcoming roadmap to provide complete support.

1. <http://sobek.ufl.edu> [↑](#footnote-ref-1)