
Background:
In the summer of 2012, Texas A&M University Libraries uploaded more than 16,000 retrospectively-digitized masters-level theses, dating from 1922 to 2004, into our DSpace institutional repository. Item records for the Retrospective Theses collection were created by mapping existing MARC records, then transforming and enhancing this metadata. Records included fields encoded in our Qualified Dublin Core schema, as well as the custom Thesis schema developed by the TDL member consortium. MODS metadata records were also generated, to be stored as bitstreams.

More than eight decades of MARC cataloging preceded the transformation and enhancement of metadata for these in cataloging practices, including:
• Moving away from full subject analysis, owing to staffing shortages and the novelty of certain subjects addressed in student works.
• The consistent use of a local call number system rather than Library of Congress classification numbers.
• Experimentation with technology to add abstracts to bibliographic records.
• Fields used in the catalog but not submitted to OCLC.

Crosswalking MARC TD Metadata to DC and MODS Metadata:
Repository-extractable metadata were generated by means of three XSL transforms applied to MARC-XML from the catalog. MARC records were extracted on the basis of call-number and subjected to these transforms to produce Dublin-Core style metadata in the dc and thesis schemas as well as a MODS record. These metadata files were packaged in the DSpace Simple Archive Format (SAF) for ingestion.

Improvements in the works:
• Removing potentially inaccurate “department” values (global collection wipe of thesis degree department).
• Adding doctype values that will facilitate featured searches in Primo.
• Investigating normalizing “major subject” terms, found in four places in DSpace metadata.
• Implementing any edits in MODS. Pie in the sky:
• Adding committee members and other metadata found in the documents themselves.
• Mapping out departmental evolution and assigning accurate department values.

Continuing challenges:
• Inconsistent (and sometimes inaccurate) metadata.
• Lack of controlled or normalized values for metadata values, including disciplines and majors.
• Additional burden of adding MODS when editing DC and Thesis DSpace metadata.
• Lessons (learned): lack of consistent metadata creates additional usability issues.

Over eight decades, though, we see shifts in A&M’s MARC cataloging, as well as a new body of metadata. Over eight decades, though, we see shifts in A&M’s MARC cataloging, as well as a new body of metadata.

More than eight decades of MARC cataloging preceded the transformation and enhancement of metadata for these in cataloging practices, including:
• Moving away from full subject analysis, owing to staffing shortages and the novelty of certain subjects addressed in student works.
• The consistent use of a local call number system rather than Library of Congress classification numbers.
• Experimentation with technology to add abstracts to bibliographic records.
• Fields used in the catalog but not submitted to OCLC.

Crosswalking MARC TD Metadata to DC and MODS Metadata:
Repository-extractable metadata were generated by means of three XSL transforms applied to MARC-XML from the catalog. MARC records were extracted on the basis of call-number and subjected to these transforms to produce Dublin-Core style metadata in the dc and thesis schemas as well as a MODS record. These metadata files were packaged in the DSpace Simple Archive Format (SAF) for ingestion.

Improvements in the works:
• Removing potentially inaccurate “department” values (global collection wipe of thesis degree department).
• Adding doctype values that will facilitate featured searches in Primo.
• Investigating normalizing “major subject” terms, found in four places in DSpace metadata.
• Implementing any edits in MODS. Pie in the sky:
• Adding committee members and other metadata found in the documents themselves.
• Mapping out departmental evolution and assigning accurate department values.

Continuing challenges:
• Inconsistent (and sometimes inaccurate) metadata.
• Lack of controlled or normalized values for metadata values, including disciplines and majors.
• Additional burden of adding MODS when editing DC and Thesis DSpace metadata.
• Lessons (learned): lack of consistent metadata creates additional usability issues.