

Uncovering the Mysteries of Metadata Harvesting: Optimizing Digital Library Content for Summon Discovery Tool Access

Kelsey Brett, Resource Discovery Systems Fellow & Santi Thompson, Metadata Coordinator

UNIVERSITY of HOUSTON | LIBRARIES

Motivation for Project

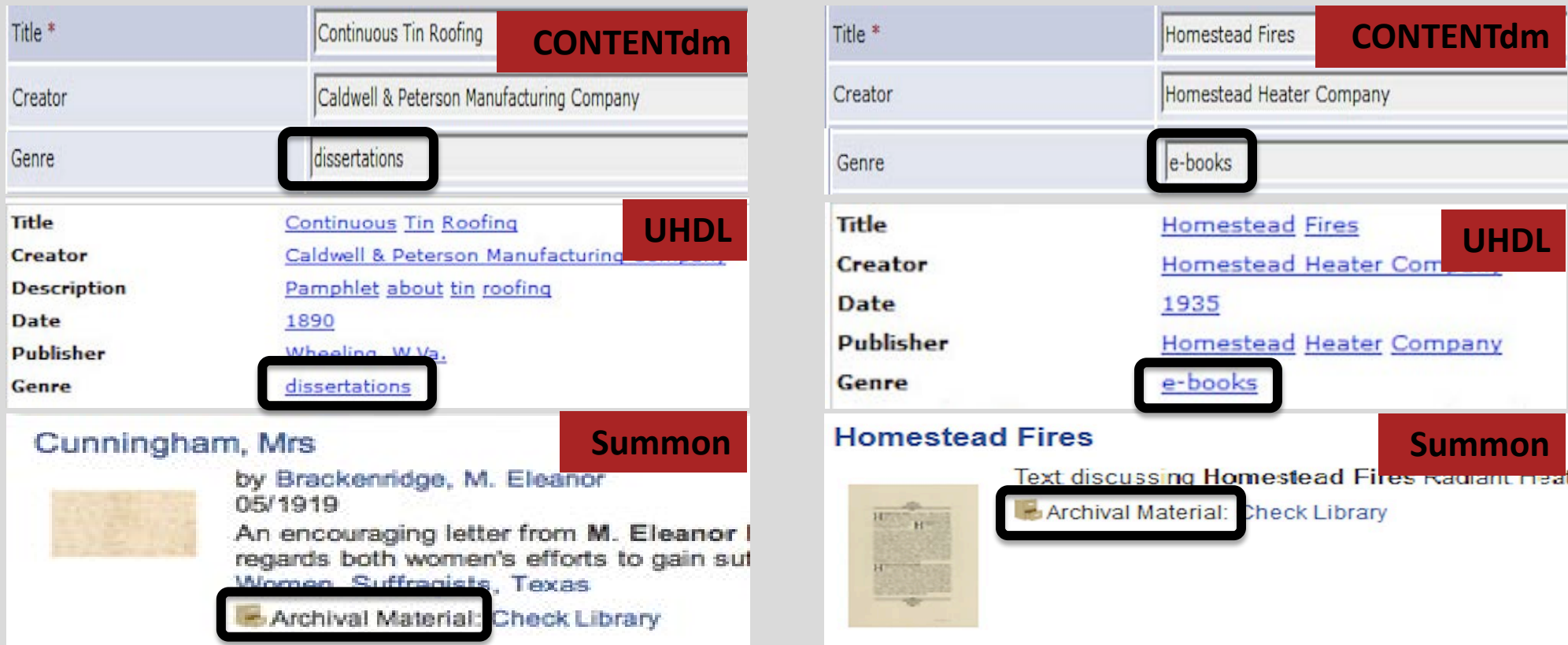
- Part of a larger UH Digital Library (UHDL) metadata upgrade project
- Inconsistent and overly-general descriptions of UHDL content in Summon
- Inability to narrow results to only UHDL materials

What We Knew

- UHDL uses Getty AAT vocabulary for the genre field
- Summon uses its own vocabulary for content type
- CONTENTdm uses Dublin Core (DC) elements to map fields for harvest
- Summon reads the field associated with the dc:type element in a CONTENTdm record to determine the content type in the Summon record
- Nearly all UHDL material is attributed to the Summon “Archival Material” Content Type
- Some UHDL items have a more specific content type in Summon, such as “Photograph”

Testing & Results

- We communicated with Serials Solutions to understand how Summon determines content type
- Meanwhile, we tested our own theory of what was happening during the harvest process
 - We discovered that items with a Summon Content Type “Photograph” contained the word photograph in the genre field of its CONTENTdm record
 - We wanted to know what happens when we used other Summon vocabulary in the CONTENTdm genre field



- We reported our unsuccessful test to Serials Solutions. They provided us with the following information:
 - During the Summon implementation, their metadata librarians “used their best judgment to assign content types based upon the values in the dc:type field”
 - Here is the table they created and use:

Value in dc:type	Summon Mapping
black-and-white photographs	Photograph
color photographs	Photograph
architectural drawings	Architectural Drawing
photographic postcards	Postcard
drawings	Drawing
models	Model
If none of these values, Archival Material	

Challenges

- Working within the constraints of two commercial vendors who do not allow for extensive manipulation
- Serials Solutions was very slow when responding to our inquiries
- No direct communication with the Summon Metadata Librarians
- Waiting for Summon’s monthly harvest to see our test results

Future Plans

- Since Summon uses a table to map the dc:type element to a Summon content type, we will create a more complete mapping table for all of the values used in the UHDL

Other Accomplishments

- We wanted a way for users to limit a Summon search to UHDL items only
- We successfully created a location facet for “UH Digital Library” to do so

