



# Iteration, Not Perfection

## The “Long Game” of Retrospective Implementation of Faceted Vocabularies

IFLA Subject Analysis and Access webinar: “Fascinating Facets”

Thursday, May 19

Casey A. Mullin, Western Washington University, Bellingham, WA, USA



# Introduction: The Changing Landscape of Faceted Vocabularies (in the USA)

- ▶ *Library of Congress Subject Headings* (LCSH) as a legacy, “all-purpose” system
  - ▶ Conflation of purposes: describing what a resource is **about** vs. what it **is**
  - ▶ *Faceted Application of Subject Terminology* (FAST: a derivative of LCSH)
- ▶ *LC Genre/Form Terms for Library and Archival Materials* (LCGFT)
  - ▶ Launched in 2007
  - ▶ Terms added in phases, by broad discipline (through 2017)
  - ▶ MARC 21 fields: 380, 655
  - ▶ Strong alignment with FAST genre/form terms (MARC 21 655)



# Introduction: The Changing Landscape of Faceted Vocabularies (in the USA)

- ▶ *LC Medium of Performance Thesaurus (LCMPT)*
  - ▶ Released in 2014
  - ▶ MARC 21 field: 382
  - ▶ No corresponding FAST vocabulary
- ▶ *LC Demographic Group Terms (LCDGT)*
  - ▶ Released in 2015 (as a pilot)
  - ▶ MARC 21 fields: 385, 386 (audience, creator/contributor characteristics)
  - ▶ No corresponding FAST vocabulary
- ▶ Geographic and chronological facets
  - ▶ MARC 21 fields: 046, 370, 388
  - ▶ Corresponding FAST vocabularies designed for “aboutness” uses (MARC 21 fields 648, 651), though they could potentially be used in other fields



# Infrastructure Needed for Full-Scale Implementation

1. Vocabularies (growing and developing over time)
  - ▶ Co-existence of LCSH, FAST, LC faceted vocabularies, and other vocabularies
2. Content designation (granular MARC fields; other ontologies)
3. Training and best practices for catalogers and metadata creators (“current” or “prospective implementation”)
4. Tools for user discovery (facets, display, indexing)
5. Enhancing legacy metadata (“retrospective implementation”)
  - ▶ Includes batch processes with metadata records new to a system
  - ▶ Iterative, requiring ongoing testing and refinement of methods

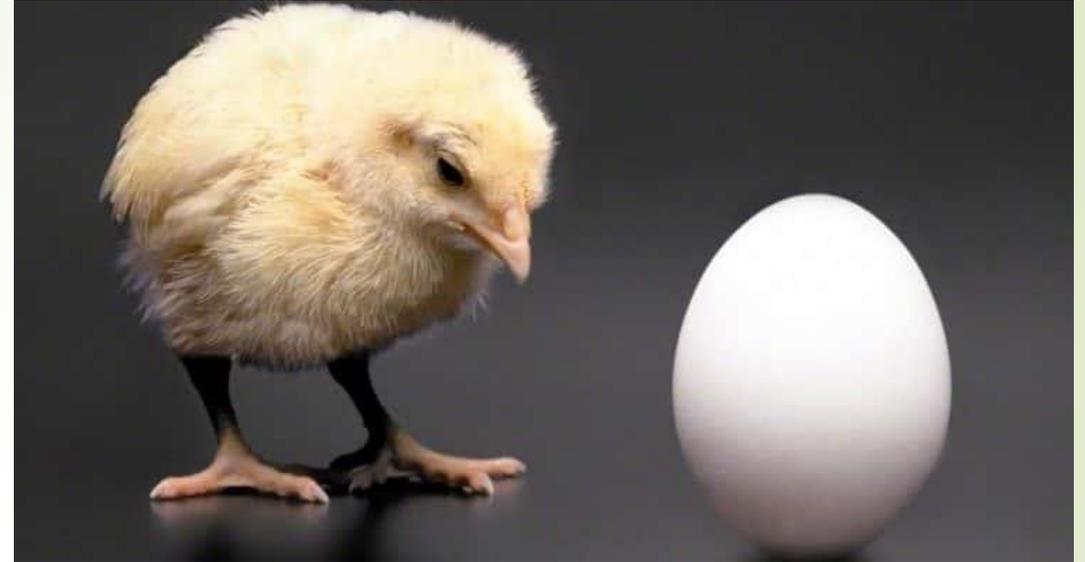


# Full-Scale Implementation: Overarching Goals

- ▶ Increase the prevalence of faceted data in library catalogs over time (in the USA and globally)
  - ▶ FAST is already highly prevalent in OCLC WorldCat
- ▶ Achieve more sophisticated user outcomes through faceted access; this requires a “critical mass” of faceted data in a catalog
- ▶ Pursue sustainable “subject” access (broadly speaking) for metadata workers and end users alike
  - ▶ Faceted vocabularies easier to learn and apply (for metadata workers)
  - ▶ Allow users to refine results set without needing to guess at subject heading syntax in advance
  - ▶ Sharpen the purpose of pre-coordinated subject strings in an alphabetical browse list (no more mixture of “about” and “is” in one place)

# Challenges/Obstacles to Full-Scale Implementation

- ▶ Chicken and egg
- ▶ Turning a ship around
- ▶ No one-size-fits-all solution
  - ▶ Different disciplines merit different approaches
- ▶ Future of legacy systems (e.g., LCSH)?





# ALA Core SSFV

## A Hub of Expertise and Advocacy

- ▶ Charge (from [SSFV website](#))
  - ▶ The Core SAC Subcommittee on Faceted Vocabularies (SSFV) facilitates the implementation and use of faceted vocabularies in library metadata. SSFV accomplishes its goals through **development of best practices** and **training materials** for catalogers/metadata creators, as well as strategies and mechanisms for **retrospective application of faceted terms in legacy metadata**. SSFV collaborates with other standards bodies such as the Library of Congress Policy and Standards Division and the Program for Cooperative Cataloging, along with specialized communities of practice such as those focused on cataloging/metadata for music, audiovisual, law and cartographic resources. Where warranted, SSFV leads or participates in the **development of new faceted vocabularies** and/or the expansion/refinement of existing vocabularies. Lastly, SSFV is concerned with the **optimization of faceted library metadata for user discovery** of library resources, and seeks opportunities to lobby developers of discovery systems for improvements to search, display and indexing thereof.
- ▶ History
  - ▶ Founded in 2017
  - ▶ Prior group: ALCTS SAC Subcommittee on Genre/Form Implementation



# Retrospective Implementation Possibilities and Parameters

- ▶ Sophisticated mapping tables and algorithms
  - ▶ Simple one-to-one mappings (or one-to-many)
  - ▶ Complex mappings (requiring advanced machine reasoning or human review)
- ▶ Computer programs and user tools built on the above
  - ▶ Scripting programs to crawl through an entire database
  - ▶ Macros (used in OCLC Connexion or other utilities)
- ▶ Testing and refinement of the above
- ▶ Role of human review
  - ▶ Inverse relationship between automation potential and the size/diversity of metadata corpus (no “one size fits all”)
- ▶ LC vocabularies are the current focus, but these methodologies can be applied to any destination vocabulary



# Retrospective Implementation Best Practices (issued by SSFV)

- ▶ Published March 25, 2022: <http://hdl.handle.net/11213/17998>
- ▶ Includes
  - ▶ Introduction (high-level considerations)
  - ▶ Modules (completed work and road map for future work)
  - ▶ Bibliography of training resources (current implementation)
  - ▶ Bibliography of relevant literature
- ▶ Ongoing revisions (“Extension plan”)
  - ▶ Summary document will likely be revised 1-2 times per year
  - ▶ Modules (mapping specifications) revised on an ongoing basis
  - ▶ Intention: to incorporate lessons learned from ongoing testing across diverse environments

# RIBP Modules Completed

- ▶ Genre/Form: mapping MARC fixed field codes to LCGFT and LCDGT terms in 655 and 385 fields
  - ▶ Example: 008/30 1 → [655 \\_7 Festschriften. \\$2 lcgft](#)
- ▶ Genre/Form: mapping LCSH form subdivisions (and select topical subdivisions) to LCGFT and LCDGT terms in 655, 385, 386 fields
  - ▶ Example: 650 \_0 ...\$v Genealogy → [655 \\_7 Family histories. \\$2 lcgft](#)
- ▶ Genre/Form: mapping LCSH music form headings to LCGFT terms in 655 fields
  - ▶ Not SSFV's own work, but a brief description of and links to the [Music Library Association's work](#)
- ▶ Medium of Performance: mapping LCSH music form headings to LCMPT terms in 382 fields
  - ▶ Not SSFV's own work, but a brief description of and links to [MLA's work](#)



# Scenario:

## Collaboration between Programmers and Librarians

- ▶ [OCLC Music Toolkit](#) developed by Gary L. Strawn, in collaboration with Music Library Association, based on MLA's specifications (algorithms)
- ▶ Mapping example
  - ▶ **650 \_0 String quartets \$v Scores** *begets*
  - ▶ **382 01 violin \$n 2 \$a viola \$n 1 \$a cello \$n 1 \$s 4 \$2 lcmpt** *and*
  - ▶ **655 \_7 Chamber music. \$2 lcgft; 655 \_7 Scores. \$2 lcgft**
- ▶ Mapping logic is complex, requiring years of iterative testing and refinement (still ongoing)
- ▶ Music Toolkit is a cataloger productivity tool as well as a testing mechanism
- ▶ Eventual goal: design some version of the MLA algorithm to run on an entire database, with limited human review



# Scenario:

## Testing at a Small Scale with Database Administrators

- ▶ OCLC testing of form subdivision mappings (began in 2021)
- ▶ Testing example
  - ▶ 650 \_0 ... \$v Dictionaries \$x [Language] → 655 \_7 Multilingual dictionaries. \$2 lcgft
  - ▶ Testing partner suggested adding additional criteria: **650 \$a must include the word “language”**
  - ▶ Not necessary to enumerate all possible languages that might appear in \$x
  - ▶ Unresolved question: if **Dictionaries** is already present in the record and **Multilingual dictionaries** is then added, does this create a problem?
    - ▶ LC policy discourages using both a broader and a narrower term to describe the same resource.



# SSFV Objectives and Future Work

- ▶ Develop additional modules
  - ▶ In progress: mapping LCSH literature form headings to LCGFT and LCDGT terms
  - ▶ Example: 650 \_0 Bruneian poetry (English) → [655 \\_7 Poetry. \\$2 lcgft](#) and [386 Bruneians \\$2 lcdgt](#)
- ▶ Encourage testing across diverse “real world” environments
- ▶ Refine mapping specifications over time (for existing modules)
- ▶ Share our work with the broader community incrementally
- ▶ Overarching goals (once again)
  - ▶ Increase the prevalence of faceted data in library catalogs over time (in the USA and globally)
  - ▶ Achieve more sophisticated user outcomes through faceted access; this requires a “critical mass”
  - ▶ Pursue sustainable “subject” access (broadly speaking) for metadata workers and users alike



# Thank you!

► Questions: [caseymullin1@outlook.com](mailto:caseymullin1@outlook.com)

