

LCDGT Pilot Project Report

Autumn Faulkner, Julia Frankosky, Tim Kiser, & Nicole Smeltekop
Submitted August 2018

Project summary

From January through July 2018, engage in experimental demographic group term (DGT) — in particular, Library of Congress Demographic Group Terms (LCDGT) — application to get a handle on how to use DGTs, and reflect, based on time investments and overall sense of value added, whether broader application is worth considering in the short- to medium-term (next 2-3 years) and which materials are likely to benefit most from that application.

- Experimentation in individual cataloging (68 records)
 - Terms applied to records for books, maps, video games, music, and archival materials
 - Mostly occupational or age terms used in 385 Audience field
 - **Boaters, Geologists, Children, Teenagers, Adults**, etc.
 - Far fewer instances of 386 Creator field
 - Examples include **Museum curators, Government employees, Mexican Americans**
- Experimentation with special groups of materials (3133 records)
 - Eunil Lee contemporary Korean literature collection — 385 **Koreans** and 386 **Koreans** applied to 518 records
 - Children's literature — applied 385 **Children** to 2615 records

Project aims

1. Develop an understanding of the application and use of DGTs in bibliographic and authority data
 2. Illuminate issues around display and utility of DGTs in the ILS
 3. Determine if there are varying levels of added value of DGTs for particular collections or formats
 4. Characterize impacts of introducing DGT application on workflow turnaround and general cataloging productivity
 5. Taking all this into account, provide recommendations on how to proceed.
-

1. Application of DGTs in bibs and authority records

Bibliographic records

In applying LCDGTs, we focused on the 385 and 386 fields. These fields describe the intended audience and creator information respectively. The coding for each is as follows:

385	386
\$a - Audience term (R) \$b - Audience code (R) \$m - Demographic group term (NR) \$n - Demographic group code (NR) \$0 - Authority record control number or standard number (R) \$1 - Real World Object URI (R) \$2 - Source (NR) \$3 - Materials specified (NR) \$6 - Linkage (NR) \$8 - Field link and sequence number (R)	\$a - Creator/contributor term (R) \$b - Creator/contributor code (R) \$i - Relationship information (R) \$m - Demographic group term (NR) \$n - Demographic group code (NR) \$0 - Authority record control number or standard number (R) \$1 - Real World Object URI (R) \$2 - Source (NR) \$3 - Materials specified (NR) \$4 - Relationship (R) \$6 - Linkage (NR) \$8 - Field link and sequence number (R) ¹

We focused on adding \$a (the LCDGT term) and a \$2 noting that the term came from LCDGT. We did not experiment with adding \$m or \$n, but the [MARC list of demographic group categories](#) is relatively short (11 terms), so these subfields could be added later with a batch process.

In adding LCDGT terms, the \$a is repeatable in the same field if all the terms came from the same source. If they did not, the cataloger can use multiple 385 and/or 386 fields as needed. In most cases our terms were from LCDGT, so multiple 385 and/or 386 fields were generally unnecessary.

Example: 386 0 \$a Kenyans \$a Government employees \$2 lcdgt

If our cataloging team decides to add these terms going forward, they would need additional guidance for applying these terms. As with authority records (see next section), recording demographic information about creators in particular can get difficult quickly and may also incite some ethical dilemmas and categorizing people in ways they may not categorize themselves. The cataloging community at large is currently wrestling with these issues, so there is not a lot of guidance available from the national bodies. Until such documentation exists, it is not feasible to implement these terms throughout our catalog.

Authority records

As with bibliographic records, LC Demographic Group Terms can be added to the 385 and 386 fields. The 385 field for audience characteristics is used for uniform title and work authority records. The 386 field is used predominantly in name authority records.

Catalogers have been adding terms of a demographic nature to authority records for some time. Because this is established practice, we do not see a need to emphasize or discourage the use of LCDGTs in authority records. However, we do encourage continued discussion of ethical issues surrounding the usage of LCDGTs in cataloging workflows in our CMS team meetings. The current focus is on gender terms, but as the cataloging community continues to add categorical terms to describe creators, additional ethical issues will probably emerge.

¹ <https://www.loc.gov/marc/bibliographic/concise/bd385.html>

The use of the 386 in addition to the 372 field (field of activity) may be worth some consideration. Many of the terms could apply to either or both fields, such as 'Composers.' Although this may seem to introduce redundancy, it should not be a deterrent to using both, because they serve different purposes. And in other cases, there are terms that could be used in the 386 that could not be reflected in the 372, and would add an additional dimension to the identity of the person being described.

2. Utility and display

Incorporating DGTs in bibliographic records presents issues related to subfield separation and display in the Innovative online catalog. Since multiple terms can be included in one field, this can create a confusing display for our users as there is neither separation between terms, nor punctuation, causing our users to see a long list of seemingly unrelated terms. This is similar to issues encountered with new MARC fields for RDA music cataloging. Additionally, there is currently no functionality in place for use of these fields in both bibliographic and authority records. While the bibliographic terms could be included through keyword indexing, this presents a new problem: mixing audience and creator characteristics (e.g. when searching for literature **by** Korean Americans, results would also include literature **for** Korean Americans). Rather than aiding users in their search, the combination of these characteristics would confound the search results and lead to more false hits.

As for authority records, the usage of these terms and their benefits are currently moderately established in the cataloging community. DGTs enable catalogers to better identify and disambiguate creators, thus making it a useful addition to cataloging workflows. However, in our current integrated library system there is no functionality in place that would enable the use of DGTs to be discoverable, so they are not useful for the discovery of materials by library users.

For a successful implementation, a separate index for the 385 and 386 fields would be required to allow for faceting. However, this route would be contingent upon several criteria being met. A critical mass of records with DGT terms would first need to exist in order to make this a worthwhile approach. How indexing and faceting are implemented would also heavily depend on the functionality of the integrated library system utilized by the library. Additionally, promotion of these features would be required; if our users do not know that they exist or how to use them, then they are meaningless. As such, instruction and reference librarians would be critical in educating library users on the most effective way to utilize these fields in their searches.

3. Added value for particular bibliographic collections or formats

We found that many of the DGTs suitable for use in our routine cataloging would not be particularly valuable to searching, often because they are in some way redundant to Library of Congress Subject Headings (LCSH) and/or Library of Congress Genre/Form Thesaurus (LCGFT) terms that already contain demographic information (implicitly or explicitly). For example:

- Audience LCDGT "Tourists" vis-à-vis LCGFT "Tourist maps"

- Audience or creator LCDGT “Geologists” vis-à-vis LCSH “Geology”

We feel that DGTs would add the most value when applied to unique or niche collections that need more exposure in MSUL’s search environments. However, with respect to some collections, the available DGTs are rather limited. When applied to the Rovi Video Game collection, the only applicable audience terms were “Teenagers” and “Adults,” terms which are duplicative of the ESRB rating found in the 521 8_ field.

In most cases, wide retrospective application of DGTs via simple, straightforward batch projects would only be feasible for terms in MARC 385 (Audience) fields. However, such projects may not add much value unless combined with projects that would simultaneously apply suitable LCGFTs — for instance, a retrospective application of LCGFTs specific to Korean literature such as Sasöl sijo, Kamyŏn’gŭk plays, Sandae drama, etc., combined with appropriate LCDGTs.

More complex batch projects to retrospectively apply terms in MARC 386 Creator fields could be feasible for collections (or segments of collections) for which distinct creator characteristics can be identified, but these would be likely to require careful attention and significant time commitments to identify suitable records. Africana posters may hold some potential for such a project, focused on creator information that is not brought out elsewhere in the record - for instance, a poster created by a Norwegian non-profit organization’s Kenyan office about their relief programs in Kenya; the Norwegian aspect of creation could be lost. The Rovi collection also may present interesting opportunities. One model we could potentially follow is a project recently begun by Illinois State University to add the LCDGT “African Americans,” first to name authority records and eventually to bibliographic records, using Wikipedia’s “List of African American writers” as a starting point. The project has involved assigning a student worker to cross-check the ISU catalog against the Wikipedia list and compile a list matching names, and hiring a graduate student to further research the suitability of “African Americans” for each name. (An ISU presentation about the project may be viewed [here](#); MSU login required).

4. Workflow and productivity (Bibliographic and authority)

While we are not recommending immediate incorporation of DGT application in daily cataloging practice (see Section 5 below), we feel the question should be revisited in a few years’ time. If and when implementation is recommended, here are considerations for any future efforts:

1. Initially, searching for applicable terms would be time-consuming; catalogers and CMS managers should expect additional time spent on records as a result. However, additional time for becoming familiar with a new vocabulary shouldn’t be a huge deterrent; exposure to this more granular descriptive process could be helpful in and of itself, as a way of preparing catalogers for future environments.
2. In general, the group decided the 385 Audience field would probably have more immediate usefulness to patrons, in terms of resource discovery, than the 386 Creator field, which seems to have more utility in authority records by helping catalogers to identify and disambiguate individuals. But if a cataloger is taking trouble to add the 385 to a bib record, a 386 might as well be recorded. Although use cases for searching by creator demographic are less readily apparent, we can at least assume some expert researchers would find it a useful tool, and that other use cases may emerge that we can’t anticipate.

3. Constant data should be employed to help with efficiency and routine adoption; also, recommended lists of terms pertaining to special formats or collections would be a helpful tool (see Section 4).

If and when implementation is recommended, we also think it would be worth seriously exploring retrospective batch work and possibly automation via mappings or APIs, in addition to routine practice by individual catalogers. This would address both the niche collection aspect discussed in Section 3, and also increase efficiency and progress toward a critical mass of these terms once a decision is made to index them and make them publicly searchable.

5. Recommendations

Concerns

We find that the DGT thesaurus has some rigidity in its terminology. For example, the only term applicable to a collection of Rovi albums produced by bands with alt-right associations is **White supremacists**. We were ultimately uneasy with applying this designation, since as far as our judgment informs us, persons associated with alt-right movements or thinking cannot consistently be identified as white supremacists. And even if there were arguments to the contrary, catalogers will never be fully equipped to sift through that level of nuance (and shouldn't be expected to). These kinds of gaps in the vocabulary—whether ethical in nature like this one, or purely logistical—could be resolved via proposals from catalogers, but obviously that would require a lot more time and energy.

A related question arose about whether the creator's intention should factor into the cataloger's choice of appropriate terms, and how much interpretation might be involved there for the cataloger. To elaborate on the earlier example, let's imagine a band with an album that ends up being embraced by white supremacists because of chance references in the lyrics or in the cover art. This band may have no intention of associating themselves with white supremacist ideology, but if down the road the album becomes strongly connected with white supremacist groups anyway, to the point where patrons might expect to find it in search results for "white supremacist music," is it the catalogers' responsibility to apply the Audience DGT? This is a highly theoretical question, but this point could use some clarification from LC before any attempts are made to seriously implement DGT application in cataloging.

Probably the most pressing concern the group had, however, is that there is not a lot of evidence of wide adoption of DGT application in cataloging practice, and that we should use caution in investing energy into an area of descriptive practice that may or may not persist.

Ultimate recommendations

1. At this time, do not expect catalogers to apply DGTs in daily practice, and continue suppressing public display of 385/386 fields in bibliographic records.
 - The vocabulary is not yet robust enough to allow for very powerful retrieval, or to escape redundancy of terminology.
 - Best practices for its application require some fleshing out; current guidance for catalogers is insufficient.
 - Indexing and display conventions in our current system would present some challenges that probably aren't worth tackling until we feel use of DGTs presents a lot more concrete benefits than it does now.

- However, if any catalogers are enthusiastic about using specific terms for special projects, this shouldn't be discouraged. Such work on niche collections could be valuable down the road.
 - For example, at the start of the Lockwood Romani collection cataloging project, we could have identified some possible DGTs to include with our cataloging which might have become useful in the future.
 - As detailed in Section 1, catalogers should only use the \$a and \$2 subfields at this point, if they wish to use DGTs in their cataloging.
- 2. Revisit this question in 2-3 years, as progress is made toward more granular cataloging in a linked data environment, and usage of a variety of vocabularies grows more common.
 - Investigate indexing and public searching possibilities. The question of adopting routine practice is inextricably bound up with the effectiveness of indexing and display.
 - Investigate one-off batch projects around special collections, which the group feels have higher potential for retrieval and exposure via DGT. Though batch application of terms would require thoughtful, time-intensive identification and curation, it could also produce some very exciting results in the public catalog.
 - Consider collections of note, like Rovi, Radicalism, African-American cookbooks, etc.