University Libraries Faculty Scholarship   University Libraries

3-31-2023

Finding Harmony: Social Tags and Name Authority Records in the Popular Music Domain

Kabel Nathan Stanwicks
University at Albany, State University of New York, kstanwicks@albany.edu

Hemalata Iyer
University at Albany, State University of New York, hiyer@albany.edu

Recommended Citation

Finding Harmony: Social Tags and Name Authority Records in the Popular Music Domain

Stanwicks, Kabel Nathan   University at Albany, SUNY, U.S.A. | kstanwicks@albany.edu

Iyer, Hemalata   University at Albany, SUNY, U.S.A. | hiyer@albany.edu

ABSTRACT

Libraries and cultural institutions use cataloging and metadata standards to describe works and those responsible for creating them. These standards are used to create name authority records (NAR) to collocate creators' works under single, authorized versions of their names. NARs employ controlled vocabulary to describe individuals and disambiguate names. NAR production is costly; therefore, few are updated when standards are revised, and controlled vocabularies are slow to adapt to a rapidly changing information landscape. Artificial intelligence provides opportunities to leverage social tagging to update and enhance NARs, while simultaneously improving user search and discovery. Our research demonstrates that social tags in the popular music domain align with the Resource Description and Access cataloging code's NAR constructs. Next steps will examine how social tags relate to controlled vocabulary terms, as well as explore the use of artificial intelligence to enhance NARs and library search interfaces.

KEYWORDS

Name Authority Records, Social Tagging, Music

INTRODUCTION

Name authority records (NAR) are used by libraries, cultural institutions, and information organizations to differentiate records for people with similar names and pull together the various spellings, translations, and pseudonyms by which someone is known under a single, authorized version of their name. The amount of time required to research and create authority records, coupled with the skill level necessary to effectively perform authority control, make authority control the most expensive component in cataloging (Boese, 2003; Russell & Spillane, 2001; Tillet, 2004). Collaborative cataloging cooperatives have emerged over the years to share the burden of this work and reduce individual institution's costs; however, authority work is still difficult to perform and costly to maintain.

PREVIOUS RESEARCH

Our prior study demonstrates that social tags in the popular music domain align with the attributes and entities used to construct NARs (Stanwicks & Iyer, 2018). For that study, we pulled a sample of user-generated tags applied to musicians and bands on the last.fm web site, developed a content analysis dictionary based on name authority data elements found in the Resource Description & Access (RDA) cataloging code, and performed content analysis on the tag sample. Intercoder reliability was calculated using Cohen's Kappa, which ranged between 0.857 and 1 for the coded attributes. Results demonstrated that user tagging focuses heavily on genre, geographic places, occupation, and gender when describing musicians and bands.

CURRENT STUDY

This paper reports on our current study, which is a work in progress and a continuation of our previous study. Our current study compares the coded data from our first study to artists' corresponding Library of Congress NARs to examine which data fields are already utilized in these NARs and explore how the user-generated tags that we coded to specific fields overlap with, or enhance, existing records.

METHODOLOGY

The authors searched the Library of Congress catalog for corresponding NARs for the 56 musician and band names sampled in this study, which were drawn from the previous one (Stanwicks & Iyer, 2018). The NARs were examined to determine which records' descriptive formatting had been updated to RDA, and each record's formatting was noted in a spreadsheet used to gather the NARs data. The authors extracted a total of 471 descriptors from the MARC 1XX through 5XX fields into this spreadsheet for analysis. A pivot table was created to obtain counts of descriptors applied to each of these records and descriptive statistics were generated for the sample. The professionally-generated descriptors were compared to the prior study's 1,716 user-generated tags that aligned with 1XX through 5XX MARC field name authority attributes to observe similarities and differences between these two approaches to resource description.

RESULTS & DISCUSSION

Controlled terms used in NARs provide precision for search and discovery, while user-generated tags contain more current terms, can enhance recall, and can improve subject access (DeZelar-Tiedman, 2011; Gorman, 2004; Hider & Steele, 2021; Vaidya & Harinarayana, 2016). Catalogs and discovery systems that integrate both controlled

vocabularies and user-generated tags have the potential to provide users with the best of both worlds. Therefore, coupling these two data sources can significantly enhance NARs. For example, the Library of Congress NAR for the musician, Snoop Dogg, includes *Rap (Music)* and *Popular music* as Field of Activity (i.e., genre) entries, whereas last.fm user-generated tags coded to this attribute include *west coast, hiphop, Hip-Hop, alternative, hip hop, Gangsta Rap, West Coast Rap, west coast hip hop, G-funk, rock, pop, indie, dance, rap, rnb, funk,* and *west coast hip hop*—these tags provide more granular description than the two broad genre terms in the NAR.

NARs updated to RDA encoding generally have more descriptors than older NARs encoded under previous cataloging codes (Moulaison, 2015). Out of the 56 NARs examined in this phase of the study, 45 were encoded according to RDA, while 11 were encoded under previous cataloging codes (non-RDA). The average number of descriptors applied to MARC 1XX through 5XX fields in each of the RDA encoded records was 9.5, whereas the average number of descriptors applied to these same MARC fields in non-RDA records was 3.9. Records varied significantly regarding the number of fields and descriptors used per record. RDA encoded records had a minimum of 2 descriptors applied to sparser records, a maximum of 22 descriptors applied to more robust records, and a mode of 7 descriptors per record. The non-RDA records had a minimum of 1 descriptor applied to sparser records, a maximum of 15 descriptors per record, and a mode of 1 descriptor per record. The sampled records contain 96 occupation descriptors, 74 variant names by which people are known, 69 place descriptors (i.e., place of birth, place of death, other associated places, associated country), 68 references from individual band members to bands in which they performed and vice versa, 47 Field of Activity descriptors primarily describing genre of music these artists perform, and 16 gender descriptors (Figure 1).

Significantly more user-generated tags were applied to these sampled names on the last.fm web site than professionally generated descriptors in the corresponding Library of Congress NARs. 1,716 user-generated tags from our last.fm tag sample aligned with 1XX through 5XX NAR MARC fields. Each last.fm artist record sampled contained an average of 29 user-generated tags aligned with NAR MARC fields. Variant names, an underpinning of NARs, were not prevalent within the tag sample. Occupation and place description tags were found among the sampled tags with 176 occupation tags and 202 place tags. Unlike the Library of Congress NARs sampled, user-generated tags contain high numbers of genre (943 tags) and gender (138 tags) tags in the study sample. Furthermore, the sampled tags describing these musicians and bands on the last.fm web site provide greater breadth and variety, as well as more granularity, than the descriptors found in corresponding Library of Congress NARs.

The average number of descriptors in non-RDA NARs was 32% less than in RDA encoded records. However, the artists and bands with non-RDA NARs had an average of 26.3 user-generated tags aligned with NAR MARC fields on each of their last.fm artist pages. This demonstrates that expanding NARs through the inclusion of user-generated tags can enhance records' currency, expand description that aligns with the RDA cataloging code, foster cultural inclusivity in library catalog records, and improve search capabilities. For example, the Library of Congress NAR for Santo & Johnny contains only two fields—the authorized corporate body name and an alternative name. However, the last.fm user-generated tags for Santo & Johnny consist of 48 tags that align with RDA NAR fields. Therefore, combining professionally generated descriptors with user-generated tags significantly enhances name description for this band (Figure 2).

FUTURE DIRECTIONS

Immediate next steps in our research will involve correlating and connecting social tags from our study's sample with professionally generated descriptors and identifying broader approaches for linking these data. Our goal is to explore and apply artificial intelligence to develop trustworthy linkages between social tags and professionally generated descriptors while eliminating the noise found in personal and subjective tags, as well as test the use of artificial intelligence to update NARs—moving beyond the sparse, basic level records used to disambiguate names, toward fuller description of the people these records describe. If we can enhance controlled vocabularies with additional cross references based on user description, we will improve the currency and cultural inclusivity in our controlled vocabularies, which, in turn, will reduce barriers in user search and discovery. Enhancing and enriching NARs has the potential to both transform how these records are used in search and discovery, as well as significantly reduce personnel resources needed to create and maintain these records.
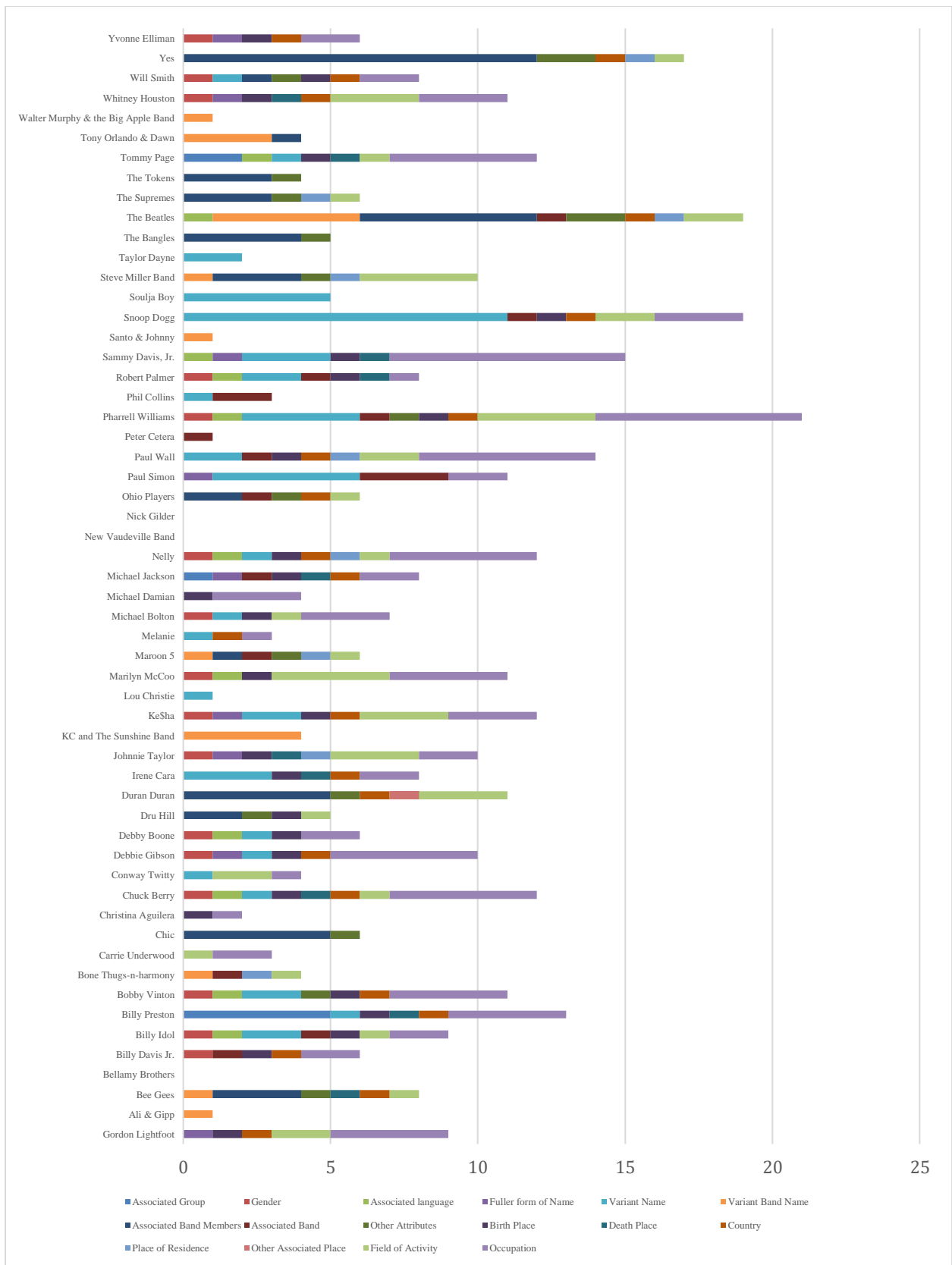
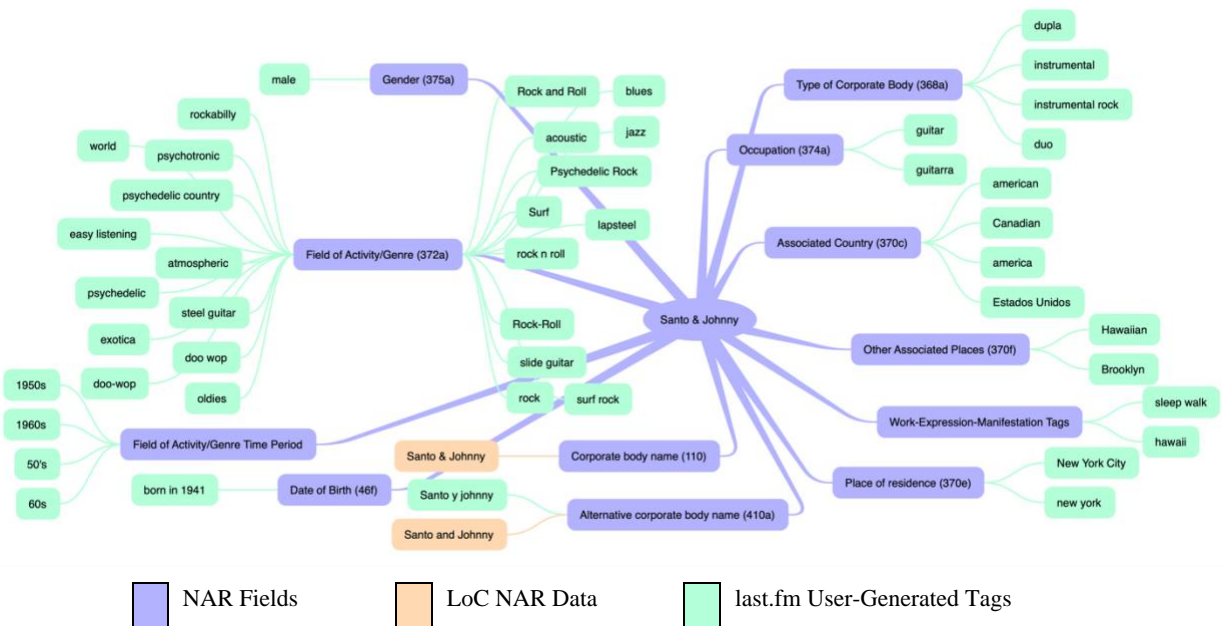**Figure 1. Library of Congress NAR Descriptor Distribution**

Figure 2. Santo & Johnny NAR Enhanced with User-Generated Tags

CONCLUSIONS

Social tagging data are both rich and messy. Tags contain the flaws inherent to natural language that information professionals strive to avoid, but most tags observed in this study align with NAR constructs. Therefore, these tags have the potential to enhance NARs. Allowing these data sources to co-exist in catalogs and databases can lower barriers to user search and discovery. Applying artificial intelligence will be necessary to realizing these results at a larger scale and throughout other knowledge domains.

REFERENCES

Boese, K. C. (2003). What's in a name: Associated costs or authority work for artists' names. The Bottom Line, 16(3), 106–110. http://dx.doi.org/10.1108/08880450310488021

DeZelar-Tiedman, C. (2011). Exploring User-Contributed Metadata's Potential to Enhance Access to Literary Works: Social Tagging in Academic Library Catalogs. *Library Resources & Technical Services*, *55*(4), 221–233.

Gorman, M. (2004). Authority Control in the Context of Bibliographic Control in the Electronic Environment. *Cataloging & Classification Quarterly*, *38*(3–4), 11–22. https://doi.org/10.1300/J104v38n03_03

Hider, P., & Steele, G. (2021). LibraryThing and Literary Works Revisited: Are Social and Library Cataloging Just as Complementary as they were a Decade Ago? *Library Resources & Technical Services*, *65*(3), 113-125. http://dx.doi.org/10.5860/lrts.65n3.113

Moulaison, H. L. (2015). The expansion of the personal name authority record under Resource Description and Access: Current status and quality considerations. *IFLA Journal*, *41*(1), 13–24. https://doi.org/10.1177/0340035215570044

Russell, B. M., & Spillane, J. L. (2001). Using the Web for Name Authority Work. *Library Resources & Technical Services*, *45*(2), 73–79. https://doi.org/10.5860/lrts.45n2.73

Stanwicks, K. N., & Iyer, H. (2018). Social tagging and name authorities: Popular music information discovery. *Proceedings of the Association for Information Science and Technology*, *55*(1), 900-901.

Tillett, B. (2004). Authority Control: State of the Art and New Perspectives. *Cataloging & Classification Quarterly*, *38*(3–4), 23–41. https://doi.org/10.1300/J104v38n03_04

Vaidya, P., & Harinarayana, N. S. (2016). The Comparative and Analytical Study of LibraryThing Tags with Library of Congress Subject Headings. *Knowledge Organization*, *43*(1), 35–43. https://doi.org/10.5771/0943-7444-2016-1-35