Wan-Chen Lee

Author information for Knowledge Organization in the digital age

Abstract

Authorship is a critical topic in knowledge organization. In the Anglo-American cataloging tradition, authorship is often used as a characteristic for arrangement and as an access point for retrieval. There is a body of literature in knowledge organization that examines the meaning of authorship. Through literature review, we can see that authorship is a cultural, social, and temporal concept, and author information in the cataloging tradition does not present all the author functions. This paper discusses challenges concerning the collection, recording, and representation of author information in the digital age, as well as the consequences of enriching author information. We argue that more information is not necessarily better for cataloging in the digital age.

Introduction

Authorship is a critical topic in knowledge organization. In the Anglo-American cataloging tradition, authorship is often used as a characteristic for arrangement and as an access point for retrieval. Author information serves both the collocating and identifying objectives of library catalogs (Cutter 1876). Tracing back to Roland Barthes’s The Death of the Author and Michel Foucault’s response What is an author? previous studies review the authorship discussion and examine authorship in different times, social and cultural contexts, and across various forms of works.

Foucault points out that the concept of author goes beyond the attributes of an individual. Identifying someone as the author of a work tells us more than the person being the content creator. Author also indicates what Foucault calls the author function, which is a complex mass of characteristics influenced by social, cultural, and temporal factors (Foucault 1970). An author may indicate certain value, a style, a mode of discourse, or historical events. Using Freud as an example, Foucault shows how an author, as a founder of a discourse, may have influence on other authors and works that go beyond the individual’s intention. The author function may group, differentiate, and relate texts, as well as characterize a discourse or specify ownership (Martínez-Ávila et al. 2015; Foucault 1970). Expanding on this, thinkers in the knowledge organization context explore and identify definitions, characteristics, and functions of the author. Through examples of studies of superworks like the Seven Epitomes, Abelard’s Works, and The French Chef (Smiraglia et al., 2011; Smiraglia and Lee 2012; Smiraglia et al. 2013; Martinez-Ávila et al. 2015; Lee 2016); as well as studies of contemporary projects like the Europeana, AustLit, the American Civil War: Letters and Diaries, and DBpedia (Moulaison et al. 2013; 2014), we can recognize some prominent observations. First, authorship is not a mere attribution. It is a culturally, socially, and temporally bound concept which is related to information objects (e.g., books). We can
explain this with an example. Recording William Shakespeare as the author of *Romeo and Juliet* addresses more than identifying him as the person responsible for the content and creation of the work. Our understanding of Shakespeare and his time provides hints for interpretations of his works, and it may shed light on relationships between *Romeo and Juliet* and other works or people. The second observation drawn from the knowledge organization literature is that author information, in the cataloging tradition, does not present all the Foucauldian author functions. The cataloging tradition treats author quite simply, as a text string. For the purpose of authority control, catalogers often use the birth date and death date, or the occupation, as qualifiers to distinguish different identities with identical names. Even if we check the name authority record, the available author information and author functions are still limited. While the classificatory function and ownership aspect of Foucauldian author function is present in this context, we do not have access to style, value, historical background, and other aspects important to Foucault (Foucault 1970).

The U.S. cataloging standard RDA (Resource Description and Access) places more emphasis on relationships. And because of this, catalogers are recording more author information. However, the author function introduced by Foucault (1970) identifies particular relationships and expansive influences of an author which go beyond the coverage of current cataloging practice. In the works reviewed above, scholars acknowledge the richness of authorship and recognize the limited recording of author information in cataloging practice. For example, Moulaison *et al.* (2013; 2014) elaborate on how events may be a critical aspect of an author’s life. An author might meet people at events and change their relationships, thoughts, or works. However, few knowledge organization systems record and present event information as part of the author information. The authors use NNBD Mapper to provide an example of visualization of event information of a person. They suggest recording more author information and making them searchable for users to realize Foucault’s author function in current knowledge organization systems.

Based on the literature reviewed, this paper recognizes the motivation of pursuing better presentation of author function in knowledge organization systems, but argues that more information is not necessarily better for cataloging in the digital age. To do this, we discuss three challenges concerning collecting and recording author information in the digital age. The first challenge focuses on representing author information. The second challenge focuses on collecting author information. The third challenge is concerned with the consequences of enriching author information.

**Representing author information**

The first challenge focuses on the representation of author information. How can we address the cultural and temporal nuances of authorship in a “standardized” manner, so institutions can share metadata? On one hand, we recognize the meaning of authorship
differs by time, culture, and social context. On the other hand, we need a modicum of standardization to enable metadata sharing.

Lee (2016) identifies six types of authorship presented in the *Seven Epitomes* (*Qilue*), which is, for imperial China, “the first classified catalog and the one that established the model for bibliography”. These six types, namely composer, speaker, transmitter, exegete/commentator, patron, and editor reflect the definitions of authorship at the time of the catalog (i.e., the Former Han dynasty, 206 B.C.E. - 8 C.E.), in the Chinese culture, and for particular social contexts, like the catalog for the imperial library. While some types of authorship, such as *composer*, may be covered by *author* in RDA, others may map with different roles contributed to the creation of the work. For example, we may map *patron* in the *Seven Epitomes* to *sponsor* in RDA. Further, there is no guarantee of a one-to-one mapping. In Lee (2017), the author discusses six mapping issues between Chinese and English role designators. There are Chinese terms indicating mixed English roles (one-to-many mapping), old or ambiguous terms, synonyms (many-to-one mapping), homographs (one-to-many mapping), homographs across languages, and Chinese terms with no match in English (one-to-zero mapping). Some of the issues are complicated and can be obscured by translation. The combination of cultural, temporal, and social nuances may be lost or misunderstood in the activities of translation, mapping, and the cataloging process.

The two examples above only present part of the challenges of representing author information. Other issues may arise when catalogers use controlled vocabularies to represent author-provided information. Thompson (2016) conducts content analysis of name authority records of authors who self-identify as trans. She shows how the limitations and values embedded in controlled vocabularies and cataloging rules may be imposed to author information. One of her examples is assigning start and end dates to a gender identity. This approach assumes that an individual’s gender shifts from one to another, rather than being fluid. The assumption imposed by the descriptors may not faithfully represent an author’s self-identification.

To improve the presentation of the author function in information systems, we may want to go beyond mapping types of authorship and role designators between different times, cultures, and societies; or go beyond representing author information using a limited controlled vocabulary. We would expect systems to provide richer author information, such as the intellectual tradition of an author, an author’s relationships with other authors, or the influences of an historical event on an author. The challenge of representing author information is one of improving the flexibility and hospitality of author descriptions to record context-specific expressions, while remaining interoperable across information systems.
Collecting author information

The second challenge focuses on collecting author information. When it becomes easier to collect information, make linkages, and specify relationships in the digital age, what are the concerns of collecting and providing author information? Moulaison et al. (2014) mention ethical concerns of both recording particular author attributes and the challenge of making the author information searchable. Thompson (2016) addresses privacy and safety concerns about collecting and providing certain author information, such as gender identity.

In addition to these concerns, another issue lies in considering the appropriate sources of information. What author information would be both practical and ethical in scope for catalogers to collect? One example case is Wikipedia, which relies on Wikipedia editors to populate information from a variety of information sources. According to the Wikipedia: Identifying reliable sources page, “Wikipedia articles should be based on reliable, published sources.” In addition, “contentious material about living persons (or, in some cases, recently deceased) that is unsourced or poorly sourced—whether the material is negative, positive, neutral, or just questionable – should be removed immediately and without waiting for discussion” (Wikipedia 2017).

We also have a case like the Library of Congress Demographic Group Terms (LCDGT), which includes eleven categories of terms (i.e., age; educational level; ethnic/cultural; gender; language; medical, psychological, and disability; national/regional; occupation/field of activity; religion; sexual orientation; social) to describe the demography of creators, contributors, and intended audiences of resources (Library of Congress 2017). Catalogers can assign LCDGT in both bibliographic records and authority records. The purpose of LCDGT is to cover the demographic terms in the LCSH. With the development of another thesaurus – the Library of Congress Genre Form Terms (LCGFT), the goal is to separate both demographic terms (LCDGT) and descriptors of what a resource is (LCGFT) from subject headings which describe what a resource is about (LCSH). According to the manual of the LCDGT (Library of Congress 2017), the practical and ethical guidance given is that the source of information for LCDGT should be limited to authors’ self-provided information in the piece described. The manual instructs catalogers to use other information sources, including previous works by the same author, with caution. It acknowledges the complexity of author information and how identity may change over time. However, the guidelines assume that authors provide information in the resources described. It is an assumption we cannot take for granted. In order to publish a work, there may be censorship, either imposed or self-applied. Further, there are questions concerning the maintenance of LCDGT. Following the guidelines, the LCDGT, which describe the author in a bibliographic record, function like a snapshot of the author’s self-identified demography at the time of publication. However, in authority records, catalogers record
author information over time. How practical it is to update the LCDGT in authority records if demographic information changes? Could what is unveiled be veiled again?

In the digital age, there is a tension between the ethics of respecting authors’ privacy and providing richer author information. Both aforementioned cases address the ethical concerns of collecting personal information with caution. The LCDGT case limits the information source to author-provided information. The Wikipedia case allows a wider variety of published information sources; and it has guidelines and policies on reliability and verifiability of information sources. However, the evolving approaches for publishing information may create grey areas. Guidelines which limit the sources of information to published works or publicly available information may not be as clear-cut as we expect.

**Consequences of enriching author information**

The third challenge is concerned with the consequences of enriching author information. Does the amount of author information available influence the visibility and access of the authors’ works? By populating author information and making connections between authors and works, there is good intention to improve the selection and interpretation of works. For instance, information about an author’s relationship with a particular school of thought could help readers interpret the author’s works. As Moulaison *et al.* (2013) argue, knowing the author’s age at the time of publication may influence the selection of works.

Besides the improvement of selection and interpretation of works, in the digital age, richer author information may result in more access points and links to an author’s information, and thus increase visibility of an author and an author’s works. However, if richer author information leads to higher visibility of an author’s works, it may render other authors less visible. And this can be seen as an ethical issue. We may want to identify the factors influencing the recording and searchability of author information. For instance, does the language and script matter? If author information in some languages were more visible than others, we could examine the knowledge organization systems to detect influential factors other than reader preferences. Is a system hospitable for information in different languages? Would international collaborations complement the local system? An expanded version of VIAF (the Virtual International Authority File) might be one possibility (Online Computer Library Center 2018).

One other concern is how the presentation of rich author information influences contemporary authors, and what are the ramifications of this to non-contemporary authors. Generally, it is easier to enrich author information for people who are currently active than for people who wrote centuries ago. If the recording of author information systematically emphasizes some groups (*e.g.*, contemporary authors) and render the other groups (*e.g.*, non-contemporary authors) less visible, enriching author
information may lead to another ethical issue.

In the digital age, we have better abilities to enrich author information. We have access to more detailed information about some authors. Technologies enable us to add more links between authors and works across languages, regions, and times. It becomes easier to populate information and create links with efficiency and at a large scale. However, we should examine the approaches and consequences of enriching author information, and take different stakeholders into account. Some stakeholders include authors of different time periods, regions, cultures, and languages; and information professionals who aim to describe authors in both ethical and practical approaches; as well as people who find, select, and interpret works using author information. Since not all stakeholders are equally vocal, it is important for information professionals to identify them and their stances, and think through the consequences of enriching author information.

Conclusion

The purpose of raising these challenges is not to deter the pursuit of presenting the author function in knowledge organization systems. Instead, the goal is to increase awareness of some concerns, and encourage discussion and studies for approaches to ameliorate potential issues. In the digital age, with the ability to share, collect, and link information at a large scale, it is important to ask if providing more information is necessarily better.

References


Thompson, Kelly J. (2016). More than a Name: a Content Analysis of Name Authority Records for Authors Who Self-Identify as Trans. Library Resources & Technical Services, 60(3): 140-155. Available at: https://doi.org/10.5860/lrts.60n3.140.
