

This is the first part of a paper to be presented at the CIG Annual Conference 2018. The second part will be published after the conference.

Disclaimer: the opinions and errors herein are personal. They should not be taken to represent the formal view of the RDA Steering Committee.

Let's talk about sex.

That's what the RDA Steering Committee (RSC) found itself doing a couple of years ago, in its previous incarnation as the Joint Steering Committee for Development of RDA. It had all started so innocently, with a request from one of the RDA constituencies to add a term to the controlled vocabulary for the element "gender".

There are around 40 elements that have controlled vocabularies in RDA: Resource description and access, a content standard for library and cultural heritage metadata. A "vocabulary encoding scheme" (VES) is a set of controlled terms with associated synonyms, numerical codes or notations, and linked data URIs. A cataloguer may select a term, its notation, or its URI as the value of the element for the specific instance of the entity being described. For example, the content type of an expression that is embodied in a printed volume may be assigned the English term "text", the notation "1020", or the URI <http://rdaregistry.info/termList/RDAContentType/1020>; the term is available in 15 languages, but the notation and URI are the same for all translations.

The "gender" element is associated with the Person entity in RDA, and at that time its VES consisted of just three terms: "female", "male", and "not known". The constituency made a proposal to add a term for "transgender", using the "fast-track" process for amending RDA. This appeared simple enough at first glance, but complications began to appear as other RDA constituencies considered the proposal. Special interest groups for LGBT (lesbian, gay, bisexual, and transgender) cataloguers and collections advised that "transgender" was too broad to reflect the complexities of non-cisgender status, and that its use would be controversial. Suggestions were made for including more terms in the vocabulary, based on various categories used by a range of LGBT communities, but there was no consensus. In fact, there are different sets of terms for covering distinct situations such as changing gender using medical assistance or identifying with a gender by adopting appropriate social and cultural roles. At one point, it was suggested that over 20 new terms should be added to the RDA vocabulary.

The RSC began to look at the element and its VES more closely. The definition of "gender" in RDA is "a gender with which a person identifies". It was originally included in RDA because it is specified as an attribute of Person in Functional Requirements for Authority Data (FRAD), one of the IFLA models on which RDA was based. The FRAD definition is "A gender by which the person is identified (e.g., male, female, unknown, other)". The difference in these definitions is striking: RDA is reflecting a general unease with the ethics of assigning gender identity, and is attempting to clarify that it is the person described by the metadata who should be responsible for the assignation, rather than the cataloguer. The RDA VES also avoids the pitfall of using "other" as a term for someone's gender. (As an aside, the use of an "other" term in a controlled vocabulary essentially makes the vocabulary closed to the addition of new terms: "other" is dependent on the coverage of the other terms in a vocabulary, so amending or adding another term changes the meaning and introduces uncertainty into the "other" values used previously.)

Some practical questions immediately arise. How does a cataloguer know the gender a person identifies with? There are two principal sources of information: the manifestation in hand that is being described; and “reference sources” which are effectively authority control systems or some other structured, controlled set of data such as an encyclopedia. It is rare for a manifestation to specifically state gender in, say, a statement of responsibility. Gender may be implied, such as in the use of the honorific “Mrs” (a historic cultural norm that is no longer relevant), or in given names (a common current cultural norm for gender differentiation). What is the gender of “J.K. Rowling”? We know that this writer deliberately conceals their gender in their books, so why should a cataloguer over-ride this and assign the value “female” after consulting a reference source? The Library of Congress Name Authority File (LCNAF) assigns “Females” to this person; the term is taken from the Library of Congress Demographic Group Terms (LCDGT).

What does a cataloguer do if a person identifies with a gender that is not in the VES? The use of LCDGT falls within the RDA guidelines, which allow any appropriate VES to be a substitute for the RDA scheme. LCDGT includes “Transgender people” and “Transsexuals”. It is worth noting the inconsistent phrasing of these terms: “Transgender people” and “Transsexual people” are variant terms respectively. This may reflect recent advice from LGBT groups to use gender terms as adjectives rather than nouns. It could also be argued that this is not an appropriate VES to use because the vocabulary is for groups of people, not gender itself. RDA also allows the use of any other suitable term if one cannot be found in the RDA or another VES, but this is treated as uncontrolled data and is likely to vary between cataloguing agencies and policies.

What does a cataloguer do if existing metadata uses a value of gender that is different from what a person currently identifies with? The LCNAF authority record for Wendy Carlos, the composer of electronic music, assigns four different values of gender: “Females”; “Males”; “Transsexuals”; “Male-to-female transsexuals”. Which of these does the person currently identify with? Is it really more than one? There is a source note for a Wikipedia article explaining some of the background to these assignments. On the other hand, the LCNAF record for Jennifer Finney Boylan does not assign any value for gender, although a source note indicates a “sex-change operation”. It is Library of Congress policy to amend or delete such information at the request of the data subject, but it is doubtful that every living person described in LCNAF knows this. Taken to its logical conclusion, to be compliant with the semantics of the RDA element the Library of Congress should be contacting a large number of people on a regular basis to verify their gender identification. Does a cataloguer replace the existing value for gender, or add a new value? In linked open data applications, where nothing is deleted, does a cataloguer deprecate the old value after adding the new value?

In both cases the person changed their name at some point, replacing a “male” given name with a “female” given name. RDA is more comfortable with this situation, and there is an instruction that when a person changes their name, the latest form should be used as the preferred name. There is an exception if an earlier form persists as the name “by which the person is better known”, with an assumption that “better known” is determined from reference sources, so there is whiff of circular reasoning. The original name may be recorded as a variant name, but this is not mandatory. It is usually such a change of name that comes to the attention of a cataloguer, triggering an update of a name authority record. RDA makes no reference to determining gender from name, or checking other data if a name changes; this must surely be the responsibility of the cataloguer and cataloguing agency.

These issues are part of a wider discussion on the principle of “representation” or how a resource manifestation describes itself. This is one of the International cataloguing principles (ICP) published by IFLA. The relevant part of the definition is “*A description should represent a resource as it appears. Controlled forms of names of persons, corporate bodies and families should be based on the way an entity describes itself*”. Much of IFLA’s content standard, the International Standard Bibliographic Description (ISBD), is based on this principle (but, curiously, the principle itself is not explicitly mentioned in ISBD). RDA describes the principle as “*The data describing a resource should reflect the resource’s representation of itself*”.

These data are important for the user task of identifying a resource, that is, a manifestation, following a search: the metadata contain transcriptions of some of the text that is written on a manifestation (or spoken in an audio recording) for identification and reference. These texts typically include titles, statements of responsibility, and information about publication, manufacture, or distribution of the manifestation. Of course, this data is not just about the manifestation, but also covers persons or groups associated with the creation of the manifestation and the creation of the works and expressions that it embodies.

Who creates this information? The creator of the work that is embodied in a manifestation may often specify the title of the work and the form of their name to be used in statements of responsibility, but in many cases it is the publisher who decides. It is not safe to assume that a name given in a statement of responsibility is “the way a [person] describes [themselves]”. For example, many publishers of scholarly works, including institutional repositories, have policies restricting given names to initials. There seems to be an inherent contradiction in the ICP, unless we make a clear articulation between the name on a manifestation and the name in an authority file. This is not an issue for ISBD, which does not cover controlled forms of name. ISBD allows names not to be recorded in a statement of responsibility. For example, if there are many names only the first few may be recorded at the discretion of the cataloguing agency, with the remainder replaced by “[et al.]”. RDA follows this practice, albeit with “and others” rather than the Latin abbreviation. RDA currently offers many instructions on using a manifestation representation as the basis of a controlled name in an access point; these are under review in the RDA Toolkit Restructure and Redesign (3R) Project which has introduced distinct elements for “name of person” and “access point for person” to clarify the articulation.

If a person represents their name on manifestations of their work in a gender-free way, as with “J.K. Rowling”, is it right for a cataloguer to add “Joanne Kathleen” to the authority record as a value for fuller form of name? As already noted, given names are differentiated by gender in many cultures. The LCNAF record also contains a “see also” reference to “Robert Galbraith”, who is presented as a separate person with gender, you guessed it, “Males”. Yet source notes in both records indicate that they are the same person using a pseudonym. Presumably J.K. Rowling wants the pseudonym to be identified as a male, otherwise it would be “R. Galbraith”. Much of this practice is derived from FRAD, which defines the Person entity as “*An individual or a persona or identity established or adopted by an individual or group*”, with a scope note “*Includes literary figures, legendary figures, divinities, and named animals as literary figures, actors, and performers*”. This explains another “see also” reference from J.K. Rowling to “Newt Scamander” in LCNAF. Newt Scamander is a fictional author of a book mentioned in the Harry Potter novels. The book was subsequently created by J.K. Rowling and published in the real world with its fictional title “*Fantastic beasts & where to find them*”, “Newt Scamander” as a statement of responsibility, and a fictional publisher. Later editions add a second statement of responsibility, “by J.K. Rowling”. Newt Scamander is treated as a pseudonym with a separate entry in LCNAF, but is not assigned the fuller form of name “*Newton Artemis Fido*” or the gender “Males” (unlike Robert Galbraith). In fact, name authority control is full of inconsistencies in the treatment of fictional persons. For example, the LCNAF entry for Arthur Conan Doyle does not give “John H. Watson” as a pseudonym, even though most of the Sherlock Holmes stories are purported to have been written by Dr. Watson. This is despite an entry for “Watson, John H. (Fictitious character)”, with gender “Males”, in LCNAF. Scamander is not identified as a fictitious character.

As FRAD points notes “*Some cataloguing rules recognize pseudonyms used by an individual as separate personas and an authorized form will be established for each persona with see also references connecting the resulting authorized forms of names for those personas. Other cataloguing rules recognize only one persona for an individual and the pseudonyms used by the person; one persona is the authorized form and the names of other personas are treated as variant forms of name for the individual.*” The difference reflects the focus of an authority control system: names of persons, or the persons themselves. LCNAF is appropriately titled!

The IFLA Library Reference Model (LRM) consolidates and supersedes FRAD and the other “functional requirements” data models, and narrows the definition of Person to “*An individual human being*” scoped as “*restricted to real persons who live or are assumed to have lived*”. The 3R Project seeks to maintain and develop its compliance with the LRM, and the RSC Fictitious Entities Working Group has been investigating the impact of this change on current cataloguing practices.

Does any of this matter? Suppose I am a male writer and my given name is “Patrick”. I provide text to accompany illustrations in children’s books. I live in a country with strong religious views on sex and gender and an antipathy to the very idea of “transgender”. But I am what I am, and eventually I change my given name to “Patricia”. After a while, the flow of new work starts to dry up. I have asked my national library to change my name in their metadata, and they are happy to replace “Patrick” with “Patricia” in the “authorized” form my name. The old version is given as a variant name in the authority file, but this is only visible if a catalogue user chooses to look at the authority record. In most cases, a prospective client, looking for works that I have contributed to, will use the form that I have given them, and find a list of my books without displaying the authority record. But then that client sees the statement of responsibility for each manifestation, and the earlier ones all refer to “Patrick”, while the “heading” refers to “Patricia”. I did not foresee the change of gender and name, and my publisher does not like to use initials because it looks less child-friendly, so I was happy for the name given by my parents to be printed. Unfortunately, the prospective client is risk-averse and wants to avoid controversy; I don’t get the commission. What can I do? The national library is sympathetic to my case, but says it cannot remove old statements of responsibility because the international standard RDA says that the first statement should always be recorded, and I usually get first billing as writer, before the illustrator. The library also points out that it would not be feasible for the publisher to recall and reissue all extant copies of my books. What about my right to “be forgotten” in this respect?

I understand that I can’t go back and change my old books, and unlike Wendy Carlos, there is little opportunity to re-publish the content in a new format. But the metadata records can be changed; they were retro-converted to machine-readable form a few years ago, and it would be easy for the library to amend the manifestation metadata now. I don’t even want “Patrick” to be changed to “Patricia”; using just the initial “P.” will help resolve my problem. Would this compromise the principle of representation?

In practice, library metadata content standards adhere to the principle by providing instructions on transcribing text that the manifestation uses to describe itself. This is not straightforward. Irrelevant parts should be omitted; sections of text may need to be re-ordered to fit standard display requirements; capitalization and punctuation may be changed for readability. ISBD allows the addition of text that does not appear in the manifestation. For example, “[sic]” may be interpolated to indicate that the transcription reflects a spelling error, and not a mistake. RDA prefers to make extrinsic notes rather than add anything to “representation”, and is being developed to accommodate a wider range of transcription rules, from ISBD to What-you-see-is-what-you-get derived from the application of digitization and optical character recognition software. Thus it seems that libraries could be more flexible on what can be omitted from a “manifestation statement” (as the LRM, and now RDA, refer to “representation” data). Maybe the national library will reconsider Patricia’s case in due course.

Why do content standards cover gender in the first place? RDA does it because FRAD did it; FRAD did it because established name authority control systems do it. The primary purpose is to disambiguate headings for persons that are otherwise similar: “Bozo the Clown (male)” and “Bozo the Clown (female)”. That’s why many other personal data are recorded: date of birth, fuller form of name, profession or occupation, and so on. This can be helpful for a name like “J. Smith”, but gender alone is insufficient: “J. Smith (male)” and “J. Smith (female)” do not achieve functional utility. The data are often recorded “just in case” someone else with the same name(s), date of birth, etc. crops up and the data become useful. In most cases, gender data is never used for this purpose. However, it is difficult to see how the gender data for Wendy Carlos can be used in this context.

The RSC discussed other uses of gender data. National libraries use it to answer queries on the national bibliographic corpus: how many female writers were published outside London between 1880 and 1900? This is important information for social and historical research. A rich catalogue can be used as an encyclopedia; conversely, Wikipedia is good for finding the works of J.K. Rowling. Linked open data further blurs the boundaries, but also reduces the need for duplicate recording. If gender is recorded in Wikipedia and other linkable datasets, why do cataloguing agencies need to create and maintain it?

Why is gender different from ethnicity or religion? It would be very useful for social and historical research if a national library could produce an answer to queries such as: how many white female Jedi Knights were published between 1995 and 2015? But we don't ask cataloguers to record this information for many reasons, ranging from privacy issues to the difficulty in obtaining the data. Perhaps we should not be recording gender information just because it seems easy to do so; as we have seen, keeping it accurate isn't so easy.

Let's not talk about sex.

It did not take long for the RSC to decide, after considering these issues, that rather than adding terms to a preferred vocabulary for gender, it would be better to have no preferred vocabulary at all, so the RDA Gender VES was removed from RDA Toolkit. The terms are now listed in an "additional resources" section in the Toolkit, for legacy purposes. RDA instructions allow any controlled or uncontrolled terms to be used as a value for the gender element, and the burden of using and maintaining gender data is now the responsibility of cataloguing agencies, who can add policy statements or application profiles to reflect local practice.

In the second part, we'll talk about separating fact from fiction.

gordon@gordondunsire.com

Further reading

Wikipedia has articles on "LGBT", "transgender", and "transsexual", available at <https://en.wikipedia.org>

Amber Billey, Emily Drabinski & K. R. Roberto (2014) What's Gender Got to Do with It? A Critique of RDA 9.7, *Cataloging & Classification Quarterly*, 52:4, 412-421, DOI: 10.1080/01639374.2014.882465

Further information about IFLA standards, including FRAD, ICP, and LRM, is available at <https://www.ifla.org/node/8750>

Presentations on the 3R Project and recent development of RDA are available at <http://www.rda-rsc.org/rscrepresentations>

Further information about the RSC, working groups, etc., is available at <http://www.rda-rsc.org/>