

AUTHORITY VERSUS AUTHENTICITY THE SHIFT FROM LABELS TO IDENTIFIERS

Gordon Dunsire

Consultant, Edinburgh, Scotland, UK

Mirna Willer

*Department of Information Sciences,
University of Zadar, Croatia*

KEYWORDS:

*authority control,
authenticity, Semantic
Web technologies, Linked
Open Data, entity person,
entity nomen, IFLA
Library Reference Model
(IFLA LRM)*

ABSTRACT

This paper describes first the concepts underlying the Semantic Web technology and Linked Open Data environment such as reification, the AAA principle, provenance, and the Open World Assumption with the aim to explore the need to make explicit the social constructs and “social machines” underlying Linked Open Data in the archives, libraries, and museums (ALM) environment, to support and benefit from the shift from human-readable labels to machine-processable identifiers.

The general question, then, is how to make explicit the social constructs, the rules that govern and underpin data production for consumption by the human user, to the “social machines”, the rules that govern and underpin data production for machine processing and semantic reasoning in applications aimed at the human user.

Assuming that this technology will be the next to be adopted for library catalogues, and in the context of the new IFLA conceptual model (IFLA LRM) for library metadata developed with “semantic web technologies in mind”, the paper analyses the choice and

form of author's name in the case of multiple identities as a case study.

The issues raised by the case study are resolved by separating labels (nomens) from the categorization of entities of interest (persons). A "personal" name is insufficient evidence that the entity labelled is a person. Conversely, every name in a statement of responsibility is the label of a person or group of persons, according to the IFLA LRM. Global interoperability and re-use of metadata recorded by different "official" agencies is better served by semantic coherence in the categorization of entities and relationships of interest than by relying on labels and manifestation statements. Semantic Web technologies support this approach, especially in the context of social construction and deconstruction. The AAA principle and Open World Assumption imply an increasing importance for the role for official, national cultural heritage organizations as providers of trustworthy provenance.

Introduction

The main goal of the paper is to explore the need to make explicit the social constructs and “social machines” underlying Linked Open Data in the ALM environment, to support and benefit from the shift from human-readable labels to machine-processable identifiers.

With that in mind, concepts underlying the Semantic Web technology and Linked Open Data environment such as reification, the AAA principle, provenance, and the Open World Assumption are described first. Then, assuming that this technology will be the next to be adopted for library catalogues, and in the context of the new IFLA conceptual model (IFLA LRM) for library metadata developed with “semantic web technologies in mind”, the paper analyses an old cataloguing issue: the choice and form of author’s name in the case of multiple identities.

Design issues for Linked Open Data

Linked Open Data (LOD) is the machine-processable data that populates the Semantic Web. The Semantic Web is based on the World Wide Web invented by Tim Berners-Lee, who states:

“The Semantic Web isn’t just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.

Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they [are] links between arbitrary things described by RDF. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs, so that they can discover more things.”¹”

These are best practice recommendations for the design of linked data. Every thing of interest, from tangible animate or inanimate object to abstract concept, is identified by one or more Uniform Resource Identifiers (URI). These are machine-processable identifiers that are unique in the domain of the World Wide Web and Internet.

URIs are linked by relationships that are themselves identified by a URI; this allows links to be established between relationships so that ontologies describing classes of things and the relationships between them can also be recorded as linked data. Well-formed ontologies provide machine-processable semantics for the data about individual things.

Internet and Web technologies are used to de-reference a URI by returning additional data about the thing in response to an online request. Good de-referencing services offer the data as a human-readable web page in hyper-text markup language (HTML) format, or as machine-processable data in Resource Description Framework (RDF) format.² RDF encodes data as a set of statements in subject-predicate-object format. Each statement, or “triple”, is a value (object) associated with an aspect or characteristic (predicate) about a thing (subject).

There are three other fundamental concepts that should be considered in the Linked Open Data environment: the AAA principle, provenance, and the Open World Assumption.

The AAA principle that “Anyone can say Anything about Any thing“ means that there is no intrinsic test of truth for the information give in a triple. It can only be tested for semantic consistency with other statements.

1 Tim Berners-Lee, *Linked data - design issues*, Date: 2006-07-27, last change: Date: 2009/06/18, <http://www.w3.org/DesignIssues/LinkedData.html>.

2 W3C. *Resource Description Framework (RDF)*, <https://www.w3.org/RDF/>.

For example, this set of four statements is always inconsistent:

1. Thing1 is-a ClassA
2. Thing2 is-a ClassB
3. Thing1 is-same-as Thing2
4. ClassA is-disjoint-with ClassB

The first three statements are data about two individual things, and the fourth statement is from an ontology. If ClassA is Dog, the class of things that are dogs, and ClassB is Cat, things that are cats, then the first three statements say that some thing identified by the URIs Thing1 and Thing2 is both a cat and a dog. The fourth statement says that no thing can be both a cat and a dog. At least one of the statements must be false, but it is impossible to determine which without further information in the form of additional triples.

In this case, most humans “know” that the fourth statement is true, so the inconsistency lies in the first three statements. The fourth statement is from an ontology that defines ClassA and ClassB as a reflection of the real world. However, the ontology could be describing a fictional world in which dogs and cats are the same species. The best that application software that uses this data can do is indicate an inconsistency; a human is required to resolve it, by flagging one or more of the statements as “false” and to be ignored. The flag itself has to be stated as a triple, with no indication of its own veracity.

The provenance of Linked Open Data is important because in the Semantic Web there are competing data from many different sources, including social networks, publishers and sellers, governments, propagandists, and so on. A triple may be “false” because of human error, deliberate misinformation, legacy context and practice, or change over time. Thus key questions for evaluating any Linked Open Data statement are “Who said that? , When was it said?, How was the value of the triple determined?, What was the context?”.

This extends the concept of provenance from archival and museum resources to Semantic Web triples. A triple is an information resource in its own right; it has a creator, a time of creation, a method of production, and a

history. Such information about a triple can be stated as another triple with the first triple as its subject. This requires the identification of the first triple with a URI, a process of reification that treats the triple as a thing.

For example, the following triple can be reified by assigning a URI to the whole statement, such as Statement1:

- Resource1 has-date-of-production “1951”
- Then the provenance of the triple can be given as:
- Statement1 has-creator Person1
- Statement1 has-date-of-production “2016”
- Statement1 has-source SourceA

Linked Open Data from archives, libraries, and museums is likely to be of much higher quality than the data from many other source, assuming they follow international, national and/or consistent (identified) local rules, standards, and vocabularies that are intended to support an ethos of neutrality and trust.

The Open World Assumption (OWA) is that

1. the absence of a statement (an RDF triple) about some thing is not a statement of absence; for example, no value for date of birth does not imply a Person was not born;
2. statements may be made about any thing in the future because:
 - Knowledge is always incomplete
 - Something new can always be stated, whether it is true or false
 - The AAA principle.

These assumptions are not made in a closed application domain or “world” of fixed sets of statements, for example using a metadata record schema where a blank value may indicate “not applicable”, “unknown” or “nothing” depending on context. Closed-world applications may also use a schema structure to distinguish sub-groups of statements, although the grouping categories are things in themselves.

In the global open world of the Semantic Web every thing must be iden-

tified by a URI. This means that closed-world sources must replace “blank nodes” (not identified by a URI) used to cluster statements inside a record, for example for a compound access point, with URIs before they can be used as Linked Open Data by the Semantic Web.

RDF is a mathematical graph of nodes (representing things) and links (representing relationships between things). The Semantic Web is expressed as a single graph of all Linked Open Data, a Giant Global Graph.³ Applications covering a wide range of things for a wide range of users are unlikely to be interested in statements with fine semantic granularity reflecting local contexts. Specific granularity can be maintained for local applications and mapped to a coarser granularity for global applications. In this process of “dumb-down”, the finer contextual semantics are replaced with broader categories of things and relationships, and detail is lost.

IFLA Library Reference Model

In February 2016, the FRBR Review Group, a standing group of the Cataloguing Section of the International Federation of Library Associations and Institutions (IFLA), published the world-wide review draft of the *FRBR-Library Reference Model*, a consolidation of three related but separately published conceptual models for the “bibliographic universe”: Functional Requirements for Bibliographic Records (FRBR), Functional Requirements for Authority Data (FRAD), and Functional Requirements for Subject Authority Data (FRSAD).⁴ A final draft based on comments from the review and subsequent discussion by the Review Group was published in March 2017 while awaiting formal approval from IFLA, which eventually happened in August 2017.⁵ The renamed model, *IFLA*

3 *Giant Global Graph*, https://en.wikipedia.org/wiki/Giant_Global_Graph.

4 Pat Riva, Patrick Le Boeuf and Maja Žumer, *FRBR-Library Reference Model*, 2016-02-21, Draft for the World-Wide Review (International Federation of Library Associations and Institutions), https://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf; information on the world-wide review, <https://www.ifla.org/node/10280?og=587>.

5 Pat Riva, Patrick Le Boeuf and Maja Žumer, *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*, December, 2017 (Den Haag: IFLA), https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf.

Library Reference Model states in its scope and objectives, “The IFLA Library Reference Model aims to be a high-level conceptual reference model developed within an enhanced entity-relationship modelling framework. The model covers bibliographic data as understood in the broad general sense.”⁶ The model, like the three previous ones, is based on the entity-relationship methodology however:

“The context has changed since the FRBR model was originally developed, and new needs have emerged, particularly in terms of reuse of data in semantic web applications, making this consideration an integral part of the initial planning of presentation of the model definition.”⁷

Two issues that raised a particularly fervent discussion during the review and revision process were the definition of the entity *person*, and the concept of Representation.⁸ Both issues are crucial for bibliographic information organisation in as much as they ensure three basic functions of the catalogue:

- identification of a resource,
- collocation of information about the author and/or work,
- navigation between entities.⁹

In this paper, we will focus on the entity *person*. LRM-E7 Person is defined as a subclass of the entity LRM-E6 Agent, and as “An individual human being”. The Scope notes state:

“The entity *person* is restricted to real persons who live or are assumed to have lived.

6 Ibid., 9.

7 Ibid., 12.

8 Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, Mass.; London: MIT Press, 2000), 71: “The principle of representation requires bibliographic descriptions to be constructed to reflect the way bibliographic entities represent themselves. This principle is used primarily to ensure accuracy of description, though it is used as well to contain costs, to prevent idiosyncratic descriptions, and to assist in the construction of operational definitions.”

9 Ibid., 17-20.

Strict proof of the existence of a *person* is not required, as long as there is a general acceptance of their probable historicity. However, figures generally considered fictional (for example, Kermit the Frog), literary (for example, Miss Jane Marple) or purely legendary (for example, the wizard Merlin) are not instances of the entity *person*.¹⁰

The objection to this definition, and, in some views, a dysfunctional restriction to the original FRBR and FRAD definitions was the content of the last sentence. It seems to break the natural association of a name with the manifestation on which it appears. The FRAD model¹⁰ for authority data defined *person* as “An individual or a persona or identity established or adopted by an individual or group.”¹¹ This was a reflection of some bibliographic and cataloguing traditions in which fictional, literary, or legendary entities and identities were considered and represented in access points as authors. It should also be noted that, according to FRAD, *person* can be also an identity which can be “established or adopted” not only by a person, but also by a group of persons. Restricting the entity *person* to a “real human being”, a type of *agent* (defined as “An entity capable of deliberate actions, of being granted rights, and of being held accountable for its actions”), the only thing that is intellectually (and/or spiritually¹²) capable of creation of a work, appeared in many comments to be extremely problematic. The concept of identity established by a group was only discussed briefly.

It is difficult to deal with this issue without taking into account its context, introduced by an attribute of the entity *manifestation*, LRM-E4-A4 Manifestation statement, and the distinction between the process of transcribing data from the resource and the process of recording data. Transcribed data may contain accidental or deliberate errors, or may be

10 *Functional Requirements for Authority Data* (Den Haag: IFLA, 2009), <http://www.ifla.org/publications/functional-requirements-for-authority-data>.

11 *Functional Requirements for Authority Data: A Conceptual Model*, Final Report December 2008, As amended and corrected through July 2013, http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf, 8.

12 The current Croatian cataloguing rules define an author in the following way: “Individual author of a work is a physical person to whom the work belongs as a spiritual ownership”. See: Eva Verona, *Pravilnik i priručnik za izradbu abecednih kataloga: Dio 1: Odrednice i redalice*, 2. izmijenjeno izd. (Zagreb: Hrvatsko bibliotekarsko društvo, 1986.), Article 4, p. 18.

incomplete or ambiguous; recorded data are corrected, completed, and disambiguated by using multiple reference sources¹³ The entity LRM-E4 Manifestation is defined as “A set of all carriers that are assumed to share the same characteristics as to intellectual or artistic content and aspects of physical form. That set is defined by both the overall content and the production plan for its carrier or carriers”, while its attribute, LRM-E4-A4 Manifestation statement as:

“A statement appearing in exemplars of the *manifestation* and deemed to be significant for users to understand how the resource represents itself”. Its Scope notes state the attribute “... is a statement normally transcribed from a source present in exemplars of a *manifestation*. Transcription conventions are codified by each implementation.”

The transcription of data thus ensures the representativity of the *manifestation* itself which supports the task of identification by the user. The recording of data based on the *manifestation statement*, on the other hand, must support the second function of the catalogue, of collocation. Fulfilling this task has to take into account other aspects of the resource and its context.

We will focus on the collocation of all of the works by a personal author and use a specific example to illustrate the issues. As already mentioned, these are contentious.

Example of Newt Scamander

In this section we discuss the question of the choice and form of author’s name in the case of multiple identities, using as an example the work of fiction titled “Fantastic Beasts & Where to Find Them”.

The “origins” of this book, a work of fiction, should be explained: Har-

13 See the discussion in: Gordon Dunsire, *RDA data capture and storage*. Presented to Committee on Cataloging: Description and Access II (CC:DA) - ALCTS CaMMS, ALA Midwinter 2016, 11 January 2016, Boston, Mass., <http://www.gordondunsire.com/pubs/pres/RDADDataCap.pptx>.

ry Potter, the main character of a series of eponymous novels, is a pupil of the Hogwarts School of Witchcraft and Wizardry in which he studies, amongst others, the book *Fantastic Beasts & Where to Find Them*, written by Newt Scamander and published by Obscurus Books, 18a Diagon Alley, London. The book, its author, publisher, and publisher address are all part of the fictional world of the novels.

As it appeared that such a book could be of interest to a wider public – the Muggle population as it is referred to in the Harry Potter series – the author of the novels, J. K. Rowling, created it as a separate work that was published in 2001, with a second edition in 2009. For a *manifestation* that is a printed volume the information that it presents about itself is usually found on the title page and its verso. The title page of the 2009 edition of *Fantastic beasts* has a pure textual transcription (Figure 1):

“FANTASTIC
BEASTS &
WHERE TO
FIND THEM

NEWT SCAMANDER

BY
J.K.Rowling

Bloomsbury
in association with
Obscurus Books
18a Diagon Alley, London”

The title page verso states:

“First published in Great Britain in 2001 by
Bloomsbury Publishing Plc
...
This edition published in 2009”

FIGURE 1 Textual transcription of the title page and the back of the title page of the book *Fantastic Beasts & Where to Find Them*, edition published in 2009

It should be noted that the first edition of 2001 has additional information on the title page, as represented in the British Library catalogue: “Newt Scamander ; special edition with a foreword by Albus Dumbledore”.

The author of the book is traditionally given by the information on the title page that is preceded by the preposition “by”, and in addition by the copyright statement usually found on the verso. In this case the title page’s “J.K.Rowling” is confirmed by the copyright assigned for the text, illustrations and hand lettering on the title page verso. However, what is the role of the name “Newt Scamander” given on the title page immediately below the title? Could a standard publisher’s layout of the title page help us here? Elaine Svenonius, discussing how the perceptions of authorial functions over time and authorship has become “increasingly diffuse and mixed”¹⁴ concludes:

“In those cases where it is useful to identify the author of a document, a constructive (as opposed to a deconstructive) approach is to capitalize on the fact that a thing is usually what it represents itself as being. Lubetzky does this when he defines *author* as “the person or corporate body *represented* as chiefly responsible for the work, i.e., the one in whose name the work is issued and who is purportedly responsible for it ... except when one has erroneously, fictitiously, or dubiously been represented as the author of the work.” This definition is open-ended in that it allows the various ways in which an author may be represented to be explored. This brings definition (again) around to enumeration, with the intriguing implication that no common or essential component may exist in all instances of authorship.”¹⁵

In considering the key to automatic author identification, in order to be able to “enumerate the various ways in which an author may be represented”, Svenonius draws from linguistic and location criteria:

14 Svenonius, *The Intellectual Foundation of Information Organization*, 44.

15 *Ibid.*, 45. The citation is from: Seymour Lubetzky, *Principles of Cataloging: final report: Phase I: Descriptive cataloging* (Los Angeles: University of California, Institute of Library Research, 1969), 29.

“Certain character strings (such as *by*) usually identify authors, while other strings (such as *edited by*) are likely to identify nonauthors. Moreover, authors’ names are likely to appear in key locations, such as in the top third of a title page.”¹⁶

In our case, we have both situations: one name represented in the top third of the title page – Newt Scamander, and another introduced by the character string *by* – J.K. Rowling. Such a situation is open to interpretation by the cataloguer when recording the name of the author by which the book (as a *manifestation*) would be identified by a user, and in deciding under which author’s name the work would be collocated with other editions or adaptations of the same *work*, or with other works by the same author! Who should the cataloguer assume that the user/reader would search for in the catalogue? Who does the reader/user consider the author to be, and, consequently, who should the reader assume they should search for: Newt Scamander or J.K. Rowling? Also, does the reader consider these to be the same person or different persons? A reader may hold both opinions, albeit in specific contexts: Rowling wrote the actual book, and Scamander wrote the fictional equivalent. This is second-guessing on a grand scale.

What does the perception of the reader match?

- the perception of the cataloguer in following particular cataloguing rules for organizing bibliographic information,
- the perception of the publisher in laying out the title page and giving publishing information,
- the perception of the author in their wish to be represented by a specific name or to assume a specific bibliographic, public identity?

The general question, then, is how to make explicit the social constructs, the rules that govern and underpin data production for consumption by the human user, to the “social machines”, the rules that govern and under-

16 Ibid.

pin data production for machine processing and semantic reasoning in applications aimed at the human user.

Cataloguing as transcription

Most national library cataloguing rules follow IFLA’s *International Standard Bibliographic Description* (ISBD) standard¹⁷ that expects the cataloguer to uniformly transcribe values for “Title proper” and “Statement of responsibility” attributes from the resource in hand; that is, an *item* as representative of its *manifestation* in the IFLA LRM model. The information transcribed in these two data elements provides guidance to the cataloguer for the next step: to record the author/creator of the work, that is, to establish authorship for a work so that the user will be able to identify the author and find collocated all their works in a catalogue.

All data transcribed from a *manifestation* is how “the resource represents itself”, so these attributes are types of LRM-E4-A4 Manifestation statement. In particular, authorship is usually evident in the “statement of responsibility”. It is useful to explore the values recorded in national library catalogues for the Statement of responsibility element of the original 2009 edition, and translations of both editions.

For the original 2009 edition:

- the British Library records “Newt Scamander [i.e.] by J.K. Rowling”
- ICCU, the Italian union catalogue records “by J.K. Rowling ; [introduction by] Newt Scamander

These two examples show two different transcriptions of the same title page information showing two interpretations (following two different cataloguing rules) by two cataloguers and therefore applying different semantics to the data. It is the difference in the cataloguers’ output that is

17 *ISBD: International Standard Bibliographic Description*. Consolidated ed. (Berlin/Boston: De Gruyter Saur, 2011); *ISBD Consolidated edition*, March 2011 (pre-print), http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf.

relevant, not the accuracy of an interpretation that is dependent on the clarity and specificity of the rules. The transcribed data is a transcription with the cataloguer's interpolation:

- [i.e.]: “Newt Scamander”, the name found on the title page in a position signifying the name of the author, is not the “real” name of the author, which is “J.K. Rowling”;
- [introduction by]: the name of the author of the *book* to which “Newt Scamander” contributed an introduction, is “J.K. Rowling”, indicated on the title page with “by”. This relegates the question of how to deal with the name “Newt Scamander”, to which the Italian cataloguer gave the credit of the writer/author of introduction, to some other considerations as the writer of the introduction would only under specific circumstances be considered worthy of bibliographic (authority) control (in Svenonius's terminology a “non-author”; see discussion about this example under For translations below);

Note: Square brackets, according to ISBD, are used to “enclose information found outside the prescribed sources of information and interpolations in the description.”¹⁸

There is also variation in the number of statement of responsibility elements:

- in the BL record, there is only one data element in line with the interpretation about “who” the names belong to, i.e., who this data element represents;
- in the ICCU record, the statement of responsibility is divided into two sequential data elements, indicated by the ISBD punctuation “_;_”, in line with the interpretation of assigning roles to the “named” entities.

18 Ibid., A.3.2.8. These are two distinct categories of added information. Use of the same mark-up of square brackets can only provide the more general indication that the data is not present on the resource itself.

For translations:

- German translation statement of responsibility records “Newt Scamander”; we can assume that this is a translation of the 2001 edition, because there is a reference to the introduction by Albus Dumbledore that was not present in the 2009 edition;
- Italian translation published in 2010 records “dove trovarli [di] Newt Scamander / J.K. Rowling”; a subtitle with cataloguer’s interpolation and a single statement of responsibility.
- Italian translation published in 2015 records a statement of responsibility “di J.K. Rowling ; Newt Scamander”. In this translation the publication statement is “Milano : Salani ; London : Obscurus Books, 2015”; with a Note that “Newt Scamander e Obscurus books sono nomi di fantasia”.
- Spanish translation statement of responsibility records “Newt Scamander ; por J. K. Rowling”.
- French translation statement of responsibility records “Newt Scamander [i.e. J. K. Rowling]”.
- Croatian translation statement of responsibility records “Newt Scamander”.

The representation of the work and names, or character strings, of “authors” by publishers of translations into national languages, and the transcription of the information from the title page varies considerably, and it can be assumed that it corresponds with local and national publishing and cataloguing traditions. The publishers assume what their readership expects, while cataloguers follow their cataloguing rules, which prescribe in what manner they have to resolve particular publishers’ resolutions.

What about the reader? The reader who wants to obtain “Newt Scamander’s book” is the least “deceived” or, more correctly, directly served by German, French and Croatian publishers who present Newt Scamander as the author of the book. Each cataloguer’s decision in those cases, however, was different: the French cataloguer intervened in transcribing the statement of responsibility from the title page by making the interpolation explaining to the reader

who the “real author” was; the German and Croatian cataloguer did not make any similar interpolation in the transcription of the title page; while the Croatian cataloguer recorded in a note the real name of the author. Faithful to the original edition, the Italian and Spanish publishers represented the two names on the title page. However, the Italian publisher, added to the 2015 edition the “original” publisher, Obscurus Books. The Italian cataloguer varied in applying the cataloguing rules in transcribing the statement of responsibility: for the 2010 edition “Newt Scamander” is transcribed as part of the subtitle with the interpolation of [*di*], while in the 2015 edition the name is recorded as the second statement of responsibility, with “J.K. Rowling” as first. In the case of the Spanish edition, the names are transcribed from the title page in the original order, but treated as first and second statement of responsibility.

Differentiating the names on the title page as first and second statement of responsibility implies that the names are considered to be of a different nature or function in terms of type of responsibility for the work. However, which name relates to the principal responsibility for a work, and which to the other, cannot be assumed from the transcription data alone. Namely, the ISBD stipulations as to transcribing the statement of responsibility from the preferred source of information prescribe:

“1.4.4.2 More than one statement of responsibility occurs when the wording shows multiple statements, as when more than one person or corporate body is represented as performing different functions and the statements are not linked by a conjunction.”

And:

“1.4.4.3 The difference between the first and subsequent statements of responsibility is merely a matter of order. It does not imply that the first statement relates to the principal responsibility for a work.”

We can assume that the two names as represented on the title page, given

the intention of the publisher and their view of how to meet a purchaser's expectations, are differently treated by the cataloguer in respect to the transcription of the data, the interpretation as to the naming of the "author of the book", and to giving different functions to a particular name. This shows to what extent particular national cataloguing rules and local cataloguer's practices diverge from international agreements to support Universal Bibliographic Control, the aim of which is to make a record of a publication once, at its origin, in order for it to be economically and efficiently re-used as necessary.

Cataloguing as recording: authorship and issues in authority control

One of the traditional "unwritten laws" of bibliographic description states that the descriptive part of the record should correspond to the choice of access points by which the catalogue is organized as a coherent structure and information tool; and *vice versa*. The descriptive part is, in our case, the transcribed statement(s) of responsibility, and, eventually notes that would give some further explanation to the reader. National cataloguing rules go into great depth in how to treat this issue, although it is untenable in a linked open data context to keep together distinct triples recording description and access data. It is also expensive to transcribe an extensive statement of responsibility, for example from a research paper with multiple authors or a motion picture with full credits, and supply an access point for every person or group that is named.

If we look at the choice of author's name as the access point in relation to the transcribed statement(s) of responsibility from the title page (Table 1), in all the examples except for the German and Spanish ones the name of the *person* J. K. Rowling is displayed as the author. Note that in all cases a space is inserted between the initials, although the author habitually uses the "J.K." form to avoid implying that these are the initials of her "real" given names. This implies that "J.K.Rowling" is intended to be a pseudonym, but the standard methods of authority control do not accommodate this situation.

The form of the author's name displayed for the user differs: J. K. Rowling; Rowling, J. K.; Rowling, Joanne Kathleen, 1965-....; Rowling, Joanne Kathleen. Searches under Scamander in these catalogues retrieve different set of records: they either display only records in which Scamander is found on

the title page (BL) or display all records related to the author J. K. Rowling.

Furthermore, the access point in the German catalogue is “Scamander, Newt”, while in the Spanish catalogue it is “Scamander, Newt (1897-),”¹⁹ which means that according to the German and Spanish cataloguing rules the name of this fictional character has gained a bibliographic, public identity of a *person*. Records retrieved under this name are related only to those cases in which the name is found on the title page. Works written by the *person* J. K. Rowling are retrieved under that name.

TABLE 1 Statement of responsibility element and access point for Newt Scamander and J.K. Rowling

Catalogue ²⁰	Statement of Responsibility	Access Point	Additional Access Point	VIAF / National authority file
<i>Original edition</i>				
BL	[i.e.] by J.K. Rowling	Rowling, J. K.		(LC) R., J. K.; <i>see also</i> S., N.; BL: search under S., N. 2 records found
ICCU	[introduction by] Newt Scamander	Rowling, J. K.	Rowling, J. K. Scamander, Newt	[No record in VIAF and in <i>voci di autorità</i> ; all 3 records found when searched under S., N.; R., J. K. collocates also other works]
<i>Publications of translations</i>				
DNB	Newt Scamander	Scamander, Newt		S., N. [all works under this name]; <i>Wirklicher Name</i> : R., J. K. [under R., J. K. only works under this name]

19 One can ask: At what point, if ever, does a death date occur, or is Newt immortal? This can cause havoc in the processing of “real-world” knowledge, e.g. statistics on the average age of authors, life-spans, copyright information, etc.

20 BL, <http://www.bl.uk/>; ICCU, <http://www.sbn.it/opacsbn/opac/iccu/free.jsp>; DNB, http://www.dnb.de/DE/Home/home_node.html; BNE, <http://www.bne.es/es/Inicio/index.html>; BnF, <http://catalogue.bnf.fr/index.do>; NUL, Zgb, <http://www.nsk.hr/>.

DNB	Newt Scamander	Scamander, Newt		Same as above DNB
ICCU, 2010	: [di] Newt Scamander ; J.K. Rowling	Rowling, J. K.		Same as above ICCU
ICCU, 2015	di J.K. Rowling ; Newt Scamander	Rowling, J. K.	Rowling, J. K. Scamander, Newt	Same as above ICCU
BNE	Newt Scamander ; por J. K. Rowling	Scamander, Newt (1897-)		S., N. [2 records found]; <i>see</i> Scamander, Newton Artemis Fido, 1897-; <i>see also</i> R., J. K. [under R., J. K. only works under this name]
BnF	Newt Scamander [i.e. J. K. Rowling]	Rowling, Joanne Kathleen (1965-)		R., J. K.; <i>see</i> S., N. [all records under R., J. K.]
NUL, Zgb	Newt Scamander	Rowling, Joanne Kathleen		R., J. K.; <i>see</i> S., N. [all records under R., J. K.]

This short analysis of cataloguing practices represented in some national library catalogues is just one more example of a known issue that has been discussed theoretically and for which solutions have been sought for decades at the international and national level.²¹ The goals remain the same throughout the evolution of the information technologies that lie behind catalogue construction: identification of a resource, collocation of information about the author and/or work and navigation between entities; the only difference now is that resolution is required not just in the local catalogue, but at global level for “universal” catalogues.

One of the solutions was found in establishing an international database that would help cataloguers in establishing the author’s form of

21 See for example: Mirna Willer, “Name Authority Control Paradigm Shift in the Network Environment,” in *Frameworks for ICT Policy: Government, Social and Legal Issues*, edited by Esharenana E. Adomi (Hershey, PA: IGI Global, 2010), 182-205, DOI: 10.4018/978-1-61692-012-8.ch012.

name at national level – a kind of riposte to the Universal Bibliographic Control programme. VIAF, the Virtual International Authority File,²² is the international union catalogue of authority data started by the Library of Congress and the Deutsche Nationalbibliothek, with which other national libraries soon joined, together with some other non-library institutions. VIAF records different forms of name for the same entity, for example *person*, from different sources. VIAF records also the ISNI, the International Standard Name Identifier for personal *names*. The ISNI (ISO 27729) is:

- “the ISO certified global standard number for identifying the millions of contributors to creative works and those active in their distribution, including researchers, inventors, writers, artists, visual creators, performers, producers, publishers, aggregators, and more. It is part of a family of international standard identifiers that includes identifiers of works, recordings, products and right holders in all repertoires, e.g. DOI, ISAN, ISBN, ISRC, ISSN, ISTC, and ISWC.
- The mission of the ISNI International Authority (ISNI-IA) is to assign to the **public name(s)** of a researcher, inventor, writer, artist, performer, publisher, etc. a **persistent unique identifying number** in order to resolve the problem of name ambiguity in search and discovery; and **diffuse each assigned ISNI** across all repertoires in the global supply chain so that every published work can be **unambiguously attributed to its creator** wherever that work is described.
- By achieving these goals the ISNI will act as a **bridge identifier** across multiple domains and become a critical component in **Linked Data** and **Semantic Web** applications.”^{23,24}

22 VIAF: the Virtual International Authority File, <http://viaf.org/>.

23 International Standard Name Identifier (ISO 27729), <http://www.isni.org/>.

24 On ISNI see: Andrew MacEwan, Anila Angjeli and Janifer Gatenby, “The International Standard Name Identifier (ISNI): The Evolving Future of Name Authority Control,” *Cataloging & Classification Quarterly*, 51:1-3 (2013), 55-71, DOI: 10.1080/01639374.2012.730601.

How do VIAF and ISNI solve the issue of multiple identities? They both record the state-of-the-art, that is, VIAF displays all forms of a name with links between the same names as used by national libraries who provide the data, and ISNI applies as many identifiers as there are bibliographic identities, that is, public names. In the case of our example, ISNI 0000 0001 2148 628X is assigned for the name “J. K. Rowling”, in alphabetical order of forms of name, and another for related identities. One of them is “Scamander, Newt” with a bracketed explanation – “other identity, same person”. The ISNI for “Scamander, Newt” is ISNI 0000 0000 7049 5783, recording two forms of the name: “Scamander, Newt” and “Scamander, Newt (Pseud.)”, with information under Dates: “1965-“. As an ISNI identifier has an associated URI, the name, irrespective of recorded forms, represents one and same bibliographic, public identity that stands for a specific entity, while at the same time enables linking to other names representing other identities of the same entity. This is the answer of the new technology to the *See also* reference of the book, card and online catalogue!

The date of birth of “Newt Scamander”, recorded in the Spanish and Korean national library catalogues as “1897“, gives us the opportunity to discuss here another issue, the one of how to treat names of fictional characters when assuming authorship. IFLA LRM makes it clear that an author cannot be a non-human entity: “... the entity *person* is restricted to real persons who live or are assumed to have lived”. The LRM offers a solution for those cataloguing communities that traditionally assign authorship to fictional, literary or legendary figures. The entity *nomen* is the concept of “name” redefined from FRAD via FRSAD. It is linked to any thing, any entity, with the relationship “has appellation”. LRM uses a super-entity of all other entities, *Res*, to allow the modelling of high-level structures that can be inherited by sub-entities. The definitions of the two entities and their relationship is:

- LRM-E1 *Res*: “Any entity in the universe of discourse.”; the Scope notes say “... *Res* includes both material or physical things and concepts.”

- LRM-E9 Nomen: “An association between an entity and a designation that refers to it.”
- LRM-R13: RES has appellation NOMEN: “This relationship links an entity with a sign or combination of signs or symbols through which that entity is referred to within a given scheme or context.”

In our case, the cataloguer has to decide with what entity and role the name “Newt Scamander” is associated. The name is a *nomen* that is an appellation of a specific entity. On the one hand, the entity is the *person* J.K. Rowling who authored the book, while on the other it is the fictional character (a concept) associated with the book in some other way. The LRM treats each as a separate *nomen* of two different entities, even though the “name”, the LRM’s *nomen* string attribute, is the same.²⁵

In other words, the IFLA LRM clearly separates the “referent”, the thing being referred to or labelled (*person*), from the “reference”, the signs and symbols that label the thing (*nomen*). This is a standard approach in modern linguistics, semantics, and computing.

Discussion

The case study illustrates a benign example of the impact of the Semantic Web’s AAA principle. There is a real person known as “J.K. Rowling”, capable of making a statement that falsely and deliberately implies that her work was created by a non-human, fictitious entity with the name “Newt Scamander”. She uses the social construct of the established layout of a printed volume to make the statement, by placing the name immediately after the title. She also rectifies the statement using the same construct by using the preposition “by” with her pen-name for the Harry Potter works. This pen-name is based on another social construct. The use of initials conceals personal information such as sex, religion, and ethnicity. J.K. Rowling also uses the pen-name “Robert Galbraith”, using the same con-

25 See further the discussion of this issue in *IFLA LRM*, op. cit., 5.4 Nomens in a Library Context.

struct to falsely and deliberately imply that she is male.²⁶

The broader social construct is the “universe of discourse” confined, like the “bibliographic universe”, to persistent information carriers; that is, *manifestations* that may consist of one or more copies or *items*. In this universe or domain, a *manifestation* must embody an *expression* that realizes a *work* created by one or more *persons* or real human beings. A *manifestation* and its ur-creator are labelled with a title and a name; another social construct that differentiates a *person* from a product by categorizing the referent. Both appellations are usually assigned by the same *person* and used for representation in *manifestations* within the universe.

The IFLA LRM is compatible with the Open World Assumption: the *nomen* of a creator may not be known, but this does not imply that there is no creator. Conversely, the presence of a *nomen* in the context of a statement of responsibility for creation of the embodied *work* does not imply that the creator is anything other than a *person*. This may appear to conflict with the social constructs of specific groups of users. In this case, children at the lower age limit of the intended audience of the *work* may not agree that a non-human entity could not have created it; clearly, in the universe of discourse of Harry Potter, non-human entities are capable of creation and production. In other cases, religious or ethnic groups may include transcendent entities that are believed to participate in discourse via a human being as “persons”. The separation of an appellation from its entity allows a *nomen* to be processed in “name authority” applications without requiring the identification of the type of entity of which it is the appellation, following the Open World Assumption.

The presentation of library and other metadata to its users is also a social construct. This has its modern origins in the catalogue card and printed bibliography, at the same time as the layout of the printed volumes that the metadata describes, in the second half of the nineteenth century. The presentation of metadata is the focus of ISBD, which is based on a recapitulation of the layout of the self-describing data in a *manifestation*. The utility of this construct is severely diminished by the impact of digital

²⁶ Wikipedia, https://en.wikipedia.org/wiki/J._K._Rowling.

technologies: presentation of data and its metadata is localized to specific groups of creators, products, and audiences. The Semantic Web does not require a universal display format when atomic, single statements can be assembled into molecular sets optimized for any specific application. The concept of a “record”, in the sense of a fixed set of metadata recorded and displayed as a definitive representation of a *manifestation*, an “official record”, is at odds with the Open World Assumption.

More generally, there is no requirement to make any descriptive statement at all about an entity. Its URI is sufficient for it to be part of a chain of links between other entities of interest. A *manifestation* of a *work*, irrespective of its popularity with the audience, may not offer a single “title proper” or any statement of responsibility, or any of the conventional representation data that form the basis of a description.

Conclusion

The issues raised by the case study are resolved by separating labels from the categorization of entities of interest. A “personal” name is insufficient evidence that the entity labelled is a person. Conversely, every name in a statement of responsibility is the label of a person or group of persons, according to the LRM. Global interoperability of metadata recorded by different “official” agencies is better served by semantic coherence in the categorization of entities and relationships of interest than by relying on labels and manifestation statements. Semantic Web technologies support this approach, especially in the context of social construction and deconstruction. The context of the AAA principle and Open World Assumption implies an increasing important of the role of official, national cultural heritage organizations as providers of trustworthy provenance.

REFERENCES

- BERNERS-LEE, TIM. *Linked data - design issues*,” Date: 2006-07-27, last change: Date: 2009/06/18, <http://www.w3.org/DesignIssues/LinkedData.html>.
- DUNSIRE, GORDON. *RDA data capture and storage*. Presented to Committee on Cataloging: Description and Access II (CC:DA) - ALCTS CaMMS, ALA Midwinter 2016, 11 January 2016, Boston, Mass., <http://www.gordondunsire.com/pubs/pres/RDADDataCap.pptx> “Functional Requirements for Authority Data,” <http://www.ifla.org/publications/functional-requirements-for-authority-data>.
- Functional Requirements for Authority Data: A Conceptual Model*, Final Report December 2008, As amended and corrected through July 2013, http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf.
- Giant Global Graph*, https://en.wikipedia.org/wiki/Giant_Global_Graph.
- International Standard Name Identifier (ISO 27729)*, <http://www.isni.org/>.
- ISBD: International Standard Bibliographic Description*. Consolidated ed. Berlin/Boston: De Gruyter Saur, 2011; *ISBD Consolidated edition*, March 2011 (pre-print), http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf.
- LUBETZKY, SEYMOUR. *Principles of Cataloging: final report: Phase I: Descriptive cataloguing*. Los Angeles: University of California, Institute of Library Research, 1969.
- MacEwan, Andrew, Anila Angjeli and Janifer Gatenby. “The International Standard Name Identifier (ISNI): The Evolving Future of Name Authority Control,” *Cataloging & Classification Quarterly*, 51:1-3 (2013), 55-71, DOI: 10.1080/01639374.2012.730601.
- RIVA, PAT, PATRICK LE BOEUF and MAJA ŽUMER. *FRBR-Library Reference Model*, 2016-02-21, Draft for the World-Wide Review (International Federation of Library Associations and Institutions), https://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf; information on the world-wide review, <https://www.ifla.org/node/10280?og=587>.

- RIVA, PAT, PATRICK LE BOEUF and MAJA ŽUMER. *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*, December, 2017. Den Haag: IFLA, https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf.
- SVENONIUS, ELAINE. *The Intellectual Foundation of Information Organization*. Cambridge, Mass.; London: MIT Press, 2000.
- Verona, Eva. *Pravilnik i priručnik za izradbu abecednih kataloga: Dio 1: Odrednice i redalice*, 2. izmijenjeno izd. Zagreb: Hrvatsko bibliotekarsko društvo, 1986.
- VIAF: *the Virtual International Authority File*, <http://viaf.org/>.
- WILLER, MIRNA. "Name Authority Control Paradigm Shift in the Network Environment," in *Frameworks for ICT Policy: Government, Social and Legal Issues*, edited by Esharenana E. Adomi. Hershey, PA: IGI Global, 2010, 182-205, DOI: 10.4018/978-1-61692-012-8.ch012.
- W3C. *Resource Description Framework (RDF)*, <https://www.w3.org/RDF/>.
- Wikipedia, https://en.wikipedia.org/wiki/J._K._Rowling.

FURTHER READINGS

- BERNERS-LEE, TIM, MARK FISCHETTI. *Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor*. London: Orion Business, 1999.
- HENDLER, JIM and TIM BERNERS-LEE. "From the Semantic Web to social machines: A research challenge for AI on the World Wide Web," *Artificial Intelligence*, October 2009, doi:10.1016/j.artint.2009.11.010, <http://web.stanford.edu/class/cs227/Readings/hendler-berners-lee-semantic-web.pdf>.
- WILLER, MIRNA and GORDON DUNSIRE. *Bibliographic Information Organization in the Semantic Web*. Oxford: Chandos publishing, 2013.
- W3C. *Resource Description Framework (RDF)*, <https://www.w3.org/RDF/>.
- W3C. *Simple Knowledge Organization System (SKOS): Reference: W3C Recommendation*, 18 August 2009, <http://www.w3.org/TR/skos-reference/>.
- W3C. *Web Ontology Language (OWL)*, <https://www.w3.org/2001/sw/wiki/OWL>.

AUTORITET NASUPROT AUTENTIČNOSTI PRIJELAZ S OZNAKA NA IDENTIFIKATORE

KLJUČNE RIJEČI:

autoritet, autentičnost, tehnologije semantičkog weba, povezani otvoreni podaci (LOD), entitet osoba, entitet nomen, IFLA Library Reference Model (IFLA LRM)

SAŽETAK

U radu se prvo opisuju temeljni koncepti tehnologije semantičkog weba i otvorenih povezanih podataka (Linked Open Data: LOD) kao što su reifikacija, načelo AAA (Anyone can say Anything about Anything), provenijencija i Pretpostavka otvorenog svijeta (Open World Assumption: OWA) s ciljem istraživanja potrebe jasnog iskazivanja društvenih konstrukata i „društvenih mašina“ na kojima se temelje otvoreni povezani podaci u području djelovanja arhiva, knjižnica i muzeja (AKM) kako bi se potpomogao, a time i iskoristio prijelaz s ljudski-čitljivih oznaka na strojno-obrađive identifikatore. Glavno je pitanje, stoga, kako jasno iskazati društvene konstrukte, pravila koja upravljaju i podupiru proizvodnju podataka pogodnih za strojnu obradu i semantičko razumijevanje u aplikacijama koje su namijenjene čovjeku kao korisniku.

Pod pretpostavkom da će se ta tehnologija primijeniti za izgradnju budućih knjižničnih kataloga, a u kontekstu novoga IFLA-inog konceptualnog modela (IFLA LRM) za knjižnične metapodatke koji je razvijen „vodeći računa o tehnologijama semantičkog weba“, ovaj rad istražuje izbor i oblik autorova imena s višestrukim identitetima kao analizu slučaja.

U analizi slučaja, postavljena pitanja rješavaju se odjeljivanjem oznaka (nomena) od kategorizacije entiteta koji su predmet zanimanja (osobe). „Osobno ime“ nije dovoljan dokaz da je označeni entitet osoba. Suprotno tomu, prema modelu IFLA LRM svako ime u podacima o odgovornosti kataložnog opisa jest oznaka osobe ili skupine osoba. Globalna interoperacija

bilnost i ponovna upotreba metapodataka koje bilježe različite „službene“ agencije mogu se bolje provoditi semantičkim usklađivanjem kategorizacije entiteta i odnosa, nego li oslanjanjem na oznake i podatke o pojavnim oblicima. Tehnologije semantičkog weba podržavaju takav pristup, posebno u kontekstu društvenih konstrukcija i dekonstrukcija. Načelo AAA i Pretpostavka otvorenog svijeta ukazuju na sve veću važnost službenih nacionalnih baštinskih organizacija kao autoriteta koji daju vjerodostojnu provenijenciju metapodacima.