Inheritance and Lamination in the Representation of Bibliographic Relationships

Abstract:
This paper uses the concept of lamination implied in the term “discovery layer” to explore how domain knowledge could be applied to large federated search environments. Using the publishing history of Daniel Defoe’s *Robinson Crusoe* as a case study, we use bibliographic scholarship in the English literary studies community to establish lines of inheritance on all four levels of the FRBR paradigm: work, expression, manifestation and item. We created a small demonstration of a visualization generated from linked data extracted from the scholarly literature, to show how literary scholarship, when encoded as linked data, can create lines of inheritance and influence that enable users to fulfil the fifth user task of the new Library Reference Model: exploration.

1.0 Introduction
Knowledge organization, as a field of both theory and practice, regularly grapples with the place of specialized domain knowledge within information systems that purport to be “universal” in breadth and use. While the teams that maintain the Decimal Classification, the Universal Decimal Classification and the Library of Congress Classification rely on subject experts to help them maintain various sections of these tools, the relationship between domain knowledge on the one hand, and principles of knowledge organization that purport to be *a priori* assumptions on the other, is by no means clear (Hjorland 2015). In this paper, we address this conflict in a different way by focusing on bibliographic rather than thesaural relationships: in particular, we examine the patterns by which one bibliographic entity influences others over time. By considering the possibilities of using linked data to encode relationships across federated information resources, we explore new ways in which specialized domain knowledge can be integrated into library catalogues.

2.0 Federated Searching and the Concept of Lamination
The concept of federated searching—simultaneously searching multiple information sources—has a long history in librarianship, dating back at least to the National Union Catalogue of the Library of Congress. Efforts to establish global standards of information description—“Universal Bibliographic Control”—rest at least partly on the dream of searching diverse catalogues through the standardization of description methods. On the Web, metasearch engines and metadata harvesting initiatives have led to the call of Tim Berners-Lee for data that is freed from its local context and permitted to combine in productive and exciting ways (Berners-Lee 2009).

The concept of the “discovery layer” has signalled a new environment in which federated searching in libraries has gone to a new level of complexity and power. Discovery layers are search platforms that integrate online catalogue data with data from other sources, enabling libraries not only to provide access to a wide range of resources through a single interface or portal, but also to customize search results in specialized ways (Ramsay & Chamberlain 2012). Not only do these discovery layers permit...
searching across a range of different catalogues; they can also create interfaces that mimic popular search engine interfaces, permit faceted filtering of search results, and allow enhanced means of user feedback and participation in community search practices.

Much of the challenge with discovery layers lies, naturally enough, with problems of data harmonization: how to enable metadata structured and encoded for one resource to combine with metadata from a different resource entirely. The promise of linked data, combined with the emergence of the Dublin Core as a *lingua franca* to which multiple diverse systems can convert, has made such harmonization possible, producing interfaces such as Western Libraries’ new federated search interface, OMNI, which allows users to search 14 Ontario university libraries for books, articles, videos, music and databases using one common interface.

The term “discovery layer,” however, evokes another mental image as well. While proprietary platforms such as ProQuest’s Summon offer limited opportunities for customization, the rise of open source platforms such as VuFind and Blacklight offer greater opportunities for innovative adaptation (Barber, Holden & Mayo 2016, 182). Such discovery layers make it possible for users to add their own tags to records, creating folksonomies that are effectively “laminated” over the library catalogue, creating bibliographic relationships of specific utility. The concept of lamination—the superimposition of multiple data layers upon a foundational piece of data—was invoked by Ranganathan to describe the process of adding isolates to a base number to create a compound subject (Ranganathan 1967, 354). Later, the rise of geographic information systems popularized the process of overlaying geographical areas with successive layers of data indicating different facets of geographical interest (Hawkins 1994, 94).

This paper describes a case study which explores the feasibility of incorporating specialized domain knowledge into a linked data discovery layer that could be laminated over bibliographic data—particularly union catalogues—in order to encode and visualize bibliographic relationships in fresh ways.

3.0 Bibliographic Relationships

Much of the development in bibliographic description since the heyday of the *Anglo-American Cataloguing Rules* has involved the development of richer and more varied bibliographic relationships. The digital environment has enabled us to move beyond those relationships defined in the Paris Principles and which had to be primarily implied through main and added entries that determined where a record appeared (or did not appear) in a library’s card catalogue. The paradigm of the *Functional Requirements of Bibliographic Records* (FRBR), which serves as the basis for AACR2’s replacement, *Resource Description and Access* (RDA), separates a resource into four distinct entities, Work, Expression, Manifestation and Item, each of which exists in a one-to-many relationship with its predecessor. This paradigm is exploited in *RDA* to facilitate four primary objectives: finding resources, selecting them, identifying them and obtaining them (RDA 0.4.2.1) In addition, *RDA* provides a set of relationship designators for each level of the paradigm that can be used to encode relationships between different resources, such as abridgments, adaptations, digests, inspirations, remakes and variations (RDA Appendix J).
With IFLA’s recently-released *Library Reference Model* (LRM), library catalogue standards have admitted a new objective. In addition to existing objectives to enable the user to find, identify, select and obtain, the LRM has added “Explore” as a fifth important user task: “to discover resources using the relationships between them and thus place the resource in context” (Riva, Le Boeuf & Žumer 2017, 15, emphasis added).

The question lurking behind this new functionality is, of course: who is going to do all of this extra encoding? One option would be descriptive cataloguers themselves: the existing relationship designators would enable cataloguers to do a certain amount of it, embedding the relationships directly into the bibliographic records. In stark contrast, many discovery layers suggest another option by including user tagging affordances in their features, enabling library users to recognize and flag relationships of meaning to themselves.

This paper investigates a middle approach between these two extremes: the use of professional domain knowledge encoded to create bibliographic relationships between different resources: relationships that can exist at various levels of the FRBR paradigm. In this particular case, we are examining relationships defined in the field of literary studies, drawing on published literary and bibliographic scholarship to define relationships that could be encoded using linked data that refers to existing bibliographic and authority records, thereby enable users to follow temporal progressions of influence and development in literary history. Patterns of influence and inheritance would be effectively “laminated” over the bibliographic data when needed.

### 4.0 Temporal Progression

The use of temporal progression as a guiding principle in this case study requires some justification: not all literary studies scholarship is rigidly defined by chronology. Nonetheless, a significant amount of literary scholarship analyzes patterns of influence, both documented and implied, patterns of development of certain genres and themes, and verification of the accuracy of certain statements. Time therefore figures prominently in many literary studies, particularly studies of literary bibliography.

Classification theorists acknowledge the omnipresence and importance of time as a facet, and Fairthorne anticipated a growing awareness, within and beyond information science, that important relationships of inheritance can only be understood through a close attention to the temporal aspects of classification, whether natural or bibliographical (Fairthorne 1985, 363). More recently, scholars have acknowledged a shift in scientific thinking away from Linnaean taxonomies to a cladistic paradigm that recognizes lines of descent across time, thereby redefining category membership in ways that Linnaean classification does not allow (Hjorland 2015).

While this case study does not adopt a formally cladistic approach, we proceed from the assumption that charting scholarship in literary studies and literary bibliography along a temporal dimension provides a fruitful means of overlaying domain expertise upon a large information store for purposes of bibliographic exploration.

### 5.0 The Case of *Robinson Crusoe*

Daniel Defoe (1661?–1731) is an author who cries out for scholarly assistance in the act of bibliographic control. An extraordinarily prolific author, Defoe wrote voluminously on a dizzying range of subjects, at a time when authorial attribution was
highly unreliable. As a result, the number of works attributed to Defoe has grown over the years, only to be cut back ruthlessly in the 1990s when Furbank and Owens aggressively refuted many of the attributions (1998). As a result, conventional bibliographic control can be perilous in the case of Defoe: *The Life of Mrs. Christian Davis*, a sensational history of a cross-dressing Irish matriarch who distinguished herself as a foot soldier before succumbing to dropsy and scurvy, was mistakenly attributed to Defoe for some time, and the bibliographic metadata of at least one published version reflects this mistake.

*Robinson Crusoe* lends itself particularly to scholarly assistance. It was “a milestone in literary history” that was immediately embraced (and sometimes ridiculed) well beyond England (Backscheider 1989, 412). What is more, the novel exhibits complexities, both in its publishing and its influence, that touch on all four levels of the FRBR paradigm. Defoe himself wrote two sequels: *The Further Adventures of Robinson Crusoe* later in 1719, and *Serious Reflections During the Life and Surprising Adventures of Robinson Crusoe* in 1720. There were many translations of the novel into other languages, as well as multiple abridgments, some unauthorized. The novel’s premise of a man shipwrecked on an island and growing in wisdom and knowledge inspired a long tradition of adaptations known as “Robinsonades,” some for adults and many for children.

In exploring these complexities, we envisioned a federated search environment that would include the catalogues of numerous academic libraries, together with their extensive authority records, and such databases as the English Short Title Catalogue. Using authoritative scholarly and bibliographic guides to the publishing history of *Robinson Crusoe*, we began tracing inheritances at all four levels of the FRBR paradigm: Work, Expression, Manifestation and Item.

### 5.1 Work Relationships

*Robinson Crusoe* was first published in 1719. The Library of Congress Name Authority File contains a name-title authority record for the work:

- **Defoe, Daniel, 1661?-1731. Robinson Crusoe**
- **LC Authority**: [https://lccn.loc.gov/n81045585](https://lccn.loc.gov/n81045585)
- **Publication date**: 1719

Using the Name Authority File together with scholarly data, we can isolate at least two lines of influence at the Work level: sequels and Robinsonades.

#### 5.1.1 Sequels.

Hutchins (1925), Lovett (1991) and Furbank and Owens (1998) all agree that Defoe published two sequels to *Robinson Crusoe*, one in 1719 and one in 1720. The LC Authority File contains name-title access points for both:

- **Defoe, Daniel, 1661?-1731. Farther adventures of Robinson Crusoe**
- **LC Authority**: [https://lccn.loc.gov/n94091299](https://lccn.loc.gov/n94091299)
- **Publication date**: 1719

- **Defoe, Daniel, 1661?-1731. Serious reflections during the life and surprising adventures of Robinson Crusoe.**
- **LC Authority**: [https://lccn.loc.gov/n94091314](https://lccn.loc.gov/n94091314)
5.1.2 Robinsonades.

Searching the MLA International Bibliography using the term “Robinsonades” produces a number of scholarly articles that identify works drawing on *Robinson Crusoe*. Among those with LC Name-Title Authority Records are:

- Dalayrac, N. (Nicolas), 1753-1809. Azémia
  - LC Authority: https://lccn.loc.gov/no93006886
  - First Performed: 1786
  - A French comic opera

- Ducray-Duminil, M. (François Guillaume), 1761-1819. Lolotte et Fanfan.
  - English
  - LC Authority for English Translation: https://lccn.loc.gov/n00075009
  - First published: 1788
  - Novel, known in England as Ambrose and Eleanor

- Campe, Joachim Heinrich, 1746-1818. Robinson der Jüngere
  - LC Authority: https://lccn.loc.gov/n95060785
  - First published: 1788
  - Instruction book for children

  - LC Authority: https://lccn.loc.gov/no2015153207
  - First published: 1915
  - Novel

- Golding, William, 1911-1993. Lord of the flies
  - LC Authority: https://lccn.loc.gov/no2006021893
  - First published: 1954
  - Novel

5.2 Expression Relationships

5.2.1 Translations

*Robinson Crusoe* was translated into many languages. A recent exhibition at Indiana University’s Lilly Library (http://www.indiana.edu/~liblilly/defoe/translations.html) provides the following examples of early translations:

- First French Translation, 1720:
Italian Translation, 1731:
Translation based on the French version above.

Second Dutch Translation, 1735:
Het Leven en de wonderbare Lotgevallen van Robinson Crusoe, behelzende onder andere ongehoorde uitkomsten een verhaal van zijn acht-en-twintigjarig verblijf op een onbewoond eiland, gelegen op de kust van America, bij de mond van de rivier Oronoone. Alles door hemzelf beschreven. Door G. Schreuders. Amsterdam, 1735-1736.

Arabic Translation, 1835:
Qi•s•sah R¯ubin•sun Kr¯uz¯i.Malta, 1835.

Persian Translation, 1878:

5.2 Abridgments
At least three abridgments of Robinson Crusoe appear in the English Short Title Catalogue:

The Midwinter Abridgment, 1722:
The life and most surprizing adventures of Robinson Crusoe, of York, mariner. Who lived eight and twenty years in an uninhabited island on the [co]ast of America, lying near the mouth [of] the great river of Oroonoque: ... The whole three volumes faithfully abridged, ..
English Short Title Catalogue (ESTC) Number: 006343293

An Abridgment of the Midwinter Abridgment, 1734:
The wonderful life, and most surprizing adventures of Robinson Crusoe, of York, mariner ... Faithfully epitomized from the three volumes, and adorned with cutts suited to the most remarkable stories .. printed for A. Bettesworth and C. Hitch, at the Red-Lyon; and J. Osborn, at the Golden-Ball in Paternoster-Row; R. Ware in Amen-Corner, and J. Hodges at the Looking-Glass on London-bridge, 1734.
ESTC Number: 066477403
Abridgment, 1790:
A concise abstract of the wonderful life, and surprising adventures of that renowned hero, Robinson Crusoe, who lived twenty-eight years on an unhabited island, and was afterwards released by pirates. Adorned with cuts. London: printed for, and sold by all the stationary and toy shops in town and country, [1790?]
ESTC Number: 006061120

5.3 Manifestation Relationships
Manifestation relationships have a significant potential to assist scholars in eighteenth-century literary studies by encoding important textual decisions made by modern editors. Modern literary studies frequently establish “standard editions” of canonical authors: editions which scholars prefer to consult and cite, and which often prove the basis for subsequent editions.

In the case of Robinson Crusoe, the de facto “standard edition” is the 2009 edition of the novel which appears in the Chatto and Windus complete edition of Defoe Novels. This edition was created by consulting Defoe’s first edition together with Defoe’s manuscript (Owens 326). But other lines occur as well. The first edition of 1719 was reprinted in 1827 as the Shakespeare Head Press edition, which served as the base text for the Norton Critical Edition, published in 1975 (Shinagel 1994). The Broadview Press edition of Robinson Crusoe studied a handful of editions published in Defoe’s lifetime (Davis 2014, 37).

In many instances in literary history, significant differences exist between early editions, such as the difference between the first and fourth editions of Samuel Richardson’s Clarissa, or the 1799, 1805 and 1850 editions of Wordsworth’s The Prelude. In such cases scholars and instructors, when selecting a text for use or for teaching, would very much want to trace modern editions that are based on a particular earlier edition.

5.4 Item Relationships
In some cases, textual decisions are made, not just on the basis of a particular edition, but on a particular copy of an edition: one which contains perhaps an author’s handwritten marginal notes, or which contains a half-sheet imposition correcting a compositor’s error. In the case of Robinson Crusoe, an editor of a modern edition, skeptical of the vagaries of text reproduction in the eighteenth and nineteenth centuries, might prefer to overlook the Shakespeare Head reprint of 1827 and go directly to the copy of the first edition held in the British Museum.

6.0 Visualization
As a preliminary experiment in how these relationships might be visualized, we created a sample visualization. While this visualization was created manually, the later steps of actually rendering the image can potentially be automated. The visualization and its creation thus serve as a proof of concept for a future discovery layer.

To create this visualization we created a dedicated ‘catalogue wiki’ using the MediaWiki software along with the Semantic MediaWiki extension. Next, we gathered the access points for a sample of works referred to in scholarly articles as Robinsonades:
we used the name-title access points in the Library of Congress Name Authority File. This information was then added to the catalogue wiki, creating simplified authority records in a wiki environment.

We then added to the catalogue wiki a ‘SourceOf’ relationship from *Robinson Crusoe* to the Robinsonades group. The Semantic MediaWiki extension allows for the encoding of these sorts of semantic web linkages within the MediaWiki environment.

The following steps to create the visualization were done manually. We exported the authority records in a standardized RDF format to visualization software. In this case Gephi was used as it has a web RDF import extension that is easily enabled. Once the RDF information was imported into Gephi extraneous information was removed and labels were simplified for human readability. One of the data points preserved on all records was the date of the work’s creation, allowing the records to be arranged chronologically.

The final result is an image that is derived from the content of the authority records (see Fig. 1). With the addition of the authoritative relationships, the final image contains some elements reminiscent of a family tree, and other elements reminiscent of a cladogram.

![Figure 1: Robinsonade Descendents of Robinson Crusoe.](image)

### 7.0. Conclusion

Most of these bibliographic relationships can be encoded by cataloguers in existing MARC records using the relationship designators provided by RDA. However, the task would be overwhelming. A laminated set of linked data relationships established by knowledgeable domain experts, however, could conceivably be employed to draw certain paths through a large federated collection of bibliographic records and authority files. Such paths would facilitate the new objective of “Exploration” advocated by the Library Reference Model, and also support a variety of scholarly approaches to a specific knowledge domain.
References