Presenting Compounds to End Users in Search Results

Trond Aalberg, Pat Riva & Maja Žumer

Published online: 01 Jul 2024.

Submit your article to this journal ⬚

View related articles ⬚

View Crossmark data ⬚

Presenting Compounds to End Users in Search Results

Trond Aalberg[a,b] 🔲, Pat Riva[c] 🔲 and Maja Žumer[d] 🔲

aDepartment of Archivistics, Library and Information Science, Oslo Metropolitan University, Oslo, Norway; bDepartment of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway; cConcordia University Library, Concordia University, Montreal, QC, Canada; dDepartment of Library and Information Science and Book Studies, University of Ljubljana, Ljubljana, Slovenia

ABSTRACT

This article presents a prototype and a user study designed in order to understand how users perceive bibliographic compounds and what is a helpful and clear way of presenting search results in this case. The aim of this paper is to contribute knowledge of what features are significant in the display of these result sets, grounded in user feedback.

Introduction

The introduction of bibliographic data models that identify intellectual and artistic creations in the form of entities is an important prerequisite for content-centric management of the resources that libraries collect. Rather than simply describing each publication as one resource, models such as the IFLA Library Reference Model (IFLA LRM) enable the description of a publication using the constructs of a *manifestation* embodying one or more *expressions* each of which realize a *work*.[1] Although it is very common that a manifestation only embodies a single expression, such as the case of a book containing a single novel, it is also very frequent that a manifestation embodies multiple expressions, such as a collection of short stories or essays, a book that contains multiple novels by the same or different authors, a CD that includes different symphonies performed by the same or different orchestras. In addition to composite structure that occurs when a manifestation embodies more than one expression, IFLA LRM also defines part-of relationships between entities of the same type. A work may have other works as parts, such as a poetry collection with poems, a trilogy of novels or a coherent series of stories. Part-of relationships can also be defined between expressions as well as manifestations. The picture is further complicated by the concept of aggregating

work and aggregating expression, which in IFLA LRM is described as the concept or plan for the selection, assembly, and ordering of the expressions embodied in an aggregate manifestation.

In this study, we use the term "compound" to refer to the various compositions of entities that include works with parts and manifestations with multiple expressions and anything in between. While the theoretical modeling of these cases with IFLA LRM is relatively straightforward, more knowledge and experience is needed to guide the description of compounds and determine how they can be supported and utilized in search and result presentation. In the end, data and systems are intended to support users in their search and discovery of resources, and we also need to explore how end users interpret and utilize the results they are presented with. To address these challenges, we have adopted a design science research method with an exploratory user study. Bibliographic test collections with rich and consistent semantic data for a variety of compound entities were developed and used in the experiments. A system for searching and representing LRM-based data with support for user feedback was developed, with special attention to how compounds and parts should be utilized and presented. Finally, an exploratory user study was conducted where users inspected search results and gave feedback according to a set of predefined search tasks.

Background

IFLA Library Reference Model

The IFLA LRM expresses the commonalities and underlying structure of bibliographic resources and is intended to serve as a guide or basis on which to formulate cataloging rules and implement bibliographic data and systems. The model is developed using the scope of end-user needs (tasks) and covers all types of resources generally of interest to libraries. It is a consolidation and harmonization of a family of models that was initiated by the IFLA Functional Requirements for Bibliographic Records published in 1997 which again built upon a history of reflections and discussions on bibliographic concepts.[2] The so-called WEMI entities (work, expression, manifestation, item), and how they relate to each other, defines the core structure of the IFLA LRM. These entities represent four levels of abstraction of the intellectual products that make up the bibliographic universe. Work is defined as "The intellectual or artistic content of a distinct creation" and represents the highest level of abstraction at which we identify and conceptualize intellectual content such as a literary work or musical composition. An expression realizes a work by conveying the content with a distinct combination of signs such as a particular text of a literary work (e.g., translation) or a particular recorded performance of a musical

composition. Expressions are embodied in manifestations which are 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." Publication products in the form of printed books or music CDs are the most obvious examples of manifestations. Finally, manifestations are exemplified by items which are "an object or objects carrying signs intended to convey intellectual or artistic content" and typically will be the exemplars of printed books or music CDs found in the holdings of libraries. The WEMI entities and core relationships are depicted in Figure 1.

The WEMI entities have been confirmed with user studies, although most studies were concerned with the hierarchical grouping enabled by the core structure of works being realized in one or more expressions which then are embodied in one or more manifestations.[3] User studies by Pisanski and Žumer show that the conceptualization of users aligns reasonably well with the WEMI entities when users are presented with the task of organizing publications into groups.[4] The study performed by Tallerås et al. on derivatives also indicates that users' conceptualizations resemble the LRM Work entity, although alternative patterns can also be identified as significant for grouping derivatives, such as superwork and fictional world.[5] A study by Arastoopoor confirms that users have a hierarchical, top-down view when they have no particular document in mind
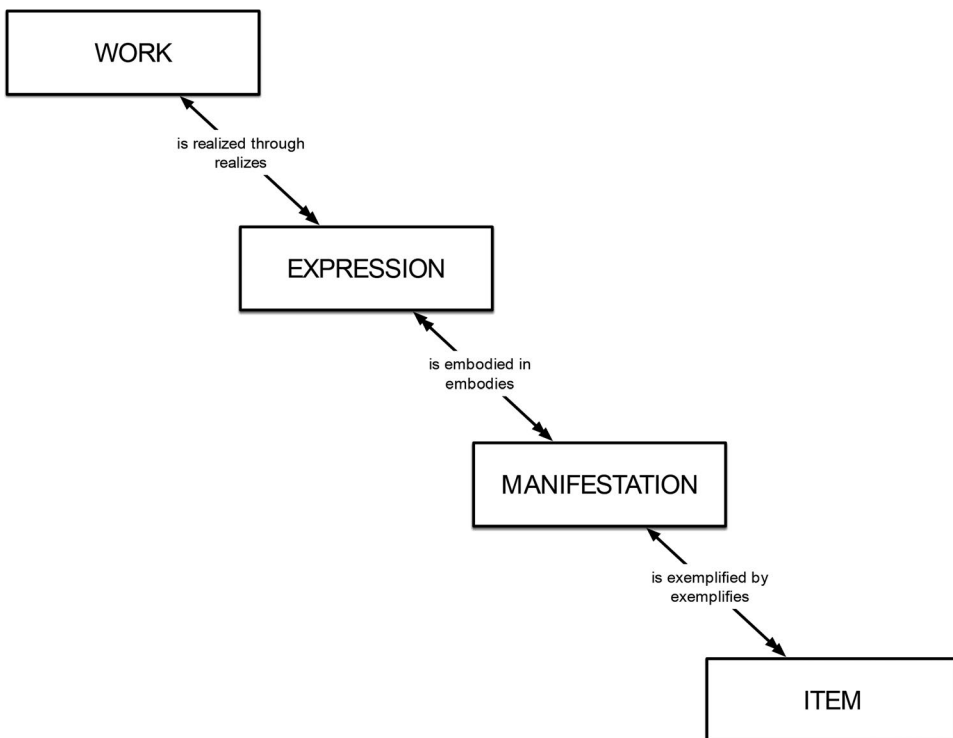
**Figure 1.** WEMI entities and core relationships.

but assume a bottom-up view when the need is for a specific expression and format.[6]

Aggregates and works with parts

The many-to-many relationship between a manifestation and an expression (LRM-R3 embodies) is an aspect that makes IFLA LRM unique and sufficiently powerful to model a broad range of carrier types and their content. A manifestation embodying multiple expressions is in IFLA LRM referred to as an aggregate and the modeling of aggregates follows the earlier recommendations from the IFLA Working Group on Aggregates.[7] There are three types of aggregates:

- Aggregate collections of expressions, which are sets of multiple independently created expressions, 'published' together in a single manifestation. The individual works are usually similar in type and/or genre, such as a collection of novels by a particular author, songs by a particular artist, or an anthology of a genre of poetry. In other cases, an aggregate may be a collection of expressions assembled for rather ad hoc reasons.
- Aggregates resulting from augmentations, which occur when an expression is supplemented with additional material that is not integral to the original work and does not significantly change the original expression. Forewords, introductions, illustrations, notes, etc. are examples of augmenting works.
- Aggregates of parallel expressions, which are manifestations embodying multiple, parallel expressions of the same work such as manuals and official documents for multilingual environments, or a DVD containing a motion picture with a choice of spoken languages and subtitle languages.

IFLA LRM argues that the process of aggregating expressions may itself be an intellectual or artistic effort, which could be represented as a work entity and for this purpose introduces the concept of aggregating work and a corresponding aggregating expression. These are not distinct entity types but merely an explanation of how a work instance can be introduced to represent the selection and arrangement of content (shown in Figure 2). A particular relationship *aggregated* (LRM-R25) is defined to support describing that a specific expression of a work was chosen as part of the plan of an aggregating expression, but this is merely intended for making explicit the connections that otherwise are implied with the embodies relationship. An aggregate should not be confused with works that were created with parts, such as multipart novels.
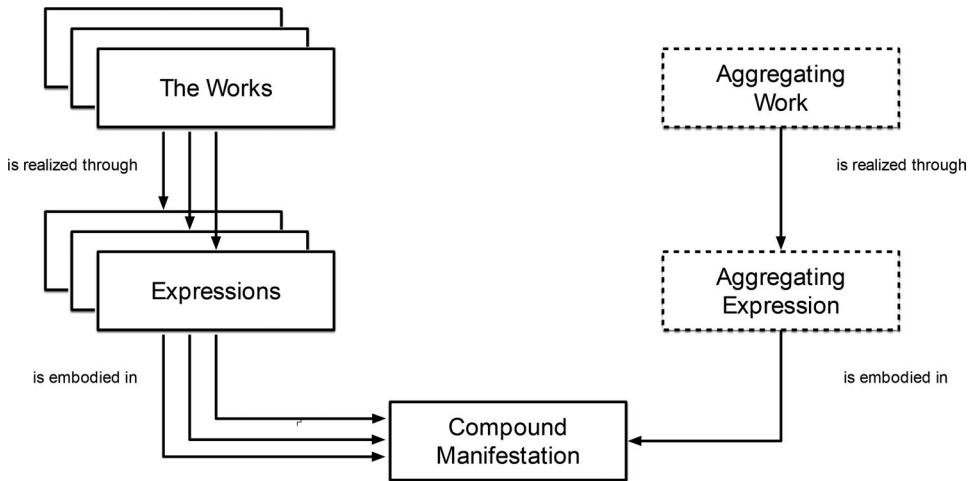
Figure 2. Modeling aggregates in LRM.

The aggregating work does not contain the aggregated works themselves and the concept of whole-part relationship is not applicable. The concept of an aggregating work can also be compared with other constructs introduced for managing composite resources at an abstract level, such as continuing resources in the form of serials or integrating resources.[8]

Aggregates occur rather frequently in the resources described in library records, according to a study by O'Neill et al.[9] In current RDA-based library practice, aggregates are emphasized as manifestations embodying an aggregating expression and practice tends to prioritize describing the aggregating work rather than the parts.[10] The implementation of the constructs needed to fully express all aspects of aggregations has been discussed and explored in some research but there is a lack of studies using realistic collections and systems.[11]

Most works are perceived as relatively simple, distinct, and individual units. However, some works are more complex, exhibiting internal structure or composition of some kind, or they are made into parts of compositions and structures they originally were not created for. The *part of* relationship is generally acknowledged as an important construct in discussions of bibliographic relationships[12] and IFLA LRM defines the part-of relationship (LRM-R18) between works where the content of one is a component of the other. Part-of relationships can also be defined between expressions using the LRM-R23 *has part* relationship, and often the part-of structure at the expression level mirrors the part-of structure at the work level. The work *The Lord of the Rings* by J.R.R. Tolkien has three parts at the work level, which also are presented as three parts at the expression level. When recording a performance of a symphony, the movements defined at the work level result in separate recordings, which form part of the overall recording of the performance. However, the structure at the expression

level does not always follow the exact same part-of structure as at the work level. A new edition of a poetry collection may include additions and removals of poems when compared with the first published version without being considered a new work. Popular music albums are often published as extended versions including extra tracks, also without being considered as a new work.

Although the formal definition and theoretical discussions clearly distinguish between aggregates and part-of structures at the work and expression level, the practice on how they should be applied is unclear and the practical use of the aggregating work and expression is not properly explored. Taniguchi points out that current cataloging rules (*RDA: Resource Description and Access*) lack proper identification of the type in the description of works/expressions but give instructions for access points supporting aggregate and component works/expressions, which may cause a variety of patterns to be applied in the cataloging of aggregates.[13]

Traditional practice for describing compounds

Cataloging practice has a tradition of various ways of describing and identifying compounds, although there is a great variety in the level of detail and choice of coding practice. When parts of a work are published separately, traditional catalogs nevertheless include the information about the larger works they belong to. This is partly also due to the information being visible from the title page, such as 'part 2 of …' being included as the subtitle or even as part of the main title. On the other hand, when the whole work is published, information about the parts is often missing entirely from the bibliographic data or is recorded only as contents notes. Contents notes are not necessarily consistently structured, with practices varying over time, thus reducing the potential for automatically establishing whole-part relationships. Some content is recorded using added entries and analytical entries; the use of this is generally sparse and varies between library catalogs. Consequently, users are often not referred to publications of the whole work when searching for one of its parts, even if it is the only option for obtaining the resource.

Traditional library practice is still primarily focused on the container (the publication in hand) and not on the content. The assumption is that a manifestation embodies only one expression, and it therefore represents a single work. This practice is quite visible in the treatment of collections. They are typically cataloged as simple works, with the work title taken from the manifestation (often generic titles such as 'Collected works') and the compiler recorded as the creator. The information about the works included is, at best, listed in contents notes or even superficially described in general notes. This information is often incomplete, particularly when

many works are in the collection. When works by a single author are collected, this author is typically recorded as creator of the collection, whether they were directly involved in assembling the collection or not, as opposed to the practice for collections of works by different authors. This practice results, on the one hand, in many authors being listed as the creators of their own posthumously compiled collected or selected works, and missing information about which works are included in collections on the other.

Treating a collection as a single work created by the compiler has another unwanted consequence: the works collected are subsumed into the collection and come to be treated as its parts. Since the whole-part relationship is of a permanent, inherent nature, being a part of a collection contradicts the existence of other independent publications of the same work that do not relate to the collection.

Other models

The modeling of entities having parts is a common element in most data models that are intended for, or include support for, bibliographic data. The BIBFRAME model that the Library of Congress initiated, defines the entity types of Work (bf:Work), Instance (bf:Instance), and Item (bf:Item), where the Work is defined as "reflecting a conceptual essence of a cataloging resource" and the Instance is "reflecting an individual, material embodiment of a Work."[14] The property "has part/part of" can be used on both the Work and Instance levels and allows for expressing that, for example, one instance of bf:Work has other instances of bf:Work as parts. BIBFRAME, however, does not typically envisage an instance being an instance of multiple works, which would be the equivalent to the embodies relationship (LRM-R3) in IFLA LRM. Cardinality constraints – which define how many instances of one entity can be related to instances of another – are not specified in the BIBFRAME RDF definition and consequently there is no formal rule preventing an instance being an instance of multiple works. The naming of the relationship and lack of examples indicating otherwise, however, suggest that BIBFRAME consolidates the traditional practice of interpreting the publication as a whole to be an instance of bf:Work, with composition only described for entities of the same type.

Schema.org has its main application outside the library domain but has the mission "to create, maintain, and promote schemas for structured data on the Internet."[15] It is developed by people from the founding companies: Google, Microsoft, Yahoo, and Yandex, as well as with the participation of the larger web community. The type CreativeWork is the one that most closely compares with LRM Work, but it is defined as the "most generic

kind of creative work, including books, movies, photographs, software programs, etc." This definition conflates the content and the format it appears in. In Schema.org, the types that can best be compared with the LRM manifestation entity are defined as subclasses of CreativeWork, implying that any book is also a CreativeWork, although it also supports stating that a Book is related to a work using the exampleOfWork property. The isPartOf/hasPart property can be used to describe the composition relationship between entities of the same type, and documentation indicates that this property can also be used between Book and CreativeWork to model content.

Wikidata is a manually curated knowledge base that covers all kinds of intellectual and artistic content and is used as a source for the display of facts in Wikipedia.[16] It is based on a rich and rather complex vocabulary and many of the classes are linked to external types such as IFLA LRM entity types. In Wikidata, we find the type Work (Q386724), which is defined as a broader class than LRM Work, and some of the properties indicate that it also compares with LRM Expression. Wikidata has the same hierarchical subclass perspective as Schema.org and defines the classes that are more LRM manifestation-like, such as Document, Book, and Film, to be subtypes of Work. A *part-of* relationship can be expressed between classes of the same type, but there is no designated property to make statements about what content an instance of class Book contains.

A general observation when comparing alternative models and type vocabularies with IFLA LRM is that LRM is unique in distinguishing between the work, as the most abstract conceptualization of content, and the expression, as a distinct unit conveying intellectual or artistic content. This distinction has been absent from BIBFRAME, although it is recently introduced in part with the Hub entity[17] (as well as the Opus entity defined in the Share-VDE project[18]). Support for expressing that an entity has parts is well captured in all models, but none of the alternatives support a property equivalent to the LRM-R3 embodies relationship in LRM.

IFLA LRM search and presentation

Search, presentation, and interaction with LRM-based bibliographic data have been the focus of some research and focused development, but so far there is no system available that demonstrates the full aspects of implementing the model. Grouping an author's publications by work is the most explored feature and predates the formalization of the WEMI-model.[19] An FRBR-based prototype was, already in 2001, described and supported with the Library of Congress FRBR Display Tool for presenting authors, works, and expressions in a hierarchical list.[20] A more well-developed system was OCLC FictionFinder, which used the OCLC

work-set algorithm to group WorldCat records for fiction into work-like clusters.[21] Some library system vendors were early implementors of the model and various protypes and evaluations were published,[22] but the interest in supporting the model waned throughout the first decade after the FRBR report was published in 1998. One reason could be that automatic grouping based on existing library data has many challenges caused by the lack of original titles for translations, the ambiguous use of added entries, etc.[23] Currently, there seems to be an increase in applying the semantic web and RDF as a platform for library data, which has also spurred new enthusiasm for bibliographic models.

Studies on the effectiveness of WEMI-based displays such as the visualized user interface explored by Merčun et al. and Salaba et al. show improvements over a baseline traditional listing, particularly in the case of complex works and exploratory tasks.[24] A study of a more realistic system for search and display of WEMI-entities was performed by Aalberg et al. and included different views for work, expression and manifestation and studies of different scenarios. Findings from the user studies indicate that each type of display is useful for some scenarios, but not for all of them.[25]

Implementations of WEMI in search and presentation have mainly explored hierarchical presentation along the axis of the core relationships between works, expressions, and manifestations. What entities to display in the results listing, and how to present them, is a main design choice that all LRM-based implementations face. Displays can favor one type of entity with a hierarchical listing of associated "child" entities. A work-based display will naturally result in the most compact listing, but it needs a more elaborate display to present the expressions and the manifestations available for each work. An expression-based display will typically result in a lengthier listing, but it requires less complex presentation of the associated manifestations. A different approach is the implementation of results listings that include a mix of entity types, such as presenting the expression whenever it is embodied in multiple manifestations, but otherwise presenting the manifestation. Of importance is also the challenge of how to display and enable users to interact with various relationships.

Types of compounds

In this study, we use the term "compounds" to refer to the various compositions of entities, including works with parts and manifestations with multiple expressions - and anything in between. The *part of* relationship between works reflects a structure at an abstract level, which is an inherent characteristic whether the work is realized as a whole or as a component.

An aggregate collection, on the other hand, is defined as a set of individually created expressions that do not imply any dependency between the works, although neither does it exclude this. In the following, we do not address and discuss aggregates resulting from augmentations or aggregates of parallels, as this is outside the scope of this investigation.

To better explain the nature of compounds, we adopt a categorization that differentiates between types of compounds according to how well established the identity of the compound is. We assume that parts can always be recognized as component works and differentiate between:

- Work with parts
- Established collections
- Ad hoc collections

Works with parts

This includes works with components which are works in their own right. Both the whole and each of the parts have a clear identity and, in most cases, the same creator(s). The publication history typically includes individual publications of the parts as well as publications of the whole. The titles of all works are generally known, and both the whole and the parts are commonly referred to in bibliographies and in common discourse. The part works are usually exclusively bound to the whole and will never be reused as parts of other works.

We can use Tolkien's *The Lord of the Rings* as an example. In addition to the whole trilogy, there are three parts: *The Fellowship of the Ring*, *The Two Towers,* and *The Return of the King*. All four are works, all created by J.R.R. Tolkien. Each of the parts has the 'is part of' (LRM-R18) relationship with *The Lord of the Rings*. The identity of the whole and of its parts is well known and shared globally. To establish the right sequence of parts, the relationships 'precedes' and 'succeeds' (LRM-R19) can be used.

The trilogy has been translated into many languages and published and republished extensively. Among the publications one may find boxed sets including all three novels, e-books containing all three parts as well as the parts individually. Quite often, additional materials can be found in the publications, such as illustrations, commentaries, introductions, or reader's companions. These are considered works in their own right and do not become part of the trilogy simply by being issued with it. (They are augmentations appearing in aggregates resulting from augmentation of a primary work.) Quite often, the novel *The Hobbit* is published with *The Lord of the Rings*. Written by the same author, it is a prequel and not part of the trilogy. The relationship 'precedes' (LRM-R19) can be used

to connect it to the trilogy, or, to be more precise, with part one, *The Fellowship of the Ring*.

Established collections

This includes compounds with clear identity, but where the whole is usually established as an individual entity at a later time than the parts. The component works were independently created and were combined into a collection, which was named at a later time, possibly by the original creator or, more often, by somebody else, a compiler, such as a publisher, literary agent, or anthologist. The compiler is known and typically recorded in bibliographic information systems, and so is the title of the collection. The components keep their individual identity despite being issued in the collection.

We can use *Masterpieces: The Best Science Fiction of the Twentieth Century* compiled by Orson Scott Card as one example. It features the short stories of many famous authors "who have forged a permanent place for science fiction in the popular culture of today… and tomorrow." Each short story was previously published independently and is included in various collections; therefore, they are all independent works. We must also acknowledge the contribution of the compiler: Card selected the works and arranged them in a meaningful order. He is, therefore, the creator of the aggregating work, and an expression of that work is also embodied in the manifestation, together with the expressions of the short stories. The compiler's name is clearly indicated on the cover.

Ad hoc collections

These are compounds with no clear identity as a whole, such as musical CDs or several short texts published together for the purpose of a single publication, including more than one independent work. The compilation is usually done by the publisher and the compiler is typically not recorded or may not even be known. The collection may have no title or have only a generic title such as "Two concertos" or "Three plays." In some cases, there is a distinct title (for example, "Poirot in the Orient," "Murders to die for"), but it is not well known, and these collections do not have an established identity. The intellectual novelty or effort behind many of these collections is marginal and can easily be duplicated. An ad hoc collection may become established if it becomes known broadly through republishing or critical acclaim; often, this also implies that it has or is given a more meaningful title by which it can be referred.

Mark Twain: 10 books is one example of an ad hoc collection. The compiler is not known, the publisher simply combined the novels and published

**Figure 3.** A CD containing two violin concertos with no collective title.

them together. Classical music CDs are also often examples of ad hoc collections with no title, only listing the works recorded (as shown in Figure 3).

Identifying compounds

While, in theory, works with parts and collections are two distinct groups, the borders are fuzzy in practice, due to the lack of documentation of the creation and publication processes (Figure 4). On the one hand, works with parts and ad hoc collections are clearly disjoint, on the other, the distinction between works with parts and established collections is not always so obvious and an ad hoc collection may evolve into an established collection if it becomes commonly recognized.

When works with parts were clearly created as such, there is no doubt. But there are cases when, for instance, three independently created novels developing the same story have been published later as a trilogy and given a title by the publisher (for example Stieg Larsson's "*Millenium Trilogy*," which contains the first three novels of what was supposed to be a planned series of ten). When the creation and publication process is not completely known, we mostly rely on a commonly shared understanding of the nature of such works.

The distinction between established and ad hoc collections is not that clear either. What seems to be an ad hoc collection at first, particularly when it has a distinctive title, may later be republished or will gain recognition and be referred to in literary theory, for example, thus acquiring its own identity.
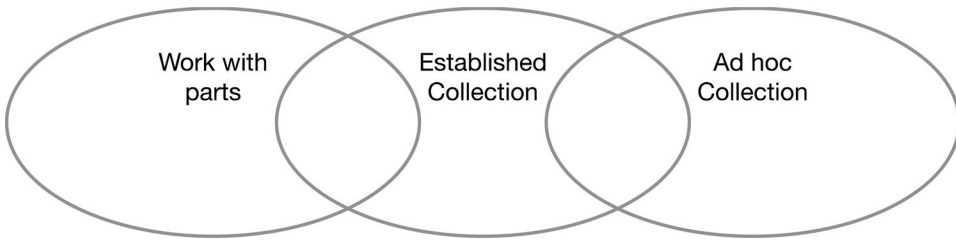
Figure 4. Types of compounds.

Since the modeling and, consequently, the presentation of information to the user, depends on the type of compound, we can use the following indicators to distinguish between these types.

Works with parts:

- Known creator(s), the same for all of the parts and the whole
- Established titles
- Titles of parts indicating their relationships to the whole
- Being part of the whole work is an inherent property of the part (it always holds for the part, whether it is published separately or as part of the whole)

Established collections:

- Known compiler, indicated prominently on the publication and recorded in bibliographic data
- Distinctive title
- Component works clearly indicated (title, creator)
- Component works may be published in other collections as well as separately

Ad hoc collections:

- Compiler not known or not clearly credited and not recorded in bibliographic data
- Generic title or no title
- Grouping by convenience (e.g., extent of parts)
- Rarely republished

Methodology

The main objective of this research is to contribute to the theoretical understanding of compounds as well as to increase our understanding of how they can or should be identified, described, supported, and

utilized in search and result presentation. We have adopted a design science research method with an exploratory user study. According to Johannesson and Perjons, design science is "the scientific study and creation of artefacts as they are developed and used by people with the goal of solving practical problems of general interest."[26] The artefacts developed in this study include bibliographic test collections with rich and consistent LRM-based data showcasing a variety of compounds, and a system for searching and presenting this data. An exploratory user study is conducted where users inspect and give feedback according to a set of predefined search tasks.

The main research questions are:

1. How can compound entities be expressed in data and utilized in search and result presentations?
2. What behavior or interpretations can be identified from inspecting users interacting with the results?

The intended research outcome is explanatory knowledge, which is characterized by a focus on 'how' and 'why'. The test collections developed can be used to answer how compounds and parts should be described in terms of entities, relationships, and attributes. The system developed can contribute with techniques and methods for presenting compounds and parts. Data from the user study will indicate selections and priorities users make in the context of the data and system. Knowledge about why users behave in certain ways given identifiable scenarios and conditions is valuable input in the design of systems.

Test collections

Data used in the experiments are bibliographic records retrieved from various catalogs using Z39.50 and the MarcEdit tool. Most English language records are from The British Library Integrated Catalogue and Library of Congress. Records in Norwegian are from the database of the BIBSYS consortium. Records were identified by performing searches on author and title and manually selecting a set that potentially would be relevant. The test collections vary in size and scope and have been tailored to the purpose of these experiments. Test collections are identified by the author's last name. Retrieved records were elaborated and enhanced to facilitate the processing and presentation of data. This typically included enhancing records with URI identifiers for works and persons using VIAF, Wikidata, ISNI, the Internet Movie Database (IMDb), and the Internet Speculative Fiction Database (ISFDB), as well as other authoritative sources for identifiers. Works and persons for which we

did not find a relevant identifier were given locally assigned URIs. Relationship types, such as those between a person and a work, expression, or manifestation, are coded using RDA registry types. To solve problems related to the ambiguous nature of added entries in MARC records, we used the technique of linking fields. The enhanced records were transformed into RDF using a rule-based conversion system, employing classes and properties from the RDA registry. Works and agents are identified using the URIs coded into the source records. Expressions are identified by automatically generated URIs constructed from the work identifier combined with additional data, such as language, content type, and numbers identifying significant agents related to the expression. The resulting test collections fully implement the IFLA LRM model, except that we do not include item entities in our data. Although essential to fulfill the user task "obtain" in an implemented system, item-level data is not relevant for the purpose of this user study. Pragmatic considerations relating to item format or availability data might influence end users' selections in response to the tasks, making it harder to discern their understanding of the compounds presented in the search interface.

The test collections include various examples of publications illustrating different aspects of compounds: works with parts, established collections, and ad hoc collections. The size of each test collection in terms of entities is presented in Table 1. For some test collections, the number of works and expressions are similar due to a lack of multiple expressions realizing the same work (same content type, same language). The number of source MARC records is equal to the number of manifestations.

For the Tolkien collection, the main example is the fantasy trilogy *The Lord of the Rings*, where both the work as a whole and each part are well known and published as individual volumes. In addition, we included boxed editions that include additional works and one combining the parts of the *Lord of the Rings* in three volumes with a readers' companion as fourth volume. Although the latter may be interpreted as an augmentation to the first, it has also been published individually. The other boxed set includes the parts of the *Lord of the Rings* boxed together with *The Hobbit* as the fourth volume.

Table 1. Size of the collections.

Collection	#works	#expressions	#manifestations	#agents
Tolkien	9	20	25	18
Christie	64	65	36	28
Mark Twain	51	51	27	29
Ballard	203	203	32	30
Hamsun	112	337	1032	194
McCarthy	62	76	135	32

For the Agatha Christie example, the emphasis is on her individual crime novels, which are often published as separate volumes but frequently also in ad hoc collections based on topical similarity, time period, main character, etc.

In the Mark Twain example, we focused on publications of Tom Sawyer stories (including Huckleberry Finn). This test collection is characterized by fewer individual novels and different collections that are variations of the ad hoc category. The Mark Twain collection includes novels published individually, publications including several of his novels, and collections of novels by different authors, as well as short story collections.

J.G. Ballard was selected to include examples of short stories that may appear in multiple collections. Some collections were originally compiled and published by the author under a specific title, other collections were constructed in retrospect, such as *Best of* collections or *Complete short stories*.

The Hamsun example is selected to explore the issue of collected works. These are Norwegian records that describe the separate volumes in different editions of the complete works, some selected works, and numerous publications of individual works. Additionally, there are collections put together by some specific criteria, such as early works or later works.

The McCarthy example includes a trilogy, which is published as a single volume as well as individually. Additional trilogies can be found in the Hamsun test collection. Although these are referred to as trilogies and mentioned in bibliographic listings of works by the author, they are not often published in single volumes.

Search system

To facilitate experiments in which users inspect search results and give feedback online, we developed an interactive prototype search system for searching and retrieval of LRM-based data, adapted to the research purpose of this project. The main idea of the prototype is to show search results to the user and allow users to mark items as relevant or non-relevant according to some given tasks. The overall design is expression-centric, which means that a search will retrieve and present a list of expressions that also includes work-level data and with an expandable list of publications (manifestations) in which this expression is available. If there is only one publication, we show a presentation that integrates information about all levels. Collections are represented using "pseudo-expressions" that represent the publication as an expression, allowing them to be presented alongside other expressions without having to add special presentation logic for this purpose.

At the backend, we store our RDF data in a Neo4J graph database. The actual data stored is a sub-selection and slightly simplified version of the RDF data generated when transforming the library records, such as using simple names for properties rather than URIs. The main reason for this setup is the availability of the Neo4J open source GraphQL Library that enables rapid and flexible API development with little code. Neo4J also comes with support for full-text indexing of nodes, which is applied to expressions in our prototype. For each expression, the full-text index includes all literal properties of an expression together with titles, names, and other relevant information from any associated entities to optimize the findability of the expression.

The frontend developed is a simplified user interface with a search bar above the presentation of the results, implemented using the React UI library and Google's Material Design. Results are presented on one page without any type of dynamic loading to avoid unnecessary complexity related to user marking of the results, and because our results never exceed what is normal to render on a single page. The user interface does not include any support for filtering or sorting, to keep the user interface simple and tailored to the experiment. However, additional search criteria can be added as parameters to the search URI to produce a more focused result set for search tasks (such as specifying the language of results or the type of content).

Collections and their parts are presented in two ways in the search system:

All collections are listed as separate results regardless of whether they are a "proper expression" or not. The main motivation for this choice is to enable user studies on the relevance of presenting collections as separate items in a result listing. Since a pseudo-expression is created for all collections, multiple publications of the same collection can be listed in the expandable list. Most collections with more than one publication would, however, be either a work with parts or an established collection rather than an ad hoc collection.

All parts of a collection also appear as separate entries in the result listing, along with a description of the publications each part is embodied in. If the part is included in multiple publications, we use an expandable list. This is comparable to how an individual expression that is embodied in multiple manifestations is presented (see Figure 5).

For parts of works, we present the *part-of* relationship with a label for the target as a clickable link (shown in Figure 6) where the novel "The Crossing" is "Part of" the collection "The Border Trilogy."

Collections presented as a result item include a contents listing, prefixed with "Includes," if this information is available in the record. We deliberately avoided the use of "Contents" since this may be understood as the content of a book. For entries representing a part of a work, we present

Figure 5. Individual novel with expanded list of publications, and ad hoc collection.

Figure 6. Work with parts.

the title of what it is a part of as a clickable link prefixed with "Part of." This is implemented for parts born as works and for some established collections, but we avoid this information for parts of ad hoc collections, since these works will potentially be included in many ad hoc collections.

Each result item is presented with an icon for the content type, or the content type that is representative of most of its parts. The language code is added to the icon for language-based material. Expressions are presented with the title that is most appropriate for that expression (the one in the language of the text) often derived from the 245$a field (title proper), but in some cases, we have hard-coded expression level titles in the data using 740-fields (title added entry) linked to the appropriate entry in the record. Under the title, we present agents associated with the work, as

well as agents associated with the expression. For compounds, we also describe the contents of this expression. We additionally present any expression-to-expression or work-to-work relationships available for the expression. Finally, under each expression we list each publication that this expression is embodied in, including its title and statement of responsibility as well as typical information such as extent and publisher data. Additionally, for compound manifestations we also present a list of contents. Since contents listings, in some cases, are very long, we have added a show more/show less feature. When presenting the search results, we highlight keywords from the query found in the display by rendering them using bold.

For marking results as relevant or non-relevant, we implemented buttons visualized as up and down arrows underneath the content type icons to the left. The up-arrow indicates a positive response, and the down-arrow a negative response from the user. After marking, the icon area of the result item turns green when positive and red when negative to indicate to the user the markings that he/she has made. Markings can be canceled by selecting the cancel button for each marked item, or by using a cancel button at the top of the results page. When a user is done with a task, the results are submitted using the submit button, which presents a short feedback dialogue where the user can indicate knowledge and confidence and is informed about giving consent by accepting the response to be submitted. Successful submission is indicated by a green checkmark on the results page. Results are stored in the backend database as JSON data. The full interface including the submission dialogue, is shown in Figure 7.

Figure 7. The user interface and submission dialogue.

Search tasks

The search experiment focuses on two categories of search tasks:

The first is a combination of <u>author name and title</u> or significant words in the title. This compares with the concept of known item search, although it may be better characterized as a known entity search. The intended context is a user searching for a known title by a known author, with the purpose of accessing a specific publication.

Searches included are:

1. You want to read Lord of the Rings by J.R.R. Tolkien and have made a search using title and author's name.
2. You want to read Murder on the Links by Agatha Christie and have made a search using title and author's name.
3. You want to read Murder on the Orient Express by Agatha Christie and have made a search using title and author's name.
4. You want to explore what is available by Mark Twain on the character Tom Sawyer and have made a search using author's name and name of the character.
5. You have been told by a friend to read the story Terminal Beach by J.G. Ballard and have made a search using title and author's name.
6. You want to read the novel Mysterier by Knut Hamsun and have made a search using title and author's name.
7. You want to read All the Pretty Horses by Cormac McCarthy and have made a search using title and author's name.

To capture different aspects of "usability," we created two subtasks for author-title searches with different wordings:

1. Based on your initial impression, mark the results that you find most interesting in the context of this search. Mark at least one, but feel free to mark more.
2. Mark the results you find most useful for identifying what is available for this title. Mark at least one, but feel free to mark more.

Respondents are presented with one of each of these subtasks combined with a query randomly selected from the list of predefined queries. The WEMI terminology is not used in the wording of the search tasks and none of these modeling constructs are explained to the users in the preliminary information seen when entering the system. This was to avoid having the respondents' behavior affected by knowledge of theory.

The second category of search is using only the author's name, which also can be described as a known entity search but is more exploratory or open.

- You are compiling a list of titles by "NN" for a personal reading list. The result page shows a selection of what can be found. Mark results you would include in your list of titles by this author. Mark at least one, but feel free to mark more.

The wording is chosen deliberately to encourage users to look for titles they naturally would conceptualize as "by this author" and make references to in a discourse. We wanted to avoid asking explicitly for "works by" as this wording, in preliminary user testing, proved to be interpreted very differently. Authors included are the same as listed in the queries above:

- J. R. R. Tolkien
- Agatha Christie
- Mark Twain
- J. G. Ballard
- Knut Hamsun
- Cormac McCarthy

During the preparation, different ways of wording the tasks were tried on trial users (colleagues) and revised based on their feedback. The actual query and results for these tasks are designed to give a representative results listing including various types of collections, parts of collections, and individual results that are not elements of any collection. We limited the search to text as content type and to a single language to avoid ambiguity in the task with respect to variant expressions. The searches are further tailored to produce a limited results set to avoid users being overwhelmed by results. To compensate for any bias caused by position and order of results, we randomly select a sort order based on either fulltext, pagerank, or random sort.

After marking results according to the task, users are presented with a dialog for entering additional information and giving final consent to store the data upon submission. The data collected is completely anonymous and requires no ethical approval according to the Norwegian regulations managed by the Norwegian Agency for Shared Services in Education and Research.

Recruitment of respondents was based on a convenience sampling method. We have, on the one hand, targeted users with little or no specialized knowledge in library data and library search systems by announcing the survey on social media and distributing it by email to contacts. On the other hand, we invited participants with library expertise, both practitioners, educators, and students. Most participants were recruited from Norway, but among the participants we also find others.

Since the user's identity is anonymous and session-based, we are only able to count individual participation sessions, which is reasonably comparable to individual participants since it is less likely that users would

retake the survey. The survey was online during December 2023 and January 2024.

Results and analysis

Implementation experience

Elaborating the records in the collection included the tasks of interpreting records, identifying works, coding parts as analytical added entries (in the MARC record) and adding contents notes for all compounds. The downloaded records reflect variable practice in describing compound publications both because they are retrieved from different library catalogs with different practices and because they were created over a long span of time. In general, the contents are more likely to be described when the parts are distinct novels with an established identity, and less likely to be described in collections of short stories. Distinct novels that are part of a collection are sometimes described in a contents note, as analytical added entries, or sometimes only mentioned in the title statement. Short stories in a collection are rarely described in ways other than through contents notes, although authors are sometimes listed using added entries.

The task of finding existing URIs for the works was rather straightforward for most individual novels using VIAF or Wikidata. Most of the novels in the test collections have a comprehensive publication history and are likely to have a global identity, either automatically or manually created. For the different compounds, however, we rarely find any preexisting URIs unless it is a well-known work with parts, such as trilogies, or a well-established short story collection with a significant publication history. Short stories are, in general, poorly covered by VIAF, and other sources such as Wikidata are more relevant for retrieving identifiers for these. For the Ballard test collection, we used URIs to the Internet Speculative Fiction Database, since it is a comprehensive database of short stories and published collections within this genre.

Our display is based on expressions as the main entries in the results listing, with additional information retrieved from the associated work and with a subordinate listing of associated manifestations. Additionally, we present all compound entities as separate entries in the results listing regardless of the type they represent. The main motivation for this choice is to avoid a potential discrepancy between what a user is searching for and what is presented in the list. If a user searches for the title of an ad hoc collection, it is natural to present the collection as an individual result entry, rather than only the parts of the collection. In practice, we have found that it is difficult to distinguish between established collections and

ad hoc collections as this is a matter of publication history, degree of referencing, availability of established identifiers, etc. Both categories of collections are presented in the same way. Expressions of works with parts, on the other hand, are easier to recognize. These are displayed in a manner comparable to other compounds, but with the relationships to the parent displayed for each part.

Listing the contents of compounds is another implementation issue that turned out to be challenging to implement. In traditional library records, this is often recorded as formatted contents notes whereas in a linked data representation this ideally should be recorded as a manifestation with multiple "embodies" relationships to the content expressions. Unfortunately, the support for ordered properties is somewhat problematic in linked data. Although RDF has support for sequences of properties, it is difficult to create such sequences consistently from existing MARC records, as the order of added entry fields may be random. It is currently also challenging to utilize such sequences when querying and reading the data due to insufficiency in SPARQL. An additional challenge is that a proper listing of contents also needs to include titles that are not often treated as distinct expressions, such as forewords. Our solution was to rely on contents notes for all types of collections, in addition to the explicit representation of relationships. These two representations serve different purposes. The entity representation is used for indexing and determines what to display, whereas the descriptive version is used in the presentation of entities. For all types of collections, we present the contents listing as expression information, and not with each manifestation. For part-of compounds, we present the contents listing at the manifestation level as these may have a different set of embodied expressions. Relying on contents notes also makes it easier to record compact contents notes in line with ISBD formatting rules.

Another important lesson learned from the implementation is the need to differentiate between the part-of relationship that exists between the whole and the parts for works with parts, and the relationship that is expressed between the components and the whole for established and ad hoc collections. These relationships need to be presented differently. Whereas the first will typically be informative to end users, e.g., as a link from the part to the whole, we do not provide the same link for components of established or ad hoc collections. A particular short story may be included in numerous established and ad hoc collections, and it will not be convenient or user friendly to add links from the story to all the collections it appears in. In our implementation, we use the RDA registry properties for works with parts and also record part-of relationships at the expression level. In the presentation of part-of relationships, we first look for

expression level properties, and if not present, we use the work level property. This allows us to easily present the whole using a title in the same language as the part. For ad hoc and established collections, we only record and use the aggregates property on the expression level, although it can be discussed whether this is the intended use of this relationship property.

The user study

At the end of the survey period, a total of 314 task submissions had been recorded. The data includes 92 participants identified by unique IDs, but this is not an exact number of individual participants, as the system allows the same user to revisit the survey system, receiving a different participant ID if the web browser tab had been closed between visits. Each task was submitted separately, and users may not have completed all the assigned tasks during their session.

Characteristics of the sample, based on self-reporting from the submission dialogue, are shown in Figure 8.

Results from the two first tasks are presented using the following data:

- T_1 is the count of respondents that have marked this result when answering task 1 (most interesting)
- T_2 is the count of respondents that have marked this result when answering task 2 (most useful)
- T_{1+2} is the sum counts for task 1 (T_1) and task 2 (T_2)
- K_{at} = Average of respondent's reported knowledge of author or title on this task, on a scale from 1 (poor) to 5 (excellent)
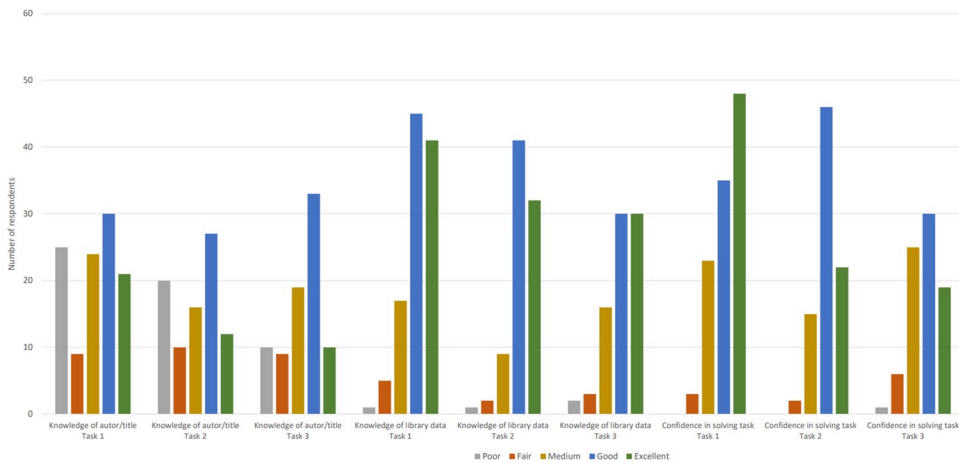

Figure 8. Distribution of participant responses for each task.

- K_{ld} = Average of respondent's reported knowledge of library data on this task, on a scale from 1 (poor) to 5 (excellent)
- T_c = Average of respondent's reported confidence on this task, on a scale from 1 (poor) to 5 (excellent)

Although the data gathered in this study is quantitative and descriptive by nature, each task has a limited number of respondents and statistical analysis beyond simple aggregation is less relevant. The analysis applied is more qualitative and exploratory. The general observed trends, such as which entries are most frequently selected and which are less frequently selected, are discussed in the context of the task, how the results are presented, the type of compound, and other reasonable explanations.

To distinguish between types of results, we categorize results as:
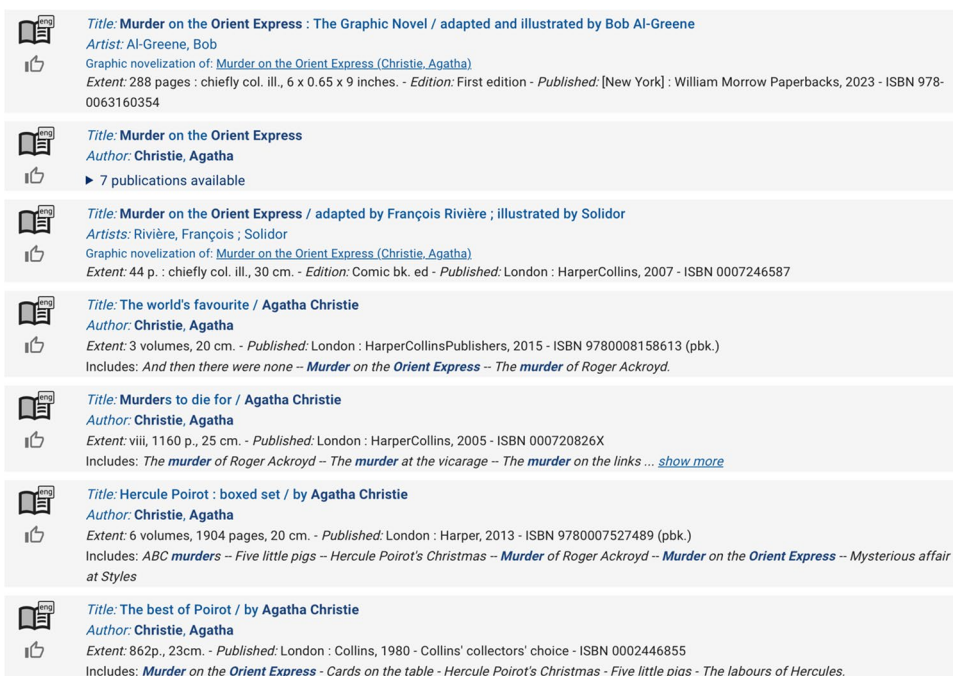
- ad hoc and established collections are labeled as "compound"
- works with parts are labeled as "parents"
- results that appear as parts of any type of collection are labeled as "part"
- results that appear as single embodied expressions are labeled as "standalone"

In some cases, a title is both standalone and part, which implies that it is published both individually and separately, such as the separate parts of *The Lord of the Rings*. Results labeled only as parts are not published individually among the publications in our example data. Typically, these are short stories, although there are also examples of short stories in our data that appear as individual publications.

Screenshots for the selected tasks are shown in Figures 9–11 to illustrate the results that users are presented with and the various cases.

Search 1: All the Pretty Horses, Cormac McCarthy

In this task (Table 2), we find that most users select the result that corresponds with the title in the given task, but additionally also select the compound it is a part of (the trilogy). It is rather evident from the contents listing and part-of relationships that *All the Pretty Horses* is part of *The Border Trilogy* and consequently both will match with the described information need. Results demonstrate the applicability of proper information about compounds and their parts for users to make informed decisions. No major difference is found between the two subtask questions, although some users make selections that are not directly relevant to the task when they are asked to select what is the most interesting. A total number of 20 participants completed this task.

Title: **Murder** on the **Orient Express** : The Graphic Novel / adapted and illustrated by Bob Al-Greene
Artist: Al-Greene, Bob
Graphic novelization of: Murder on the Orient Express (Christie, Agatha)
Extent: 288 pages : chiefly col. ill., 6 x 0.65 x 9 inches. - *Edition:* First edition - *Published:* [New York] : William Morrow Paperbacks, 2023 - ISBN 978-0063160354

Title: **Murder** on the **Orient Express**
Author: **Christie, Agatha**
▶ 7 publications available

Title: **Murder** on the **Orient Express** / adapted by François Rivière ; illustrated by Solidor
Artists: Rivière, François ; Solidor
Graphic novelization of: Murder on the Orient Express (Christie, Agatha)
Extent: 44 p. : chiefly col. ill., 30 cm. - *Edition:* Comic bk. ed - *Published:* London : HarperCollins, 2007 - ISBN 0007246587

Title: The world's favourite / **Agatha Christie**
Author: **Christie, Agatha**
Extent: 3 volumes, 20 cm. - *Published:* London : HarperCollinsPublishers, 2015 - ISBN 9780008158613 (pbk.)
Includes: *And then there were none -- **Murder** on the **Orient Express** -- The **murder** of Roger Ackroyd.*

Title: **Murders** to die for / **Agatha Christie**
Author: **Christie, Agatha**
Extent: viii, 1160 p., 25 cm. - *Published:* London : HarperCollins, 2005 - ISBN 000720826X
Includes: *The **murder** of Roger Ackroyd -- The **murder** at the vicarage -- The **murder** on the links ... show more*

Title: Hercule Poirot : boxed set / by **Agatha Christie**
Author: **Christie, Agatha**
Extent: 6 volumes, 1904 pages, 20 cm. - *Published:* London : Harper, 2013 - ISBN 9780007527489 (pbk.)
Includes: *ABC **murders** -- Five little pigs -- Hercule Poirot's Christmas -- **Murder** of Roger Ackroyd -- **Murder** on the **Orient Express** -- Mysterious affair at Styles*

Title: The best of Poirot / by **Agatha Christie**
Author: **Christie, Agatha**
Extent: 862p., 23cm. - *Published:* London : Collins, 1980 - Collins' collectors' choice - ISBN 0002446855
Includes: ***Murder** on the **Orient Express** - Cards on the table - Hercule Poirot's Christmas - Five little pigs - The labours of Hercules.*

Figure 9. Result page for "Murder on the Orient Express" by Agatha Christie.

Title: The **terminal beach**
Author: **Ballard**, J. G.
▶ 5 publications available

Title: The best of J. G. **Ballard** / J. G. **Ballard**
Author: **Ballard**, J. G.
Extent: 411 pages, 18 cm. --. - *Published:* London : Futura Publications Limited, 1977 - An Orbit book - ISBN 0860079430
Includes: *Introduction (The Best Short Stories of J. G. **Ballard**) / essay by Anthony Burgess ... show more*

Title: The best short stories of J.G. **Ballard**
Author: **Ballard**, J. G.
Includes: *Introduction (The Best Short Stories of J. G. **Ballard**) / essay by Anthony Burgess ... show more*
▶ 2 publications available

Title: The complete stories of J.G. **Ballard** / [J.G. **Ballard** ; introduction by Martin Amis]
Author: **Ballard**, J. G.
Extent: xv, 1199 pages : portrait, 25 cm. - *Edition:* 1st American ed - *Published:* New York ; London : W.W. Norton, 2009 - ISBN 9780393072624 (hbk.)
Includes: *Introduction / by Martin Amis -- Author's introduction -- Prima belladonna -- Escapement ... show more*

Title: Chronopolis : and other stories / by J. G. **Ballard**
Author: **Ballard**, J. G.
Extent: 319 pages, 21 cm. - *Published:* New York : Putnam, [1971]
Includes: *The voices of time -- The drowned giant -- The **terminal beach** -- Manhole 69 -- Storm-bird, storm-dreamer ... show more*

Figure 10. Result page for "Terminal Beach" by J.G. Ballard.

Search 2: Murder on the Links, Agatha Christie

The results for this search (Table 3) exemplify a novel for which there is no individual publication containing only this novel, but it is nevertheless presented as an individual result item. The individual novel is, however, more often selected regardless of this. The title used in the search result listing has a perfect match with the title given in the search task.

Figure 11. Result page for "Mysterier" by Knut Hamsun.

Table 2. Results for search task on All The Pretty Horses, Cormac McCarty ($n = 20$).

Title	Times selected			Mean score			Type
	T_1	T_2	T_{1+2}	K_{at}	K_{ld}	T_c	
All the pretty horses	11	9	20	2.2	4.2	4.2	Part, standalone
The border trilogy	7	5	12	2.6	4.1	4.2	Parent
The crossing	4	0	4	2.2	3.8	3.0	Part, standalone
Cities of the plain	1	0	1	3.0	3.0	3.0	Part, standalone

Note: T_1 – most interesting; T_2 – most useful.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

Table 3. Results for search task on Murder on the Links, Agatha Christie ($n = 16$).

Title	Times selected			Mean score			Type
	T_1	T_2	T_{1+2}	K_{at}	K_{ld}	T_c	
Murder on the links	10	6	16	3.1	4.6	4.6	Part
Poirot : the French collection	6	6	12	3.2	4.7	4.4	Compound
Agatha Christie omnibus: 1920s	6	4	10	3.3	4.5	4.5	Compound
Murders to die for	7	3	10	3.7	4.5	4.5	Compound
Classic tales of mystery	5	3	8	3.6	4.4	4.4	Compound

Note: T_1 – most interesting; T_2 – most useful.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

Although numbers for this task are low, they demonstrate that users will make use of and explore parts that are listed as separate results. Other results selected have in common that the title in the search task is listed in either the title statement or the includes statement. The visibility of the search title, such as its placement in the content listings, seem to

correlate with the frequency of selections. A total number of 16 participants completed this task.

Search 3: Murder on the Orient Express, Agatha Christie

This task confirms the pattern that users prefer result entries that directly match the query (see Table 4). None of the participants have marked the graphical novels, which is understandable given they both are displayed with the artist as the creator and a relationship indicating they are adaptations. The four compounds included have all been selected, but the relevant title may be less visible in these results compared to the *Murder on the Links* case. Content listings must be expanded to reveal the relevant title, and the word "murder" is highlighted in many irrelevant titles. The case indicates that the display of individual expression is valued by the users. A total number of 15 participants completed this task.

Search 4: Terminal Beach, J. G. Ballard

In this case, the query title refers to a short story which is only included in collections (see Table 5). Naturally, most participants choose the result that corresponds directly with the query title. Again, this demonstrates that compound parts have a purpose and are valued as individual entries in results listings. The compounds listed are all ad hoc collections containing the short story. Although all collections would offer access to the same short story, the ones where they have to expand the contents listing are less frequently selected. A total number of 34 participants completed this task.

Table 4. Results for search task on Murder on the Orient Express, Agatha Christie ($n = 15$).

	Times selected			Mean score			
Title	T_1	T_2	T_{1+2}	K_{at}	K_{ld}	T_c	Type
Murder on the Orient Express	10	5	15	4.3	4.1	4.1	Part, standalone
Hercule Poirot : boxed set	4	2	6	4.5	4.5	3.5	Compound
The world's favorite	2	3	5	4.2	4.4	3.8	Compound
The best of Poirot	2	2	4	4.2	4.2	3.2	Compound
Murder on the Orient Express: The graphic novel	1	0	1	5.0	5.0	5.0	Standalone
Murders to die for	0	1	1	3.0	4.0	3.0	Compound
Murder on the Orient Express (graphical novel)	0	0	0	0	0	0	Standalone

Note: T_1 – most interesting; T_2 – most useful.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

Table 5. Results for search task on Terminal Beach, J. G. Ballard ($n = 34$).

Title	Times selected			Mean score			Type
	T_1	T_2	T_{1+2}	K_{at}	K_{ld}	T_c	
The terminal beach	15	19	34	1.6	4.0	4.0	Part
Chronopolis : and other stories	9	10	19	1.9	4.0	4.0	Compound
The complete stories of J.G. Ballard	4	8	12	1.2	4.2	4.2	Compound
The best short stories of J.G. Ballard	8	4	12	1.3	4.3	4.2	Compound
The best of J. G. Ballard	6	5	11	1.5	4.2	3.9	Compound

Note: T_1 – most interesting; T_2 – most useful.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

Table 6. Results for search task on Lord of the Rings, J. R. R. Tolkien ($n = 28$).

Title	Times selected			Mean score			Type
	T_1	T_2	T_{1+2}	K_{at}	K_{ld}	T_c	
Lord of the rings	14	14	28	3.9	4.1	4.1	Part, parent
Lord of the rings : with a readers companion	6	10	16	4.4	4.5	4.1	Compound
The Hobbit ; Lord of the rings	6	8	14	3.8	4.1	4.1	Compound
Fellowship of the ring	4	5	9	4.6	4.2	4.3	Part, standalone
Return of the king	3	2	5	4.2	4.0	4.2	Part, standalone
Two towers	3	2	5	4.2	4.0	4.2	Part, standalone
The war of the ring	0	2	2	4.5	4.0	4.0	Standalone

Note: T_1 – most interesting; T_2 – most useful.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

Search 5: Lord of the Rings, J. R. R. Tolkien

For the search for *Lord of the Rings* (Table 6), we also find that the result with this specific title is selected by most. There is also a reasonably high number of participants selecting the boxed editions, which can probably be explained by the fact that these editions still are dominated by *Lord of the Rings*, both in the title and the included statement. Less participants marked the parts as relevant. In practice, the three parts could replace the compound publication, but end-users indicate them as less relevant in terms of what they immediately find interesting and useful for the specific query. A total number of 28 participants completed this task.

Search 6: Mysterier, Knut Hamsun

The case is based on a Norwegian title by a well-known Norwegian author, and again we find that the result with this specific title is selected by most users (see Table 7). This novel is also included in the compounds presented, all of which are less selected. We may see a slight preference for collections where users understand more quickly that the task title is included. A total number of 47 participants completed this task.

Table 7. Results for search task on Mysterier, Knut Hamsun ($n = 47$).

	Times selected			Mean score			
Title	T_1	T_2	T_{1+2}	K_{at}	K_{ld}	T_c	Type
Mysterier	32	15	47	3.6	4.2	4.2	Part, standalone
Samlede verker, Bind 2	8	7	15	3.7	4.4	4.1	Compound
Samlede Romaner og Fortællinger B. 1	8	6	14	3.6	4.4	4.2	Compound
Hamsuns beste : ungdomsverker	8	5	13	3.2	3.8	3.8	Compound
Victoria ; Mysterier	7	6	13	3.6	4.3	4.3	Compound
Samlede verker, Bind 1	6	5	11	3.3	4.4	4.5	Compound

Note: T_1 – most interesting; T_2 – most useful.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

Table 8. Results for search task on Tom Sawyer, Mark Twain ($n = 16$).

	Times selected			Mean score			
Title	T_1	T_2	T_{1+2}	K_{at}	K_{ld}	T_c	Type
The adventures of Tom Sawyer	8	8	16	2.9	4.1	3.9	Part, standalone
Mark Twain: 10 books in 1	7	6	13	2.8	4.1	4.1	Compound
Mark Twain : selected novels	6	6	12	2.9	4.3	4.1	Compound
Tom Sawyer abroad	4	6	10	2.6	4.4	4.0	Part
Tom Sawyer, detective	4	6	10	2.8	4.4	4.0	Part
Great novels of Mark Twain	5	1	6	2.8	4.3	3.8	Compound
Mark Twain	5	1	6	2.5	4.5	3.8	Compound
The adventures of Huckleberry Finn	2	0	2	3.0	3.5	3.0	Part, standalone
The gilded age and later novels	1	1	2	2.0	5.0	4.5	Compound

Note: T_1 – most interesting; T_2 – most useful.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

Search 7: Tom Sawyer, Mark Twain

This case is slightly different from the others, as the search task relates to a character that is mentioned in various titles, but without a particular perfect match that stands out (see Table 8). No clear conclusions can be derived from the numbers, but if we inspect the two most selected results, they all have in common that "Tom Sawyer" appears in several of the titles listed as contents. Given that the task was to explore what was available, this may be a reason for the slight preference of collections over individually published novels in this case. A total number of 16 participants completed this task.

Author search

For the author search tasks, users were asked to mark results for their own reading list. We find that users do not have any strong preference when it comes to individual expressions vs. collections. We have only

included results from the tasks for Knut Hamsun (28 completed tasks), Agatha Christie (32 completed tasks), and Mark Twain (13 completed tasks) in the listings (Tables 9–11). For Hamsun there is a certain preference for individual expressions, but we also find that users will select compounds quite often. Collections with a more topical title such as "Best of Hamsun. Early works" seem to draw more attention than publications with titles such as collected works, volume X, but the numbers are not conclusive. We also find some individual expressions that are less selected than compounds. For Agatha Christie, we find that users to a certain extent prefer individual expressions but will also select compounds. This may also be caused by these titles looking less interesting or familiar to the users. For Mark Twain we find that the most selected results are compounds, even those with rather generic titles such as "Mark Twain: 10 books in 1." The average knowledge of the author is for all cases above medium, which indicates that many users are familiar with these authors, but for Agatha Christie the familiarity seems to be consistently higher. All in all, results show that users have different preferences for individual expressions vs. compounds, particularly when comparing Agatha Christie with Mark Twain. In general, this indicates that both compounds and parts have a purpose in results listings and that it is difficult to predict what users prefer.

Table 9. Results for search task on Knut Hamsun ($n = 28$).

	Times selected	Mean score			
Title	T_3	K_{at}	K_{ld}	T_c	Type
Benoni	25	3.4	4.2	3.8	Part, standalone
Rosa	25	3.4	4.1	3.8	Part, standalone
August-trilogien	22	3.5	4.1	3.8	Parent
Segelfoss by	20	3.2	4.0	3.6	Part, standalone
Ringen sluttet	19	3.2	4.1	3.6	Standalone
Blandt dyr	16	3.0	4.3	3.8	Part
Markens grøde	16	3.6	4.1	3.9	Part, standalone
Mysterier	16	3.4	3.9	3.8	Part, standalone
Hamsuns beste : ungdomsverker	15	3.3	4.1	3.5	Compound
Hamsuns beste : senere verker	15	3.3	3.9	3.6	Compound
Sult : roman	14	3.6	4.1	3.8	Part, standalone
Samlede verker, Bind 19	14	3.0	4.0	3.6	Compound
Landstrykere	13	3.3	4.2	4.1	Part, standalone
En fløjte lød i mit blod	12	3.2	4.2	3.8	Part
Samlede Romaner og Fortællinger B. 2	11	3.1	4.2	3.6	Compound
Brev til Marie	8	3.2	4.4	3.9	Standalone
Samlede verker, Bind 7	7	2.9	4.0	3.4	Compound
Samlede Romaner og Fortællinger B. 1	4	2.2	3.2	3.2	Compound

Note: T_3 – selected for personal reading list.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

Table 10. Results for search task on Agatha Christie ($n = 32$).

Title	Times selected	Mean score			Type
	T_3	K_{at}	K_{ld}	T_c	
Murder on the Orient Express	24	3.3	4.3	4.1	Part, standalone
Sleeping murder: Miss Marple's last case	23	3.6	4.3	4.0	Part, standalone
Death in the clouds	21	3.5	4.3	4.0	Part, standalone
Murder on the links	21	3.5	4.3	4.0	Part
And then there were none	21	3.6	4.2	4.0	Part
4:50 from Paddington	20	3.6	4.3	4.0	Part, standalone
Secret adversary	20	3.6	4.2	3.9	Part
A Miss Marple quartet	20	3.6	4.4	3.8	Compound
Hercule Poirot : boxed set	19	3.5	4.4	3.6	Compound
Hercule Poirot's Christmas	19	3.5	4.2	3.8	Part
Murder in Mesopotamia	19	3.5	4.3	4.0	Part
A pocket full of rye	18	3.5	4.3	4.0	Part
The Hound of Death	16	3.6	4.3	4.0	Part, standalone
The Fourth Man	16	3.6	4.3	4.0	Part
The Red Signal	16	3.5	4.2	3.9	Part
Murders to die for	15	3.4	4.2	3.5	Compound
The hound of death, and other stories	15	3.5	4.3	4.0	Parent
Cat among the pigeons	14	3.9	4.4	3.9	Part
Poirot in the Orient	13	3.5	4.2	3.2	Compound
The world's favorite	13	3.2	4.4	3.9	Compound
Nemesis	13	3.8	4.3	3.8	Part
Poirot investigates	11	3.9	4.4	3.7	Part
1920s omnibus	11	3.9	4.3	3.3	Compound
Classic tales of mystery	5	3.0	4.6	3.8	Compound

Note: T_3 – selected for personal reading list.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

The impact of users' knowledge and confidence

The tables include data about users' knowledge of bibliographic data, knowledge of author or title, as well as confidence in performing the task. The numbers are presented as average scores among the ones that have selected a particular result. The reported level of knowledge of bibliographic data is in most cases in the range between good (4) and excellent (5), which can be explained by the fact that many respondents are either library students or library professionals. Users familiar with library data and search will more easily be able to interpret the results and make decisions. Knowledge of author or title varies between the search tasks and naturally corresponds with how commonly known an author is. Confidence in solving the task seems to generally correspond with the level of knowledge of bibliographic data, indicating that confidence may be more related to understanding the results presentation than knowledge of the author or title. The data about users is included to contextualize the findings, but more users would be required for a more detailed analysis.

Table 11. Results for search task on Mark Twain ($n = 13$).

Title	Times selected T_3	Mean score K_{at}	K_{ld}	T_c	Type
Mark Twain: 10 books in 1	12	3.2	4.1	3.9	Compound
The gilded age and later novels	11	3.2	4.2	3.8	Compound
The stolen white elephant	10	3.0	3.9	3.5	Compound
Pudd'nhead Wilson: Those extraordinary twins ; The man that corrupted Hadleyburg	9	3.6	4.3	4.0	Compound
The notorious jumping frog of Calaveras County	7	3.4	3.7	3.7	Part
The story of the good little boy who did not prosper	6	2.7	4.5	3.7	Part
Mark Twain : selected novels	6	2.8	4.3	4.0	Compound
The prince and the pauper	6	2.8	4.3	3.7	Part, standalone
A Connecticut Yankee in King Arthur's court	6	3.8	4.5	4.5	Part, standalone
Great novels of Mark Twain	4	3.0	5.0	4.0	Compound
Mark Twain	4	3.0	5.0	4.0	Compound
The gilded age	3	3.7	4.3	4.3	Part, standalone
The celebrated jumping frog of Calaveras County	3	3.7	4.3	4.3	Part
American midnight : tales of the dark	3	2.7	4.7	3.3	Compound
The adventures of Huckleberry Finn	3	3.7	4.3	4.3	Part, standalone
Ghost story	3	4.0	4.7	4.7	Part
Tom Sawyer abroad	3	3.7	4.3	4.3	Part
Tom Sawyer, detective	3	3.7	4.3	4.3	Part
The adventures of Tom Sawyer	3	3.7	4.3	4.3	Part, standalone
Those extraordinary twins	2	4.0	5.0	5.0	Part

Note: T_3 – selected for personal reading list.
K_{at} – knowledge of author/title; K_{ld} – knowledge of library data; T_c – task confidence.

Limitations

Our research method is a combination of experimental system design combined with a user study that collects data from an online prototype. The setup is complex and there are limitations and biases introduced that may influence the results. The overall setup of the experiment, however, also represents a methodological contribution to the field of researching aspects of implementing the IFLA LRM model in library systems.

For the experiments, we used test collections designed for the purpose of this study. Records were selected to illustrate the challenges of compound publications and parts. This reflects what a user may, or ideally should, find, but the overall results presented for each search are not necessarily representative of what users may find when in a real catalog. Particularly, we do not include the nonrelevant results that will often be returned if partial matching is applied and, for some authors, we used a limited selection of their works. Additionally, to avoid confusion caused by variant expressions, records selected were restricted to a single language for some test collections or the results presented were filtered accordingly

if the collection consisted of publications in multiple languages. Given a more comprehensive result listing, users may have applied a different strategy such as selecting less or more results.

The search system also introduces some variables that may impact the results. First, the results presentation is expression-oriented, which is only one possible solution for presenting the results. A work-oriented display would have presented results rather differently. To avoid bias caused by topmost items getting the most attention, different ordering methods were implemented and applied randomly to the participants, but this also means that less relevant results may get the first attention of the user. It is also quite evident that the position of each result has an impact on how users perceive the results set. The user interface highlights search terms found in the results presentation, which will easily draw attention to relevant titles in the contents note. For lengthy contents notes, an expand-hide feature was implemented but this also implies that titles hidden by default are harder to spot than those in the initial display. Other elements in the user interface may also have an impact on how users interpret the results, such as the links presented for relationships or the ordering of agents.

For the participants, the intent was to recruit a balanced user pool, including both ordinary users with no knowledge of library systems or data, as well as users familiar with library systems and data. In the end, it turned out that it was much easier to get responses from persons having an interest in library data and many participants are probably library students or librarians. Whether they have a different understanding of the results presented is not clear, and what kind of bias this may have introduced is not identified. However, most librarians participating would be from public libraries (the target group for the mailing list used). They have knowledge of library data and will be familiar with search and result listings, which also means they are more likely to make qualified selections.

Presenting constructed user needs to participants will always be a challenge compared to users evaluating results according to realistic searches. The use of well-defined queries, some for authors and titles that are generally well known, compensates to a certain degree but there is no guarantee that participants will respond and rank in the same way for a predefined query as they would if the query and information need was their own. We also see that users may have skipped certain tasks, as there are significant differences in the number of responses to some tasks, but we do not consider this to have an important effect.

The system for giving responses may also have had an impact on the results. When testing the system, we explored different alternatives, such as using both positive and negative markings or limiting the number of markings. Preliminary versions of the system were tested and discussed with colleagues. The solution applied in the experiments is what we

identified as the most intuitive compromise. From the results we see that some users have marked most of the results, which may indicate that they are less critical in their selection than what was intended in the survey. An upper limit on the number could have prevented this.

How to phrase the tasks was another challenge that was addressed when developing the experiment. For the title/author task we attempted to indicate interesting and useful as key criteria but have not explored whether users actually apply these criteria in their selection or use a more general understanding of relevance. For the author task, the intended outcome was a list of distinct expressions/works, without asking specifically using any of these terms. Results may indicate that the users understood this task somewhat differently, since the number of compounds selected was significant. The findings, however, indicate quite clearly what entities the users would see as recognizable entities.

Conclusion and future work

The main research questions addressed in this project are:

How can compound entities be expressed in data and utilized in search and result presentations?

The test collections and prototype presented in this study demonstrate that compounds and parts can easily coexist in results presentations and that the main bottleneck is lack of rich bibliographic data for this purpose. The demonstrated systems present expressions with an expandable list of the publications that embody each expression, which appears to be a convenient display technique that is easily supplemented with contents listings on the expression or for each manifestation. The implementation demonstrates that there is a need to clearly distinguish between compounds that are "proper" works with parts and established and ad hoc collections, particularly with respect to the presentation of relationships from the component to the whole which only is relevant for part of works. This indicates that library practice should take care in distinguishing between works with parts and other types of compounds.

What behavior or interpretations can be identified from inspecting users interacting with the results?

For the second research question we have found that users tend to prefer individual expressions when the title corresponds with the search task. When presented as individual entries in result listings, users value and utilize individual expressions. This preference is likely influenced by how

well the expression matches the query in the search task, but individual expressions will be found relevant regardless of whether they appear as separate publications or not. Secondly, users also show an interest in compounds, where they in some way may identify that the content of interest is a part, which indicates the importance of properly describing the contents of publications and systematically utilizing this information in search result presentations. For author search, we do not see any clear pattern in what users prefer. In some tasks, they tend to prefer individual expressions over compounds. In other tasks, results indicate a preference for compounds. All in all, findings indicate that it can be beneficial for end users if both the individual expressions as well as the compounds are presented in result listings.

Future work includes exploring different test cases and user tasks. Other collections may reveal other challenges related to presentation and user preferences. Traditional library collections may not be the ideal examples for exploring how to best implement support for compounds and components, and alternative resources such as dedicated bibliographic databases may be better examples. The user study performed is based on users marking their preferences according to assigned tasks. The system can also be used in combination with interviews or think aloud protocols to gather more qualitative data, which can give insight into how users interpret the various elements of the display and the rationale behind their decisions. The overall methodology of prototyping and gathering online responses to different scenarios can be expanded with additional logging, such as recording whether they inspect the expandable contents note and list of publications. The use of an interactive search system with logging and user responses can easily be applied to other aspects of implementing IFLA LRM. Particularly, we would like to explore how users make use of relationships if they are presented.

Disclosure statement

Funding

ORCID

Trond Aalberg http://orcid.org/0000-0001-5593-0860
Pat Riva http://orcid.org/0000-0001-6024-4320
Maja Žumer http://orcid.org/0000-0002-3756-7524

Notes

1. Pat Riva, Patrick Le Boeuf, and Maja Žumer, *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*, IFLA Publications (International Federation of Library Associations and Institutions (IFLA), 2018), https://repository.ifla.org/handle/123456789/40.
2. Martha M Yee, "What Is a Work? Part 4, Cataloging Theorists and a Definition," *Cataloging & Classification Quarterly* 20, no. 2 (1995): 3–24, https://escholarship.org/uc/item/62x0917w; Chris Holden, "The Bibliographic Work: History, Theory, and Practice," *Cataloging & Classification Quarterly* 59, no. 2–3 (2021): 77–96, https://doi.org/10.1080/01639374.2020.1850589.
3. Jan Pisanski and Maja Žumer, "Mental Models of the Bibliographic Universe. Part 1: Mental Models of Descriptions," *Journal of Documentation* 66, no. 5 (2010): 643–67, https://doi.org/10.1108/00220411011066772; Jan Pisanski and Maja Žumer, "Mental Models of the Bibliographic Universe. Part 2: Comparison Task and Conclusions," *Journal of Documentation* 66, no. 5 (1 January 2010): 668–80, https://doi.org/10.1108/00220411011066781; Jan Pisanski and Maja Žumer, "User Verification of the FRBR Conceptual Model," *Journal of Documentation* 68, no. 4 (2012): 582–92, https://doi.org/10.1108/00220411211239129; Athena Salaba and Tanja Merčun, "Visualizations of Bibliographic Information: A User Experience Study," *Journal of Librarianship and Information Science* 52, no. 1 (2020): 271–87, https://doi.org/10.1177/0961000618804014.
4. Pisanski and Žumer, "Mental Models of the Bibliographic Universe. Part 1"; Pisanski and Žumer, "Mental Models of the Bibliographic Universe. Part 2"; Pisanski and Žumer, "User Verification of the FRBR Conceptual Model."
5. Kim Tallerås, Jørn Helge B. Dahl, and Nils Pharo, "User Conceptualizations of Derivative Relationships in the Bibliographic Universe," *Journal of Documentation* 74, no. 4 (2018): 894–916, https://doi.org/10.1108/JD-10-2017-0139.
6. Sholeh Arastoopoor, "Users' Perception of Navigating Bibliographic Families from IFLA-LRM Perspective," *Library Hi Tech* 40, no. 1 (2022): 265–80, https://doi.org/10.1108/LHT-12-2019-0240.
7. The Working Group on Aggregates, "Final Report of the Working Group on Aggregates," (IFLA, September 2011), https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbrrg/AggregatesFinalReport.pdf.
8. Ed Jones, "The FRBR Model As Applied to Continuing Resources," *Library Resources & Technical Services* 49, no. 4 (2005): 227–42, https://doi.org/10.5860/lrts.49n4.227; Ed Jones, "Modeling of Serials," *Library Trends* 71, no. 1 (August 2022): 90–112, https://doi.org/10.1353/lib.2023.0006; Trond Aalberg, Edward O'Neill, and Maja Žumer, "Extending the LRM Model to Integrating Resources," *Cataloging & Classification Quarterly* 59, no. 1 (2021): 11–27, https://doi.org/10.1080/01639374.2021.1876802.
9. Edward O'Neill, Maja Žumer, and Jeffrey Mixter, "FRBR Aggregates: Their Types and Frequency in Library Collections," *Library Resources & Technical Services* 59, no. 3 (2015): 120, https://doi.org/10.5860/lrts.59n3.120.
10. Damian Iseminger, "Aggregates in the IFLA Library Reference Model and RDA" (Power-Point presentation, Library of Congress, May 2, 2019), https://www.loc.gov/aba/pcc/documents/OpCo-2019/AggregatesPCC.pptx.
11. Jacob Jett, "Towards a General Conceptual Model for Bibliographic Aggregates: Four Case Studies from Our Bibliographic Standards" (PhD diss., Urbana-Champaign, University of Illinois at Urbana-Champaign, 2019); Alan MacLennan and Agnieszka Walicka, "An Investigation into Cataloguers' Experiences with RDA," *Journal of Li-*

brarianship and Information Science 52, no. 2 (June 2020): 464–75, https://doi.org/10.1177/0961000618820655.

12. Barbara B. Tillett, "Bibliographic Relationships," in *Relationships in the Organization of Knowledge*, ed. Carol A. Bean and Rebecca Green, Information Science and Knowledge Management (Dordrecht: Springer, 2001), 19–35, https://doi.org/10.1007/978-94-015-9696-1_2; Allyson Carlyle, "Fulfilling the Second Objective in the Online Catalog: Schemes for Organizing Author and Work Records into Usable Displays," *Library Resources & Technical Services* 41, no. 2 (1997): 79–100, https://doi.org/10.5860/lrts.41n2.79.

13. Shoichi Taniguchi, "Aggregate and Component Entities in RDA: Model and Description," *Cataloging & Classification Quarterly* 51, no. 5 (2013): 580–99, https://doi.org/10.1080/01639374.2013.763316.

14. Library of Congress, "BIBFRAME - Bibliographic Framework Initiative," accessed January 15, 2024, https://www.loc.gov/bibframe/; Sally McCallum et al., "From Idea to Implementation: BIBFRAME Becomes Reality," in *Charleston Conference Proceedings 2022* (Charleston Information Group, LLC, 2023), https://doi.org/10.3998/mpub.14369519.

15. Schema.org Community Group, "Documentation - Schema.Org," accessed January 15, 2024, https://schema.org/docs/documents.html.

16. Wikidata, "Wikidata," accessed January 15, 2024, https://www.wikidata.org/wiki/Wikidata:Main_Page.

17. Ian Bigelow, "Conducting the Opera: The Evolution of the RDA Work to the Share-VDE Opus and BIBFRAME Hub," in *Linking Theory and Practice of Digital Libraries*, ed. Gianmaria Silvello et al., vol. 13541, Lecture Notes in Computer Science (Cham: Springer International Publishing, 2022), 335–43, https://doi.org/10.1007/978-3-031-16802-4_30.

18. Jim Hahn and Tiziana Possemato, "The Share-VDE Ontology: A BIBFRAME Extension for Linked Data Discovery," Zenodo, September 10, 2023, https://doi.org/10.5281/ZENODO.8332350.

19. Carlyle, "Fulfilling the Second Objective in the Online Catalog"; Allyson Carlyle, "Ordering Author and Work Records: An Evaluation of Collocation in Online Catalog Displays," *Journal of the American Society for Information Science* 47, no. 7 (1996): 538–54, https://doi.org/10.1002/(SICI)1097-4571(199607)47:7<538::AID-ASI6>3.0.CO;2-V.

20. Network Development and MARC Standards Office Library of Congress, "FRBR Display Tool, Version 2.0," accessed January 15, 2024, https://www.loc.gov/marc/marc-functional-analysis/tool.html.

21. OCLC Research, "FictionFinder: A FRBR-Based Prototype for Fiction in WorldCat," accessed January 15, 2024, https://www.oclc.org/research/areas/data-science/fictionfinder.html.

22. Yin Zhang and Athena Salaba, "What Do Users Tell Us about FRBR-Based Catalogs?" in *The FRBR Family of Conceptual Models* (Routledge, 2013); Jan Pisanski and Maja Žumer, "Functional Requirements for Bibliographic Records: An Investigation of Two Prototypes," *Program* 41, no. 4 (2 2007): 400–417, https://doi.org/10.1108/00330330710831602.

23. Trond Aalberg and Maja Žumer, "The Value of MARC Data, or, Challenges of Frbrisation," *Journal of Documentation* 69, no. 6 (2013): 851–72, https://doi.org/10.1108/JD-05-2012-0053.

24. Tanja Merčun, Maja Žumer, and Trond Aalberg, "FrbrVis: An Information Visualization Approach to Presenting FRBR Work Families," in *Theory and Practice of Digital Libraries*, ed. Panayiotis Zaphiris et al. (Berlin, Heidelberg: Springer Berlin

Heidelberg, 2012), 504–7; Tanja Merčun, Maja Žumer, and Trond Aalberg, "Presenting Bibliographic Families Using Information Visualization: Evaluation of FRBR-Based Prototype and Hierarchical Visualizations," *Journal of the Association for Information Science and Technology* 68, no. 2 (2017): 392–411, https://doi.org/10.1002/asi.23659; Salaba and Merčun, "Visualizations of Bibliographic Information: A User Experience Study"; Athena Salaba, Tanja Merčun, and Trond Aalberg, "Complexity of Work Families and Entity-Based Visualization Displays," *Cataloging & Classification Quarterly* 56, no. 7 (2018): 628–52, https://doi.org/10.1080/01639374.2018.1529008.

25. Trond Aalberg, Tanja Merčun, and Maja Žumer, "Interactive Displays for the Next Generation of Entity-Centric Bibliographic Models," in *Digital Libraries: Data, Information, and Knowledge for Digital Lives*, ed. Songphan Choemprayong, Fabio Crestani, and Sally Jo Cunningham, vol. 10647, Lecture Notes in Computer Science (Cham: Springer International Publishing, 2017), 199–211, https://doi.org/10.1007/978-3-319-70232-2_17.

26. Paul Johannesson and Erik Perjons, *An Introduction to Design Science* (Cham: Springer International Publishing, 2014), https://doi.org/10.1007/978-3-319-10632-8.