

Model for Aggregating Manga Metadata Across Institutions: Improving the Granularity of Manga Bibliographic Data on the Semantic Web

Senan Kiryakos

Graduate School of Library, Information and
Media Studies University of Tsukuba

August 2015

TABLE OF CONTENTS

| | |
|---|-----------|
| TABLE OF CONTENTS | II |
| LIST OF FIGURES | IV |
| LIST OF TABLES | VI |
| | |
| 1. INTRODUCTION | 1 |
| 1.1 Background | 1 |
| 1.2 The Problem with Manga Bibliographic Data | 3 |
| 1.3 Research Goals and Objectives | 5 |
| | |
| 2. LITERATURE REVIEW AND RELATED WORKS | 7 |
| 2.1 Semantic Web and Linked Data | 7 |
| 2.1.1 Bibliographic Data and the Semantic Web | 7 |
| 2.1.2 BIBFRAME | 8 |
| 2.1.3 Data Aggregation | 10 |
| 2.2 Manga and Related Formats | 12 |
| | |
| 3. DATA COLLECTION & MODELING | 14 |
| 3.1 Manga Bibliographic Data Collection | 14 |
| 3.1.1 Data from Toppan | 14 |
| 3.1.2 Data from Monash | 15 |
| 3.1.3 Data from US Academic Libraries | 16 |
| 3.1.4 Data from the Web | 17 |
| 3.2 Property Identification & Content Analysis | 18 |
| 3.2.1 Toppan Data Analysis | 18 |
| 3.2.2 Monash Data Analysis | 26 |
| 3.2.3 US Academic Library Data Analysis | 28 |
| 3.2.4 Web Data Analysis | 32 |
| 3.3 Methods of Bibliographic Description and Aggregation | 34 |
| 3.3.1 Existing methods of Bibliographic Description | 34 |
| 3.3.2 Bibliographic Description Using BIBFRAME | 35 |
| 3.3.3 Aggregating Data using EDM | 40 |
| | |
| 4. MODELING BIBLIOGRAPHIC DATA FOR MANGA & ITS AGGREGATION | 42 |
| 4.1 Aggregation Model for Manga Metadata | 42 |
| 4.2 Bibliographic Data for Manga | 48 |
| 4.2.1 Manga Volumes Described in BIBFRAME | 48 |
| 4.2.2 Manga Works Described in EDM and DC | 53 |
| 4.3 Identifying Related Manga Data | 56 |
| | |
| 5. SUMMARY AND CONCLUSIONS | 62 |
| 5.1 Significance | 62 |
| 5.2 Limitations | 63 |

| | |
|---|----|
| 5.3 Recommendations | 65 |
| 5.4 Future Work | 66 |
| 5.5 Conclusion | 67 |
| ACKNOWLEDGEMENTS | 68 |
| REFERENCES | 69 |
| APPENDICES | 72 |
| A. Full Aggregation Model | 72 |
| B. Full Aggregation Model – Generalized | 73 |
| C. Aggregation Model – Group 1 | 75 |
| D. Aggregation Model – Group 2 | 76 |
| E. Aggregation Model – Group 3 | 77 |
| F. Aggregation Model – Group 4 | 78 |
| G. Aggregation Model – Group 5 | 79 |
| H. Aggregation Model – Group 6 | 80 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1: Screenshot of Toppan's Tankoubon Zenkan file. | 15 |
| Figure 2: Screenshot of Monash's database file. | 16 |
| Figure 3: Record display for Bleach from Stanford University library's catalogue | 17 |
| Figure 4: MARC data for Stanford's "Bleach" record | 17 |
| Figure 5: Screenshot of DBpedia's English page for the manga "One Piece" | 18 |
| Figure 6: The facet menu in OpenRefine | 21 |
| Figure 7: Results from two facet operations | 21 |
| Figure 8: Original Monash JapaneseTitle column | 26 |
| Figure 9: Split column function inside OpenRefine | 27 |
| Figure 10: Data separated into the Japanese title and its reading | 28 |
| Figure 11: Possible Linked Data opportunities found inside library records for manga | 31 |
| Figure 12: A portion of DBpedia properties for the Wikipedia entry for "One Piece" | 32 |
| Figure 13: Dual language chapter titles and a Volume 1 summary, from http://en.wikipedia.org/wiki/List_of_One_Piece_chapters_(1-186) | 33 |
| Figure 14: BIBFRAME model showing a Work to Instance relationship, and authority resources (Miller, et al. 2012) | 36 |
| Figure 15: A mapping of FRBR entities to BIBFRAME's Work and Instance (Library of Congress, 2014b) | 38 |
| Figure 16: An RDF example of BIBFRAME Authority properties (Library of Congress, 2014) | 38 |
| Figure 17: Holding information and a review attached to a BIBFRAME Instance and Work, respectively (Miller, et al. 2012) | 39 |
| Figure 18: The three core classes in EDM and the properties connecting them (Isaac, 2013) | 40 |
| Figure 19: A single provider's aggregation, with descriptive data attached to the ore:Proxy (Isaac, 2013) | 41 |
| Figure 20: Overview of the full aggregation model | 43 |
| Figure 21: The aggregation of data from multiple sources for a single manga volume | 44 |
| Figure 22: Connecting successive volumes of manga in the aggregation model | 45 |
| Figure 23: Portion of the aggregation model showing the Work level entities | 47 |
| Figure 24: Toppan data described using BIBFRAME properties | 49 |
| Figure 25: Monash data described using BIBFRAME properties | 52 |
| Figure 26: DBpedia's data attached to the Work level resource | 55 |
| Figure 27: Reconciling the Title column of Monash's data | 57 |
| Figure 28: Recommended DBpedia properties inside of the Reconciliation menu based off of the selected column data | 57 |

| | |
|---|----|
| Figure 29: A list of possible matching DBpedia articles based on the data value “Detective Conan” | 58 |
| Figure 30: A DBpedia URL column added based on the reconciled URLs from the Title column | 59 |
| Figure 31: Use of the <i>cell.cross</i> function in OpenRefine | 60 |
| Figure 32: Retrieved ID value from Toppan’s data that matches Monash’s reconciled DBpedia URL | 61 |

LIST OF TABLES

| | |
|---|----|
| Table 1. Toppan data properties in Japanese and English | 19 |
| Table 2. Levels of filled properties in a Toppan data file | 22 |
| Table 3. Used Toppan headers and their mapped BIBFRAME properties | 50 |
| Table 4. Monash headers and their BIBFRAME properties | 52 |
| Table 5. DBpedia properties and their respective EDCM DC properties, along with the literal values taken from the “Astro Boy” DBpedia pages | 54 |

1. INTRODUCTION

Manga, a Japanese style of comic, is enormously popular in Japan, and increasingly so overseas. As is common with other pop culture media forms, manga has not typically been the subject of academic interest, though with the growing appreciation for its historical and cultural significance, this is changing. As appreciation and academic interest in manga grows, so to does the need for cultural heritage institutions, such as libraries, to begin to better describe their manga collections. Some of this descriptive data, however, already exists, created by other special institutions, libraries with a focus on manga, or by hobbyists on the Web. This bibliographic data can be found in varying levels of granular detail, typically dependent on the institution creating the data. Often, these institutions are describing the same manga resources, so the exchange of data would be useful for all parties involved, as they could utilize bibliographic data created by others to enhance their own. This data, however, typically exists in isolation from other sources.

In an attempt to bring together these various descriptions about similar resources, this thesis presents a conceptual model for the aggregation of bibliographic data for manga in Linked Data formats using Semantic Web technologies. Data sourced mainly from Monash University's Japanese Studies Centre (JSC) Manga Library, herein referred to as Monash, and Toppan Printing Co. Ltd., herein referred to as Toppan, along with some academic library data from the US and some Web data, was collected and analyzed to identify their level of descriptive granularity, and identify similar resources with which to aggregate data for. New metadata schemas utilizing Dublin Core (DC), Europeana Data Model (EDM), and BIBFRAME vocabularies were then applied to the data, allowing for its modeling as Linked Data, and aggregation based on EDM. The final result is a conceptual model that enables the aggregation of institutional metadata for manga, connecting data across institutions for both individual manga volumes, and higher-level, conceptual FRBR "Work" level manga resources, while providing the basis for possible future Semantic Web applications, such as a manga data Web portal or authority resource.

1.1 Background

Until recently, the consumption of manga outside of Japan has been limited to personal collections of so-called "otaku" – a term used, in the West, used to describe someone with a love for various forms of Japanese pop culture. The inclusion of manga inside institutions such as libraries has typically been relegated to special collections focusing on Asian material, or small collections in public libraries aiming to serve the recreational reader. Thanks to both a constantly increasing interest in Japanese pop culture and growing appreciation and

acceptance of various sub-cultures, like comic books and video games, there is also a growing interest in manga by academic libraries and other institutions. Collections of manga, graphic novels, and comic books at various US University libraries, such as Cornell and Ohio State, now number in the thousands. The Japanese Studies Centre's Manga Library at Monash University contains over 7000 manga and related items. This increased scholarly interest is not limited to the West, however, as even in Asia, manga has historically not been the subject of much academic interest. Large collections, like those at Meiji University in Japan and a new 20,000-item collection at Peking University in China (Jiji) are now being established as academic interest in manga rises.

Bibliographic data for manga also exists in various forms outside of cultural heritage institutions. Corporations involved with the production or distribution of manga may maintain databases for business or archival purposes. Toppan Printing Co. Ltd., a major Japanese commercial printing and publication service provider, and one of the data providers in this study, maintains several databases containing data for manga in which they have had a role in the manufacture of. While typically these corporate databases are private, Toppan's was seen as significant enough by the Japanese Government's Agency for Cultural Affairs, that a project to publish and work with the database has been launched – the beginnings of which can be found at <http://mediaarts-db.jp>.

Finally, bibliographic data for manga exists on the Web, created mainly by hobbyists and consumers of manga. Some of this data resides in dedicated fan-sites, like www.mangaupdates.com, which boasts perhaps the most thorough collection of fan-created manga series and creator information on the Web – in English or Japanese. A significant amount of data also exists in Wikipedia, which, unlike any resource mentioned thus far, is open for Linked Data usage thanks to DBpedia, which is a site offering structured data versions of Wikipedia pages.

While bibliographic data for manga exists in some forms on the Web, this thesis deals with the lack of manga data specifically on the Semantic Web. While the Web can be said to consist mainly of documents and hypertext, the Semantic Web is a “Web of data” allowing for greater machine-readability, data interaction, the building of ontologies, and more (W3C, 2015a). The Semantic Web is essentially a set of rules and tools for structuring the enormous amount of data that exists on the web. The application of Semantic Web technologies to Web data, structuring it and applying standards in a way that allows for relationships to be made amongst interrelated datasets, can be referred to as Linked Data (W3C, 2015b). The principles for structuring this interrelated data, which are 1) the use of universal resource identifiers (URIs) as names for things, 2) the use of HTTP URIs so those names can be looked up, 3)

providing useful information when those names are looked up, and 4) include links to other URIs so more things can be discovered (Berners-Lee, 2009), represent an ideal for an interconnected “Web of data” that the Semantic Web and Linked Data aim to achieve. The previously mentioned DBpedia is perhaps the most popular example of a Semantic Web project, and is currently home to the only amount of structured Semantic Web data for manga.

While the lack of quality bibliographic data for manga, and its isolation from other resources, can be seen as problematic, the emergence of Semantic Web technologies and a varying granularity of data existing for identical manga resources represents an opportunity to address multiple issues simultaneously. The aggregation and connecting of differently sourced bibliographic data for manga not only opens this data up for exchange and reuse as Linked Data, but also allows institutions, corporations, hobbyists, and others, to utilize already existing data to add to their own, enhancing the representation of data for the entire manga fan community.

1.2 The Problem with Current Bibliographic Descriptions of Manga

Currently, manga is described in different ways by different institutions. As libraries have historically viewed manga as a resource that perhaps was not worthy of much attention outside of serving recreational readers, the level of description detail is rather shallow. Authority data, such as author and publisher names, is fairly strong, but library manga records typically lack any further granular information. For some resources, this may not be a particularly pressing issue, but fans of pop culture resources, like manga or western comic books, are typically quite interested in the minutiae of the medium (Fee, 2013), and the more detailed the catalog record, the better the needs of the patron would be met, be it a recreational reader or academic researcher. More detailed information can be found in the records of special libraries with a manga focus, corporate databases that handle manga, or hobbyist resources on the Web, but the exchange and reuse of this data is typically not enabled, hindering the betterment of records through the use of already existing data. The interest in minutiae by fans, along with varying levels of description depending on institution type, make a resource such as manga different from a scholarly journal or other more common serial type.

These different institutions are often describing the same resources, but in different ways, or to a varying degree of detail. As stated previously, an academic library may pay close attention to details such as author names in accordance with their authority files, but titles may be described in restrictive terms due to cataloguing formats and rules. Moreover, any additional details, such as plot summaries, are also generally not included; correspondence with multiple cataloguers in US University libraries has revealed that they would like to

include this type of information, but it is uneconomical to do so, and, at least, the information can be found elsewhere online. These online resources, such as dedicated fan-sites or Wikipedia contain more detailed information such as volume summaries, character information, genres, and so on, which would be of use to the manga fan patrons of libraries or other professional institutions. This data exchange would also be of use to the hobbyist resources, as they could utilize, and perhaps build on, library authority files, in strengthening the quality of their data using institutional controlled vocabularies.

In-between the general library and the hobbyist resource are institutions such as manga focused libraries or databases held by corporations that handle manga to some degree. The level of detail in bibliographic records created by these institutions are typically similar to those created by libraries, but the lack of rigid cataloguing restrictions allows for some important added descriptive details. These include properties such as translated or alternate titles, which fans of manga would find beneficial, not only as minutiae, but as even large Western manga collections, such as those at Stanford, Cornell, or Ohio State Universities, feature a significant number of Japanese language manga.

The “level” of manga instance being described also tends to vary based on institution type. In this study, level refers to different entities described most commonly by the Functional Requirements for Bibliographic Records (FRBR) model (IFLA, 1998). A library catalogue with entries for specific manga volumes or corporate database handling information about specific publication instances can be said to be describing manga in terms resembling the FRBR “Manifestation” or “Item” entities. At this level, one is able to not only describe broader properties such as author or main title, but also specifics like volume summaries, chapter titles, physical dimensions, and others. A more general library catalogue series record or Wikipedia entry describing the entire intellectual property of the manga describes it in terms closer to the FRBR “Work” level, which can describe a resource more broadly and in terms agnostic of specific publications or volumes. If one wishes to aggregate manga data from different sources, the ability to distinguish between different entity levels and describe them appropriately is important.

The problem addressed in this study is how the bibliographic descriptions of manga that are created by different sources, describe different bibliographic qualities, and are in different languages, can be connected together in a way that enables the use and exchange of this data on the Semantic Web. Recreational and academic interest in manga is growing, and it is sensible – and economical – for institutions to use existing resources to enhance their own. Establishing a basis for the collecting and connecting of bibliographic descriptions for manga in the Linked Data space allows for not only the strengthening of institutional data, but also an

expanded “collective knowledge” for the resource through the bridging of institutional and specialist community data (Gruber, 2008).

While some of the individual problems addressed in this thesis are not manga-specific, the sum of these issues creates a situation that is fairly distinctive when it comes to other materials, and in particular other serials, creating a unique situation regarding the bibliographic description of manga. The interest in minor details by manga readers (Fee, 2013) such as characters, story arcs, etc., is unlikely to be shared by readers of more scholarly serials. Similarly, there are simply more possible details in manga that exist to be described over other serials. This makes aggregation for manga more useful than other serials, as different types of data providers, for example a hobbyist fan-site and a library, are more likely to describe different details of the same resource. Lastly, the successive relationship in a manga series is important and different from most other serials. In typical scholarly journals, the key relationship is the hierarchical one that exists between the journals themselves and the articles contained within. For serialized manga, the key relationship is one of succession between volumes, as the story content within manga typically requires sequential, volume-to-volume reading to be understood – something not required for other serial types. These factors make manga a resource unique from other serial types, thus creating the need for a model that serves the specific needs of the format rather than a generalized model for all serials.

In summary, this thesis seeks to address the lack of manga metadata on the Semantic Web, insufficient Linked Data description methods for manga in order to facilitate Semantic Web use, and absence of bibliographic data sharing and aggregation among various providers for related manga resources.

1.3 Research Goals and Objectives

The primary objective of this research is to establish a metadata aggregation model for manga that is able to aggregate data from different types of institutions, for different FRBR “levels” for manga, and accommodate varying levels of descriptive granularity. In achieving this objective, an analysis of bibliographic data for manga is performed on data sourced from different institutions and of varying levels of detail granularity. New Linked Data methods for describing manga are examined, specifically those utilizing DC, EDM, and BIBFRAME vocabularies, in order to enable current bibliographic descriptions of manga to work with a Linked Data aggregation model, and in general on the Semantic Web. Data from different institutions is compared in order to identify instances of manga data that describe the same resource, and are then aggregated using the model, connecting multiple descriptions about the

same manga resources, while maintaining the volume-to-volume relationship of serialized manga. This aggregation model also aims to serve multiple levels of the conceptual FRBR model, namely the Work and Item levels. Special libraries, for example, typically manage Item level description, cataloguing individual and separate volumes of manga. Other sources, such as main Wikipedia entries, may describe the manga series as a whole, creating data that leans more towards the Work level. The model seeks to allow aggregate descriptions of both sides, while establishing the relationship between them.

As the pool of available Linked Data grows and as libraries and other institutions seek to publish their data on the Web, the assumption is that this aggregation model will allow for manga to be described by institutions in ways that better serve the enthusiastic fans of the medium, while simultaneously utilizing and contributing to hobbyist and other institutional data on the Web. This fan-created data is often the most descriptive, and the creation of this model seeks to leverage this, in conjunction with institutional data, in order to provide a more complete, authoritative, and higher quality level of data for this increasingly popular format.

2. LITERATURE REVIEW AND RELATED WORKS

2.1 Semantic Web and Linked Data

There has been a considerable amount of past study performed examining the Semantic Web and Linked Data in general. While this research is important in understanding the fundamentals and problems with data on the Semantic Web, section 2.1.1 of the literature review will focus specifically on work dealing with bibliographic data and the Semantic Web, typically from a library perspective. As manga is typically catalogued as a type of serial, some research on this format will also be discussed. Section 2.1.2 will discuss works that focus on the BIBFRAME model, as well as works about metadata aggregation.

2.1.1 Bibliographic Data and the Semantic Web

In describing how Linked Open Data (LOD) can be generated using bibliographic information from digital repositories, Konstantinou, Houssos, and Manta use a system consisting of a Resource Description Framework (RDF) quadstore, SPARQL endpoint, and faceted browsing. While the technical details of the authors' methodology are not relevant to this thesis, the realized benefits of transforming bibliographic data into LOD are noteworthy – namely increased discoverability, synthesis, inference, and reusability of data. (Konstantinou, Houssos, and Manta, 2014). The noted drawbacks are important to keep in mind as well, namely native data errors such as concept mismatching, and multiple technology-based issues, which are obstacles relevant to any LD-related project.

Similar benefits can be seen as the guiding force of Southwick's project, which looks at the transformation of digital collections metadata from the University of Nevada, Las Vegas, into Linked Data. The author lists the motivations for the project as the desire to break up the isolated data "silos" in which the digital collections metadata resides, to interconnect their data to data from other providers, and improved search capabilities of relevant data no matter its origin (Southwick, 2015). Similar to this thesis, Southwick's project utilizes both the OpenRefine software tool, and EDM, though their use of EDM is for its properties useful in modeling historical materials rather than the use of EDM as an aggregation model. Despite differences in specific details, the similarity to this thesis, namely in its general goals regarding Linked Data and workflow utilizing OpenRefine, made this a particularly valuable project to examine.

Moulaison and Million discuss the general merits of Linked Data in the library space in the context of whether Linked Data technologies constitute a "disruptive technology" according to a theory by Clayton Christensen (Moulaison and Million, 2014). This theory is

used to understand how one technology eclipses or eliminates an incumbent based on a process of adoption. The authors argue that for libraries, Linked Data is currently not disruptive – that is, not in a position to completely overtake the current bibliographic data technologies used by libraries today. In explaining this, the paper names several criticisms of Linked Data, citing concerns with the lack of widespread adoption, absence in the for-profit space, and technological maturity problems. The author's do, however, see a great “disruptive potential” of Linked Data for libraries. A recommendation to explore this potential is for libraries to utilize linked data for spin-off programs, seeking specific use-cases for Linked Data and performing tasks such as “enriching electronic content they possess with Linked Data.” Data enrichment will be explored in this thesis, and perhaps, due to the wealth of data found for manga outside libraries, there is a strong argument to be made for the use of manga materials in a spin-off operation.

In exploring the process of converting library MARC records into Linked Data, Cole, Han, Weathers, and Joyner, explore the potential for Linked Data in libraries, while addressing the lack of best practices for data transformation. The paper focuses on the process of transformation and the types of services that Linked Data can offer libraries, in their case additional information shown on a catalogue splash page based on links to Linked Data services, such as the Virtual Internet Authority File (VIAF) and the Library of Congress Subject Heading's (LCSH) Linked Data Service (Cole, Han, Weathers, & Joyner, 2013). Apart from the transformation process, much discussion is focused on making MARC data more Linked Data-friendly. The lack of a proper Linked Data carrier for bibliographic metadata, allowing for Semantic Web capabilities while maintaining the information richness and OPAC-suitability of MARC, was the most relevant topic relating to this thesis. The authors settled on a semantic set based on Schema.org, which satisfied their needs with some limitations. BIBFRAME, a semantic set used in this thesis, is mentioned briefly, though the early developmental stage that BIBFRAME existed in at the time of the paper's writing made it unsuitable for use. This thesis, in its experimentation using BIBFRAME, will investigate whether it is a more suitable semantic set in contrast to that selected by the authors.

2.1.2 BIBFRAME

BIBFRAME, standing for Bibliographic Framework, is an initiative started in 2012 by the Library of Congress in the US that is meant to be both a conceptual model for bibliographic description, and a modern replacement for the forty-plus year old MARC cataloguing standard. As a conceptual model, BIBFRAME resembles FRBR in that it consists of four high-level classes, Work, Instance, Authority, and Annotation. Similar to FRBR, a

BIBFRAME Work is meant to be a high level concept of a resource, with Instance representing some form of embodiment of a Work (Library of Congress, 2015a). As a MARC replacement, BIBFRAME relies more on relationships between resources rather than collecting information into a single record. In using identifiers for things such as people and places, emphasizing relationships between resource, and employing an RDF-compatible vocabulary, BIBFRAME is the library community's attempt to integrate their data into the Semantic Web. More BIBFRAME discussion will be had in Section 3.3.2, however a more thorough description of the model can be found in the BIBFRAME primer document (Miller, Ogbuji, Mueller, & MacDougall, 2012). Other information, along with the full BIBFRAME vocabulary, can be accessed via <http://www.loc.gov/bibframe>.

Kroeger (2013) provides a useful literature overview that outlines the reasoning and history behind the MARC to BIBFRAME transition. The paper focuses on important documents that discuss why MARC was insufficient for use in a Web-dominated world, beginning with Tennant's infamous "MARC Must Die!" article (Tennant, 2002), before moving to comment on articles and discussions that led to the creation of BIBFRAME. Many of these early articles theorized about the future of library record in relation to FRBR, and as MARC could not properly express an entity-relationship model (IFLA, 1998), it was clear a replacement was needed. Some, like Karen Coyle, even suggested that bibliographic record functions should go beyond the find, identify, select, and obtain that FRBR suggested, but also include functions such as discovery, description, and promotion (Coyle, 2004). In discussing works that stress the importance of interoperability and data sharing, managing relationships across communities, and the ability for non-library communities to add value to library data, Kroeger's article helps illustrate how vital Linked Data was during the formulation of BIBFRAME. While BIBFRAME is not ready for full implementation, this thesis hopes to act as a use-case to illustrate Kroeger's assumption that a "post-MARC Linked Data future" (Kroeger, 2013, p. 886) is near.

In a discussion about the benefits and challenges for the adoption of Linked Data in libraries, Gonzales sees BIBFRAME as a "substantial step toward the implementation of Linked Data to connect library bibliographic materials with other resources on the web" (Gonzales, 2014, p.16). The challenges to BIBFRAME's adoptions, such as a current lack of large cooperative effort towards implementation, are made clear, but the author views BIBFRAME as having the potential to make possible Linked Data for libraries, giving them the opportunity "integrate their authoritative data with user-generated data from the web" (p. 19), a core objective in this thesis.

Fallgren, Lauruhn, Reynolds, and Kaplan (2014) provide some insight into the current evolving state of Linked Data for Serials. Along with sections on Linked Data's relationship with International Standard Serial Number (ISSN), and Linked Data from the perspective of a serials publisher, Elsevier, the article discusses the use of BIBFRAME with serials. As BIBFRAME is still early in development and uses cases for serials are still few, the authors provide recommendations for what should be kept in mind when attempting to model serials using BIBFRAME. As manga are generally treated as a type of serial for the purposes of cataloguing, several of the questions raised are important in this thesis, and the later modeling of manga using BIBFRAME attempts to address them. These questions include whether different language versions are to be classified as a new BIBFRAME Work or Instance, whether content differences between print and electronic serials represent new Works, and how relationships between serial volumes should be modeled, i.e. whether each volume is a Work or an Instance when looked at hierarchically.

Baker, Coyle, and Petiya (2014) examine multi-entity models for bibliographic description – FRBR, Resource Description and Access (RDA), and BIBFRAME – that have been published as RDF vocabularies, and test the validity of these data structures in RDF, while also commenting on multi-entity bibliographic models in general. In testing the RDF validity of the data models, the authors find issues with all models when implementing unorthodox – but possible – examples of bibliographic description. These errors are mostly technical, i.e. an RDF reasoner not giving an error over incorrect or inconsistent data that is allowed within the model, and are important to keep in mind when attempting to use these new multi-entity bibliographic description models on the Semantic Web. The paper recommends data integrity constraints, made possible in BIBFRAME using “profiles”, as necessary to the success of multi-entity bibliographic description models, not only in the “library silo”, but also for interoperability on the Semantic Web, which is viewed as “not optional but a requirement” (Baker, et al., 2014, p. 581)

2.1.3 Data Aggregation

This section will focus on relevant literature discussion data aggregation on the Web, specifically those with a focus on the use of the Europeana Data Model (EDM). EDM is the underlying data model that serves the Europeana cultural heritage portal. Designed to collect information from various European cultural heritage institutions, EDM is a method of “collecting, connecting, and enriching metadata” (Europeana, 2014a, p.2). The requirements for EDM are 1) distinguish between a ‘provided item’ (painting, book) and its digital representations, 2) distinguish between an item and its descriptive metadata, and 3) allow the

ingestion of multiple records for the same item, which may contain contradictory statements (Europeana, 2014b). Here, ingestion refers to the collection of digital content from European cultural heritage data providers in order to aggregate it on Europeana's portal. While this process is a main function of Europeana, of relevance to this thesis is the data model itself, which will be used as the method of aggregation for manga metadata sourced from different institutions. Sections 4.1.3 and 5.3 of this thesis will discuss the aggregation method in more detail, however those interested in understanding EDM more thoroughly may refer documentation located at (Europeana 2014a) and (Europeana 2014b).

Agenjo, Hernández, and Viedma (2012) look at data aggregation and authority record dissemination using Linked Data, for eventual collection by Europeana. While the article goes into some detail about the Polymath Virtual Library's attempt to aggregate MARC data about authors, of particular interest is the discussion on the enrichment of authority data. Stating that authority records contain name identification, biographical information, occupation, relationships, Web resources, and others, one can compare this ideal with the reality of current library authority records for manga authors, where little more than name identification is provided. The method and content focus differs, but looking to external sources to provide more complete authority data runs parallel with this thesis.

Zapounidou, Sfakakis, and Papatheodorou have authored two related papers of particular relevance to this thesis, as both discuss the use of BIBFRAME within EDM. Central to both papers is the mapping of BIBFRAME core classes to EDM to ensure interoperability when using both models. Specifically, an attempt is made to map BIBFRAME into EDM respecting the relationship between the three EDM core classes: `edm:providedCHO` representing the Work, Expression, and Manifestation entities of FRBR, `edm:webResource` representing the digital representation of the providedCHO, and `ore:aggregation` linking the former two classes (Zapounidou, Sfakakis, and Papatheodorou, 2014a). The initial paper contains a single base mapping of the BIBFRAME core classes to the EDM core classes. The base path is given as "Work – `bf:hasInstance` – Instance" representing an `edm:providedCHO`, along with an attached BIBFRAME annotation, connected using the `bf:hasAnnotation` property, containing the path "HeldMaterial – `bf:electronicLocator` – URI", which represents the `edm:webResource`. In the following work (Zapounidou, Sfakakis, and Papatheodorou, 2014b), the authors extend the former work by mapping to different paradigms, namely a 2012 EDM library metadata alignment report, and an EDM-FRBRoo (FRBR-object oriented) profile. Crucially, this second paper utilizes the `ore:Proxy` class in EDM, which allows for different institutional descriptions of the same providedCHO – a key function in this thesis. While both works focused mainly on monographs and not serials, the core ideas behind the mappings were nonetheless influential

in the formulation of the model used in this thesis. Sections **4.1.3** and **5.3** will look at the models presented in these works in more detail.

2.2 Manga and Related Formats

This section will focus on literature discussing manga and “related formats,” which here refers to resources such as comic books and graphic novels. Comic books specifically, being a serialized pop-culture media format, share similarities to manga, which make examining related works particularly relevant. While literature on topics such as FRBR modeling for manga, or the manga in the Linked Data space has been uncommon, similar works exist for Western comic books. Moreover, in the West, manga are sometimes published in the comic book periodical or graphic novel paperback format and not the standard Japanese publication format (O’Nale, 2010, p. 385), allowing more parallels to be drawn in some publication scenarios. Relevant literature, however, does exist specifically for manga, and these will also be examined.

In looking at bibliographic data for manga, it is important to understand how institutions that have typically been in charge of creating that data handle manga and related formats. With the assertion that fans of these formats “have vastly differing search needs from the standard browsing patron. The more minutiae that can be added to the catalog record the better,” Fee (2013, p. 37) discusses the nature of and solutions to problems with comic and graphic novel classification and cataloguing in libraries. The paper also discusses issues with granularity in comic book records, stating for example that a library may choose to have a single serial catalogue record that includes all volume holdings in a single record, or create individual volume records that allow for a greater level of detail for the item and hand and create more access points for patrons, but may be difficult to create due to staffing and time issues. The importance of these more detailed records, a result of so-called “analytic cataloguing,” and particularly the role of the FRBR model in achieving them in serial records, was earlier stated by Fee (2008), claiming that the greater-detailed record produced by analytic cataloguing is “needed for the comic scholars and fans who come into our libraries looking for these items” (p. 178). As stated earlier in this thesis, contact with cataloguing librarians have indicated that staffing and time is indeed a barrier to these greater-detailed records, which are desired, so making this task easier through the use of available Linked Data resources is worth pursuing. Indeed, Fee also states that Wikipedia is a “surprisingly good source of information on comic book related subjects” (Fee, 2013, p. 39), so the leveraging of this data is logical.

The usefulness of resources found outside of libraries is echoed by Markham (2009) in his discussion of the Portland State Library's cataloguing endeavours with the Dark Horse Comics collection. Rather than simply use LCSH for describing subject and genres for comics, Portland State Library used genres available from Dark Horse's website, such as action/adventure, crime, manga, and superhero (p. 168). Greater descriptive detail for the benefit of patrons can be seen as the main reasoning behind this, with Markham's quoting of Dilevko and Gottlieb (2004) stating, "libraries should provide access to these sources using terminology that describes the specific content" (p. 44) – something that current library resources are not fully capable of achieving for manga.

Morozumi, Nomura, Nagamori and Sugimoto (2009) proposed a metadata framework for manga that examined the format's bibliographic descriptions, structural descriptions, and intellectual entities. In their approach, the authors state that manga should be able to be described in varying levels of granularity, and that there should be a clear distinction between manga as an intellectual entity and as a published instance. Their adoption of the FRBR model for use with bibliographic description of manga allows this, with FRBR's Work representing the intellectual entity level, and FRBR's Manifestation or Item representing the published instance. The approach also calls for the enabling of specific contents inside manga, such as leading characters, and the leveraging of sources such as Wikipedia to achieve this. In a related work, Wikipedia information about manga was accessed using DBpedia as a method of identifying FRBR Works using Linked Open Data (LOD) resources (He, Mihara, Nagamori, & Sugimoto, 2013). The authors used DBpedia as a pseudo-authority in order to identify Work entities of manga housed in the Kyoto International Manga Museum's catalogue. This thesis seeks to build on the ideas of multi-level manga entities and the leveraging of Linked Data resources put forth in these works.

3. DATA COLLECTION AND MODELING

Section 3 will examine the bibliographic records that were used throughout the thesis. Section 3.1 will discuss the acquisition of bibliographic data for manga from various sources, namely US University libraries, Toppan, and Monash University's JSC manga library, as well as a look at data available on the Web. Examples of records from each dataset will also be shown. Section 3.2 will look first at the properties and contents of the records contained within each dataset, then their respective description models and methods.

Section 3, along with those that follow, discuss handling data in various ways using the OpenRefine tool. OpenRefine, open source software formerly known as Google Refine, is "a powerful tool for working with messy data: cleaning it; transforming it from one format into another" (OpenRefine.org, n.d.). This tool was also utilized in a study by Southwick (2015), as well as discussed by Hooland and Verborgh (2014). The initial intended of using OpenRefine was for its ability, in conjunction with an RDF extension, to reconcile data to various web services and transform spreadsheet data into RDF. Detail about these functions will be discussed in Section 4.3, however as the majority of the data acquired was in spreadsheet formats, the software was also useful in manipulating and analyzing data in various other ways which will be discussed where applicable.

3.1 Manga Bibliographic Data Collection

3.1.1 Data from Toppan

The bibliographic records for manga used in this thesis came from several sources, with Monash and Toppan being the two main providers, both maintaining private databases containing bibliographic records for manga. Toppan, a major Japanese printing company, has kept private database files on materials that they have had a role in the production of. This data from Toppan was acquired via the research lab this thesis author is a member of, the metadata research lab at Tsukuba University's Graduate School of Library, Information, and Media Studies (see <http://mdlab.slis.tsukuba.ac.jp/index.html>). As of fiscal year 2014, Toppan is currently involved in the creation of a digital media arts database, a project funded by the Japanese Government's Agency for Cultural Affairs, the beginnings of which can be found at <http://mediaarts-db.jp/>. Toppan looked to the metadata lab for assistance in handling and modeling the data, and so it has been made available for use.

Toppan's data exists as spreadsheet files in .xls and .tsv formats, and consists of data for different pop culture mediums, such as manga, anime, and video games. While multiple files deal with manga, used in this study were there 単行本 (tankoubon, referring to individual

volumes of manga) and 単行本全巻 (tankoubon zenkan, referring to an entire manga series). Details on the properties and contents of Toppan's data can be found in Section 3.2.1, while a glance at what the spreadsheet data looks like can be seen below, in Figure 1.

| | A | B | C | D | E | F | G | H | I | J |
|----|---------------|---------------|--------------------|-----------------------------------|--------------------------------|--------------------|--------------|-----------|-----------|--------------|
| 1 | マンガ単行本全巻ID | マンガ作品ID | マンガ単行本全巻名 | マンガ単行本全巻名3記 | マンガ単行本全巻名追記 | マンガ単行本全巻名追記3記 | マンガ単行本全巻別版表示 | マンガ単行本全巻数 | 全グループビング数 | 責任表示 |
| 2 | MMG0000000001 | MMT0000000079 | TALES OF SYMPHONIA | TALES OF SYMPHONIA / テイルズオブシンフォニア | TALES OF SYMPHONIA#extra load | Tales of symphonia | | 5 | 5 | [著]巻村仁 |
| 3 | MMG0000000006 | MMT0000000079 | TALES OF SYMPHONIA | TALES OF SYMPHONIA | | | | 1 | 1 | [著]巻村仁 |
| 4 | MMG0000000007 | MMT0000000007 | うしろわだるもの | ウツロワザルモノ | プレスオブファニア4 | | | 4 | 4 | [著]巻村仁 |
| 5 | MMG0000000011 | MMT0000000011 | それすらも日々の果て | ソレ スラモ ヒビ ノ ハテ | ソレスラモ ヒビ ノ ハテ | | | 1 | 1 | [著]一条ゆかり |
| 6 | MMG0000000012 | MMT0000000011 | それすらも日々の果て | ソレ スラモ ヒビ ノ ハテ | ソレスラモ ヒビ ノ ハテ | | | 1 | 1 | [著]一条ゆかり |
| 7 | MMG0000000013 | MMT0000000013 | すくらんぶる-えっく | スクランブル エック | | | | 1 | 1 | [著]一条ゆかり / [|
| 8 | MMG0000000014 | MMT0000000013 | すくらんぶる-えっく | スクランブル エック | | | | 1 | 1 | [著]一条ゆかり |
| 9 | MMG0000000015 | MMT0000000015 | ブライド | ブライド | | | | 12 | 12 | [著]一条ゆかり |
| 10 | MMG0000000029 | MMT0000000029 | ラグナロクオンライン | ラグナロク オンライン | あなたを渡し藤 | | | 1 | 1 | [著]七六 |
| 11 | MMG0000000030 | MMT0000000030 | ロマンチックくだけい | ロマンチック クダサイ | | | | 1 | 1 | [著]一条ゆかり / [|
| 12 | MMG0000000032 | MMT0000000030 | ロマンチックくだけい | ロマンチック クダサイ | | | | 1 | 1 | [著]一条ゆかり |
| 13 | MMG0000000036 | MMT0000000036 | 七羽あむ | ナナア ム | | | | 4 | 4 | [著]七羽あむ |
| 14 | MMG0000000041 | MMT0000000041 | キスと後悔 | キスト コウカイ | | | | 1 | 1 | [著]七尾美穂 |
| 15 | MMG0000000042 | MMT0000000042 | 彼女が寝におちる理由 | カノジョ ガ カレ ニ オチル ワケ | カノジョ ガ カレ ニ オチル リユウ | | | 1 | 1 | [著]七尾美穂 |
| 16 | MMG0000000044 | MMT0000000044 | アセンチ | アセンチ | もう一つの世界 | モウ ヒトツ ノ セカイ | | 3 | 3 | [著]七角佳那 |
| 17 | MMG0000000047 | MMT0000000047 | 夢のあとさき | ユメ ノ アトサキ | | | | 1 | 1 | [著]一条ゆかり / [|
| 18 | MMG0000000048 | MMT0000000048 | KICK! | KICK / Kick / キック | We can change! | WE CAN CHANGE | | 1 | 1 | [著]七角佳那 |
| 19 | MMG0000000049 | MMT0000000049 | イヤだなんて言わせない | イヤダ ナンテ イワセナイ | | | | 2 | 2 | [著]七角佳那 |
| 20 | MMG0000000052 | MMT0000000052 | エアコイ | エアコイ / Air Koi / エア コイ | | | | 1 | 1 | [著]七角佳那 |
| 21 | MMG0000000053 | MMT0000000053 | クレイジー LOVEゲーム | クレイジー LOVE ゲーム | クレイジー love ゲーム | | | 1 | 1 | [著]七角佳那 |
| 22 | MMG0000000054 | MMT0000000054 | ココロ スランブル | ココロ スランブル | | | | 1 | 1 | [著]七角佳那 |
| 23 | MMG0000000055 | MMT0000000055 | まきもどしの底の詩 | マキモドシ ノ コイ ノ ウタ | マキモドシ ノ コイ ノ シ | | | 1 | 1 | [著]七角佳那 |
| 24 | MMG0000000056 | MMT0000000056 | 恋じゃないのダ! | コイ ジャ ナイダ | | | | 2 | 2 | [著]七角佳那 |
| 25 | MMG0000000058 | MMT0000000058 | トラファルガー | トラファルガー | | | | 1 | 1 | [著]青地保子 |
| 26 | MMG0000000059 | MMT0000000059 | 女もどち | オンナ トモダチ | オンナ トモダチ | | | 3 | 3 | [著]一条ゆかり / [|
| 27 | MMG0000000060 | MMT0000000021 | 風光る | カゼ ヒカル | | | | 44 | 44 | [原作]七三太朗 / [|
| 28 | MMG0000000063 | MMT0000000063 | 許してあげない | ユルシテ アゲナイ | | | | 1 | 1 | [著]七尾けいな |
| 29 | MMG0000000065 | MMT0000000065 | 天使のツラノカワ | テンシ ノ ツラノカワ | | | | 5 | 5 | [著]一条ゆかり |
| 30 | MMG0000000066 | MMT0000000066 | ドラッパンの騎士 | ドラッパン ノ キン | | | | 1 | 1 | [著]青地保子 |
| 31 | MMG0000000067 | MMT0000000067 | まわり道のエッセイ | マワリミチ ノ エッセイ | | | | 1 | 1 | [著]七尾秋生 |
| 32 | MMG0000000070 | MMT0000000070 | かくも不純なロマンチズム | カクモ フジュンナ ロマンチズム | | | | 1 | 1 | [著]七尾秋生 |
| 33 | MMG0000000072 | MMT0000000072 | ひそやかに加速する欲望 | ヒソヤカニ カソクスル ヨクボウ | A desire speeding up in secret | | | 1 | 1 | [著]七尾秋生 |
| 34 | MMG0000000075 | MMT0000000065 | 天使のツラノカワ | テンシ ノ ツラノカワ | | | | 3 | 3 | [著]一条ゆかり |

Figure 1: Screenshot of Toppan's Tankoubon Zenkan file

3.1.2 Data from Monash

The second major source of data in this study came from Monash University's JSC Manga library. The Manga Library is part of Monash's Japanese Studies Centre, and serves both academic and recreational users of manga. While no public catalogue is available via their website (see <http://yoyo.its.monash.edu.au/groups/mangalib/>), through contact with Tadgh Dinnage, the Manga Library's manager, they were gracious enough to provide their data for use in this study. Like Toppan, the data exists as a spreadsheet (.xls) file, and is used for managing the collection, as well as keeping track of some circulation details, such as renewals. This data will be examined further in Section 3.2.2, while a screenshot of the spreadsheet file can be seen in Figure 2.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|----|----|-----------------------------------|---------------|-----|---|----------------|----------------------------|-----------|-------------------|-----------|-------------|--------|----------|----------|-----------|-----------|---------|
| 1 | ID | Title | JapaneseTitle | Vol | Author | JapaneseAuthor | Translator | Publisher | JapanesePublisher | Published | Donated by | Rating | Language | Due Date | Loaned to | Notes | Renewal |
| 2 | 0 | | | | | | | | | | | | | | | | |
| 3 | 1 | 5031 The webcomic - Now | | 1 | Owen Heitmann | | | | | 2004 | | Red | English | | | Doujinshi | |
| 4 | 2 | Storytale | | | Julianne Ting | | | | | | | Yellow | English | | | Doujinshi | |
| 5 | 3 | 5031 The webcomic - Now | | 2 | Owen Heitmann | | | | | 2004 | | Red | English | | | Doujinshi | |
| 6 | 4 | The History of Oztaku | | | Avi Bernshaw/b/Kenneth Chan | | | | | 2005 | | Red | English | | | Doujinshi | |
| 7 | 5 | The Seven Swords | | | Scott Beattie | | | | | 2006 | | Yellow | English | | | Doujinshi | |
| 8 | 6 | 9V | | | Michael Li | | | | | 2004 | | Yellow | English | | | Doujinshi | |
| 9 | 7 | Once Upon A Time | | | various | | | | | 2004 | | Yellow | English | | | Doujinshi | |
| 10 | 8 | Multi/Maniax | | | DATE | | Manga Translation Workshop | | | 2006 | | Yellow | English | | | Doujinshi | |
| 11 | 9 | 5031 The webcomic - Now | | 3 | Owen Heitmann | | | | | 2006 | | Red | English | | | Doujinshi | |
| 12 | 10 | The Boy, the Vampire and the Bear | | | Jing/Sarah Milne | | | | | 2005 | | Yellow | English | | | Doujinshi | |
| 13 | 11 | The Princess and the Gypsy | | | Jing/Sarah Milne | | | | | | | Yellow | English | | | Doujinshi | |
| 14 | 12 | I & I | | | Avi Bernshaw/Kenneth Chan | | | | | 2004 | | Yellow | English | | | Doujinshi | |
| 15 | 13 | Nakedfella Comics | | 5 | David Blumenstein | | | | | 2001 | | Yellow | English | | | Doujinshi | |
| 16 | 14 | More Secret Pictures of You! | | | Ben Bullock | | | | | | | Black | English | | | Doujinshi | |
| 17 | 15 | ANINDO punch | | 1 | Ed. Psycho Ann | | | | | 2004 | | Yellow | English | | | Doujinshi | |
| 18 | 16 | Suburban Knights 1 Squillion | | | Jack Kirby | | | | | | | Yellow | English | | | Doujinshi | |
| 19 | 17 | Nothing Personal | | | Kenneth Chan | | | | | 2004 | | Yellow | English | | | Doujinshi | |
| 20 | 18 | Totems | | 2 | Brian Gollnick/Wendy Leong/Tiffany Johnson/Chris Hayes-Kossmann | | | | | | Mishka Kent | Yellow | English | | | Doujinshi | |
| 21 | 19 | Nakedfella Comics | | 6 | David Blumenstein | | | | | 2003 | | Yellow | English | | | Doujinshi | |
| 22 | 20 | Ordinary Eyeball | | | Mandy Ord | | | | | 2006 | | Yellow | English | | | Doujinshi | |
| 23 | 21 | Nez Shouls: School of the Damned | | | Kai Lynk | | | | | 2007 | | Yellow | English | | | Doujinshi | |
| 24 | 22 | Gay Gay Kill Kill | | | Mattbonzo | | | | | 2005 | | Red | English | | | Doujinshi | |
| 25 | 23 | Vienna | | 1 | Jon Chung/Monin Sak/Vito Leo | | | | | 2006 | | Yellow | English | | | Doujinshi | |
| 26 | 24 | Vienna | | 2 | Jon Chung/Monin Sak/Vito Leo | | | | | 2006 | | Yellow | English | | | Doujinshi | |
| 27 | 25 | Vienna | | 3 | Jon Chung/Monin Sak/Vito Leo | | | | | 2007 | | Yellow | English | | | Doujinshi | |
| 28 | 26 | Azerath | | 1 | Daniel Lawson/Ryan Wilton | | | | | 2005 | | Yellow | English | | | Doujinshi | |
| 29 | 27 | Azerath | | 2 | Daniel Lawson/Ryan Wilton | | | | | 2005 | | Yellow | English | | | Doujinshi | |
| 30 | 28 | Azerath | | 3 | Daniel Lawson/Ryan Wilton | | | | | 2005 | | Yellow | English | | | Doujinshi | |
| 31 | 29 | Azerath | | 4 | Daniel Lawson/Ryan Wilton | | | | | 2005 | | Yellow | English | | | Doujinshi | |
| 32 | 30 | Azerath | | 5 | Daniel Lawson/Ryan Wilton | | | | | 2005 | | Yellow | English | | | Doujinshi | |

Figure 2: Screenshot of Monash's database file

3.1.3 Data from US Academic Libraries

While the majority of later data modeling was done using Toppan and Monash's data, initial research focused on examining data from various US academic libraries. Bibliographic data for manga from these sources was gathered manually on a per-record basis. This data was either the human-readable catalogue record pages, or the MARC data access through these catalogue pages. The lack of a large quantity of readily available, easily gathered data from these libraries meant that data analysis to the level of Monash and Toppan was not performed, but libraries remained an important focus throughout the thesis for two main reasons. First, the importance of BIBFRAME in this study, which was developed for and aimed mainly at libraries, meant that they would remain a relevant party regardless of the amount of library data used. Second, as perhaps the most important institutions with regards to bibliographic description in general, libraries are relevant to the overall scope of this thesis. As a goal of this thesis is to better the quality of descriptive data for manga, including the betterment of manga descriptions inside library records, library data collection was nonetheless important in understanding how bibliographic data for manga was currently handled by these institutions.

Library data came from Cornell, Duke, Illinois, Ohio State, and Stanford Universities. While relatively similar, there were some differences in their data, which will be discussed more in depth in Section 3.2.3. Figure 3 and Figure 4 below show an example of a library catalogue record and MARC record display, respectively, for manga.

BLEACH = Burīchi

BLEACH = ブリーチ

AUTHOR/CREATOR Kubo, Tite.
久保 帯人.

LANGUAGE Japanese. In Japanese.

IMPRINT Tōkyō : Shūeisha, 2002-
東京 : 集英社, 2002-

PHYSICAL DESCRIPTION v. : chiefly ill. ; 19 cm.

SERIES Jump comics.

Bibliographic information

BEGINNING DATE
2002

RESPONSIBILITY
Kubo Taito.
BLEACH = ブリーチ / 久保 帯人.

TITLE VARIATION
Burīchi
ブリーチ

SERIES
Jump Comics.

NOTE
Graphic fiction.

LOCAL SUBJECT
Stanford Manga Collection.

ISBN
4088732138

Japanese Collection
Library has: v.1-8,15,17-19,21-24,26
✓ PN6790 .J34 B54 2002 V.1
✓ PN6790 .J34 B54 2002 V.2
✓ PN6790 .J34 B54 2002 V.3
✓ PN6790 .J34 B54 2002 V.4
[show all](#)

More options
[Find it at other libraries via WorldCat](#)

Figure 3: Record display for Bleach from Stanford University library's catalogue

Librarian View

| | |
|--------|---|
| LEADER | 03359cam a22005651a 4500 |
| 001 | a8218223 |
| 003 | SIRSI |
| 005 | 20150321050002.0 |
| 008 | 090807m20029999ja a 000 c jpn d |
| 020 | a 4088732138 (v.1) |
| 020 | a 9784088732138 (v.1) |
| 040 | a EMU c STF d EMU d STF |
| 049 | a STFA |
| 079 | a ocm62470604 |
| 090 | a PN6790.J34 b B54 2002 |
| 100 | 1 6 880-01 a Kubo, Tite. = ^A2337768 |
| 245 | 1 0 6 880-02 a BLEACH = b Burīchi / c Kubo Taito. |
| 246 | 3 1 6 880-03 a Burīchi |
| 260 | 6 880-04 a Tōkyō : b Shūeisha, c 2002- |
| 300 | a v. : b chiefly ill. ; c 19 cm. |
| 490 | 1 a Jump Comics. |
| 546 | a In Japanese. |
| 500 | a Graphic fiction. |
| 830 | 0 a Jump comics. = ^A2300678 |
| 994 | a Z0 b STF |
| 690 | a Stanford Manga Collection. |
| 596 | a 26 |
| 035 | a (OCoLC-M)62470604 |
| 035 | a (OCoLC-I)429693529 |
| 880 | 1 6 100-01 a 久保 帯人. |
| 880 | 1 0 6 245-02 a BLEACH = b ブリーチ / c 久保 帯人. |
| 880 | 3 1 6 246-03 a ブリーチ |

Figure 4: MARC data for Stanford's "Bleach" record

3.1.4 Data from the Web

While Web data was not gathered in large amounts, it was examined to determine what type of manga data was available. This examination proved useful, as it was important to not only discover what the situation of manga metadata on the Web was like, but also identify what, if any, kinds of bibliographic data for manga existed as Linked Data. As a result, differences in "professional" institutional and "hobbyist" Web data metadata could be established, as well as the locating of Linked Data for manga that other resources could utilize. As recognized by He, et al. (2013), Wikipedia contains a significant amount of manga metadata of varying levels of granularity. Through DBpedia, some of this is available as Linked Data. Unfortunately, the author found no other major source of manga bibliographic data available as Linked Data. Other manga Web data, such as the searchable database location at <https://www.mangaupdates.com/>, contains useful information on manga series' and their authors, though as the data has not been made available outside of HTML scraping, its collection was not a part of this thesis. Examination of Web data can be found in Section 3.2.4, with a DBpedia sample shown in Figure 5.

3.2 Property Identification & Content Analysis

This section will go into greater detail about the properties and contents of each of the previously mentioned data sources. Here, properties refer to specific aspects of a manga being described, for example title or author. The analysis of bibliographic records and their contents was a necessary preliminary step in order to discover what properties of a manga were being described, and by which institutions. This helped not only to find similarities and differences amongst different types of institutions, but was necessary as a foundation for modeling the data.

About: ONE PIECE
An Entity of Type : [Manga](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

『ONE PIECE』（ワンピース）は、尾田栄一郎による日本の少年漫画。および、これを原作としたテレビアニメ、アニメ映画、ゲームなどのメディアミックス作品。『週刊少年ジャンプ』（集英社）にて1997年34号から連載されている。略称は「OP」「ワンピ」。

| Property | Value |
|----------------------------------|--|
| dbpedia-owl:abstract | <ul style="list-style-type: none"> One Piece (ワンピース, Wan Pisu) is a Japanese manga series written and illustrated by Eiichiro Oda. It has been serialized in Weekly Shōnen Jump since August 4, 1997; the individual chapters are being published in tankōbon volumes by Shueisha, with the first released on December 24, 1997, and the 73rd volume released as of March 2014. One Piece follows the adventures of Monkey D. Luffy, a young boy whose body gains the properties of rubber after unintentionally eating a Devil Fruit, and his diverse crew of pirates, named the Straw Hat Pirates. Luffy explores the ocean in search of the world's ultimate treasure known as One Piece in order to become the next Pirate King. The chapters have been adapted into an original video animation (OVA) produced by Production I.G in 1998, and an anime series produced by Toei Animation, which began broadcasting in Japan in 1999. Since then, the still ongoing series has aired over 600 episodes. Additionally, Toei has developed eleven animated feature films, two OVA's, and five television specials. Several companies have developed various types of merchandising such as a trading card game, and a large number of video games. The manga series was licensed for an English language release in North America by Viz Media, in the United Kingdom by Gollancz Manga, and in Australia and New Zealand by Madman Entertainment. The anime series has been licensed by Funimation Entertainment for an English-language release in North America, although the series was originally licensed and distributed by 4Kids Entertainment. One Piece has received wide critical acclaim, primarily for its art, characterization, humor and story. Several volumes of the manga have broken publishing records, including highest initial print run of any book in Japan and the first book to sell over three million copies in Oricon history. As of 2013, the series had over 345 million volumes in circulation worldwide, making it the best-selling manga series in history. 『ONE PIECE』（ワンピース）は、尾田栄一郎による日本の少年漫画。および、これを原作としたテレビアニメ、アニメ映画、ゲームなどのメディアミックス作品。『週刊少年ジャンプ』（集英社）にて1997年34号から連載されている。略称は「OP」「ワンピ」。 |
| dbpedia-owl:author | dbpedia:Eiichiro_Oda |
| dbpedia-owl:firstPublicationDate | 1997-08-04 (xsd:date) |
| dbpedia-owl:magazine | dbpedia:Weekly_Shōnen_Jump |
| dbpedia-owl:numberOfVolumes | 73 (xsd:integer) |
| dbpedia-owl:publisher | dbpedia:Shueisha |
| dbpedia-owl:thumbnail | http://commons.wikimedia.org/wiki/Special:FilePath/Onepiece-weit_(2).png?width=300 |
| dbpedia-owl:type | dbpedia:Manga |
| dbpedia-owl:wikiPageExternalLink | <ul style="list-style-type: none"> http://one-piece.com/ http://onepiece.viz.com/ http://onepieceofficial.com/ http://www.j-onepiece.com/ http://www.onepiece.com.au/grandline/ http://www.toei-anim.co.jp/tv/onep/ |

Figure 5: Screenshot of DBpedia's English page for the manga "One Piece"

3.2.1 Toppan Data Analysis

As mentioned, Toppan's data is in a tabular, spreadsheet format. The main data file worked with was the 単行本, or Tankoubon file, which contained approximately 85,000 records. The properties in this tabular data exist as the header values in the spreadsheet, all of which are in Japanese. A translation of the values can be found in a table below in Figure 6. All properties will be included, with the exception of one set of two repeated properties, マンガ単行本 ID and マンガ単行本全巻情報 ID, which are of relevance to Toppan holdings information and hold no bibliographic data of relevance. The table features a number of headers that

appear similar, but end with the Japanese ヨミ (yomi). This refers to the “reading” of the Japanese characters, and is usually given in the Japanese syllabary, Katakana. As the adopted Chinese characters, or Kanji, may have many different readings, a “yomi” is sometimes given to disambiguate a word from its other possible readings. The translation of this will be given simply as “reading” in Table 1.

Table 1: Toppan data properties in Japanese and English

| Original Japanese Header Name | English Translation |
|--------------------------------------|---|
| マンガ単行本名 | Manga Volume Title |
| マンガ単行本名ヨミ | Manga Volume Title, Reading |
| マンガ単行本名追記 | Manga Volume Title Supplement |
| マンガ単行本名追記ヨミ | Manga Volume Title Supplement Reading |
| マンガ単行本別版表示 | Manga Volume Alternate Edition Title |
| 巻 | Volume |
| 巻ソート | Volume Order |
| 責任表示 | Statement of Responsibility |
| 作者・著者 | Author Name |
| 作者・著者ヨミ | Author Name Reading |
| 原作・原案 | Original Story Author |
| 原作・原案ヨミ | Original Story Author Reading |
| 協力者 | Collaborator |
| 協力者ヨミ | Collaborator Reading |
| 標目 | Author Access Point / Entry Heading |
| 著者典拠 ID | Author Authority ID |
| 初版発行年(西暦) | Year First Issued |
| 初版発行月 | Month First Issued |
| 初版発行日 | Day First Issued |
| 初版価格 | Price First Issued |
| 単行本レーベル (サブレーベル) | Publisher Label First Issued In |
| 単行本レーベルヨミ | Publisher Label First Issued In Reading |
| レーベル番号 | Label Volume ID |

| | |
|----------------|---------------------------------------|
| レーベル典拠 ID | Label Authority ID |
| シリーズ | Series |
| シリーズヨミ | Series Reading |
| 出版者名 | Publisher |
| 出版者典拠 ID | Publisher Authority ID |
| 出版地 | Place of Publication |
| ページ数 | Number of Pages |
| 縦の長さ_横の長さ | Physical Dimensions |
| ISBN | ISBN |
| 全国書誌番号 | Japanese National Bibliography Number |
| 言語区分 | Language |
| 分類 | Nippon Decimal Classification (NDC) |
| レーティング | Rating |
| マンガ単行本紹介文 | Manga Abstract / Introduction |
| マンガ単行本タグ | Manga Tag |
| マンガ単行本備考 | Manga Note |
| 画像 1 | Image 1 |
| 画像 1 の表示フラグ | Image Display Flag 1 |
| 画像 2 | Image 2 |
| 画像 2 の表示フラグ | Image Display Flag 2 |
| 画像 3 | Image 3 |
| 画像 3 の表示フラグ | Image Display Flag 3 |
| メモ | Memo |
| マンガ単行本 ID | Manga Volume ID |
| マンガ単行本所蔵情報 ID | Manga Volume Holdings ID |
| 登録番号 (館固有の ID) | Building-Specific Registration ID |
| 版数 | Version / Edition Number |
| 刷数 | Number (of copies) Printed |
| 発行年(西暦) | Publication Year |
| 発行月 | Publication Month |

| | |
|-----------------|---------------------------------|
| 発行日 | Publication Day |
| 価格 | Price |
| 判型 | Format |
| 館独自の備考 | Original Record Note |
| 所蔵情報テーブルの非表示フラグ | Holdings Information Table Flag |

As shown in Table 1, Toppan's data contains 57 separate headers, or properties (minus the two aforementioned omitted) about individual volumes of manga. At 57 headers, Toppan's data appeared to be describing several times more properties than the other sources. Much of the data, however, appeared to be unfilled. After some inquiry, it was found that the Toppan database files were initially created with the help of librarians, who suggested properties to use, but some were unfilled in the process of data entry. To determine how many properties were blank, OpenRefine was used. OpenRefine has the ability to create facets on tabular data, which summarizes the information contained in all of the rows of a specific column. This allows one to do things such as find duplicate row values or calculate the number of blank rows. The facet menu inside OpenRefine can be seen in Figure 6, with the result of two facet operations shown in Figure 7.

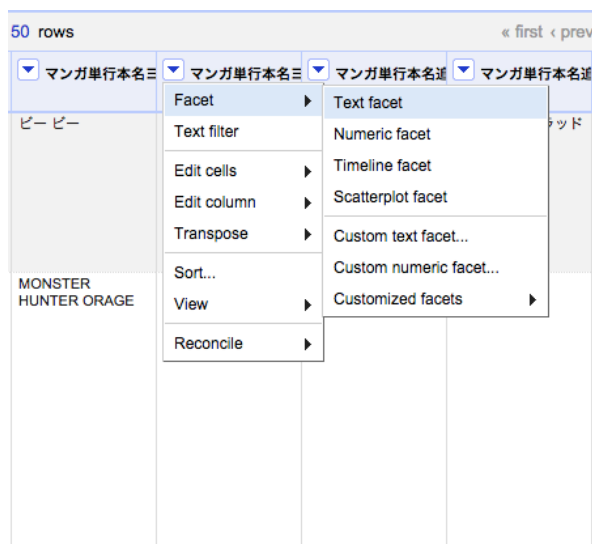


Figure 6: The facet menu in OpenRefine



Figure 7: Results from two facet operations

In the facet results of Figure 7, two facets, both of which are able to reveal the numbers of total blanks, were run on the same column. The top half shows the “Text facet” function, which produces a list of all row values in the selected column, along with the “count”, or the amount of times they appear in unique rows. At the bottom of this list, the value “(blank) 811” can be seen, showing that 811 rows – in this sample file containing 999 rows – were blank. In the bottom half of Figure 8, the custom facet function, “Facet by blank” was run. The “false” and “true” values given show how many are blank or filled, and again give the total counts. This blank faceting was performed using all of the aforementioned Toppan properties on a database file containing 85,249 rows in order to find which properties contained all or mostly blank data, and therefore were not vital in describing or modeling the data. The results are listed below in Table 2. The “Data Entry Status” column will have values of Sufficient, meaning the rows are mostly or totally filled, No Data, meaning the rows are completely blank or contain an irrelevant 0 value, or Insufficient, meaning the majority (typically 80,000+ of the 85,249 rows) are blank.

Table 2: Levels of filled properties in a Toppan data file

| Original Japanese Header Name | English Translation | Data Entry Status |
|--|--|--------------------------|
| マンガ単行本名 | Manga Volume Title | Sufficient |
| マンガ単行本名ヨミ | Manga Volume Title, Reading | Sufficient |
| マンガ単行本名追記 | Manga Volume Title Supplement | Sufficient |
| マンガ単行本名追記ヨミ | Manga Volume Title Supplement Reading | Sufficient |
| マンガ単行本別版表示 | Manga Volume Alternate Edition Title | Insufficient |
| 巻 | Volume | Sufficient |
| 巻ソート | Volume Order | Sufficient |
| 責任表示 | Statement of Responsibility | Sufficient |
| 作者・著者 | Author Name | Insufficient |
| 作者・著者ヨミ | Author Name Reading | Insufficient |

| | | |
|------------------|---|--------------|
| 原作・原案 | Original Story Author | Insufficient |
| 原作・原案ヨミ | Original Story Author Reading | Insufficient |
| 協力者 | Collaborator | Insufficient |
| 協力者ヨミ | Collaborator Reading | Insufficient |
| 標目 | Author Access Point / Entry Heading | Sufficient |
| 著者典拠 ID | Author Authority ID | Sufficient |
| 初版発行年(西暦) | Year First Issued | Sufficient |
| 初版発行月 | Month First Issued | Sufficient |
| 初版発行日 | Day First Issued | Sufficient |
| 初版価格 | Price First Issued | Insufficient |
| 単行本レーベル (サブレーベル) | Publisher Label First Issued In | Sufficient |
| 単行本レーベルヨミ | Publisher Label First Issued In Reading | Sufficient |
| レーベル番号 | Label Volume ID | Sufficient |
| レーベル典拠 ID | Label Authority ID | Sufficient |
| シリーズ | Series | Insufficient |
| シリーズヨミ | Series Reading | Insufficient |
| 出版者名 | Publisher | Sufficient |
| 出版者典拠 ID | Publisher Authority ID | Sufficient |
| 出版地 | Place of Publication | Sufficient |
| ページ数 | Number of Pages | Sufficient |
| 縦の長さ_横の長さ | Physical Dimensions | Sufficient |
| ISBN | ISBN | Sufficient |
| 全国書誌番号 | Japanese National Bibliography Number | Sufficient |
| 言語区分 | Language | Sufficient |
| 分類 | Nippon Decimal Classification (NDC) | Sufficient |

| | | |
|---------------------|--------------------------------------|--------------|
| レイティング | Rating | Insufficient |
| マンガ単行本紹介文 | Manga Abstract / Introduction | Insufficient |
| マンガ単行本タグ | Manga Tag | No Data |
| マンガ単行本備考 | Manga Note | Insufficient |
| 画像 1 | Image 1 | No Data |
| 画像 1 の表示フラグ | Image Display Flag 1 | No Data |
| 画像 2 | Image 2 | No Data |
| 画像 2 の表示フラグ | Image Display Flag 2 | No Data |
| 画像 3 | Image 3 | No Data |
| 画像 3 の表示フラグ | Image Display Flag 3 | No Data |
| メモ | Memo | Sufficient |
| マンガ単行本 ID | Manga Volume ID | No Data |
| マンガ単行本所蔵情報 ID | Manga Volume Holdings ID | No Data |
| 登録番号（館固有の ID） | Building-Specific Registration ID | No Data |
| 版数 | Version / Edition Number | No Data |
| 刷数 | Number (of copies) Printed | No Data |
| 発行年(西暦) | Publication Year | No Data |
| 発行月 | Publication Month | No Data |
| 発行日 | Publication Day | No Data |
| 価格 | Price | No Data |
| 判型 | Format | No Data |
| 館独自の備考 | Original Record Note | No Data |
| 所蔵情報テーブルの非表示フ ラグ | Holdings Information Table Flag | No Data |

Through the faceting done to produce the table in Table 2, it can be shown that 26 of the total properties contain sufficient levels of data. It is worth noting that while it may appear

strange for the “Author Name” field to contain mostly blanks, this is because the author name data, along with their role, generally appears in the “Author Access Point / Entry Heading” column instead.

Analyzing the properties and contents of Toppan’s data held reveal what was being described and to what level of granularity. The majority of sufficient level data describes fairly standard bibliographic information, such as titles, author, publication information, etc. This means that when attempting to develop a model for Toppan’s data, a model made for typical library data may also be suitable in this case. One of the more unique pieces of information found in Toppan’s data is the role that is attached to the “Statement of Responsibility” column. The data in these rows contain one or more names, with roles typically attached to each. For example, one value is listed as [著]秋本治, where the 著 kanji stands for the story’s author, 秋本治 (Akimoto, Osamu). While the majority of the row values are single names with an author role, occasionally multiple names are listed with roles such as editor or original story author, which is of interest, as role data outside of author is typically not included from other sources.

As the file holds records for individual volumes of manga, much of the data, such as ISBN, physical properties, and publisher imprint information, can be said to describe manga at the FRBR Manifestation or Item level. As mentioned in Section 1.3, a goal of the aggregation model is to serve multiple FRBR “levels” of bibliographic description; therefore Toppan can be seen as a possible provider of information that is closer to the FRBR Item side.

Importantly, the vast majority of information in Toppan’s data is in Japanese. This means that when aggregating variously sourced manga data, Toppan’s will serve specific functions, at least when the main focus of aggregation is aimed at English data. First, when aggregating English and Japanese data for a manga, the data can be used to provide the original Japanese for titles, author names, publishers, etc., for Japanese language manga that has been catalogued in English, as is often the case. This original Japanese data not only provides additional access points for users, but also contributes to the minutiae of data that fans of the medium are interested in. The second function it may serve is that the data may be used to describe a Japanese manga that is connected to its respective English translation, which is described using English data providers, assuming the different language versions can be identified and connected in a model.

In summary, it can be said that Toppan’s data describes data near the FRBR Item level, contains fairly standard bibliographic description properties that may function well with a library-based data model and vocabulary, and may serve specific important functions during aggregation due to the contents being in Japanese.

3.2.2 Monash Data Analysis

Monash's data, like Toppan's, is a spreadsheet file with no "formal" data model attached. The data properties are simply column headers in the spreadsheet file, with their contents being the row values. The file contains 7349 data rows, and like Toppan, describes individual volumes of manga.

Monash's data has a total of 17 properties. Some of these, however, are only relevant to Monash themselves and do not contain bibliographic data. These columns are: ID, Donated By, Due Date, Loaned To, Notes, and Renewal. The remaining columns, which contain relevant bibliographic data, are: Title, Japanese Title, Volume, Author, Japanese Author, Translator, Publisher, Japanese Publisher, Published In, Rating, and Language. Rating is arguable not relevant to bibliographic description, as it is a subjective content rating system unique to Monash, but it may be of use in identifying obvious adult rated content. Also, it is important to note that with the "Japanese Title," "Japanese Author," and "Japanese Publisher," these are generally the names given in the original Japanese, whereas the "Title," "Author," and "Publisher" are Monash's translations. In other words, the "Japanese Publisher" property, for example, will give the Publisher name in Japanese, but only if the work is in Japanese. If the work is an English translation published by, for example, Viz Media, an English language publisher, the Japanese Publisher column will typically not contain the original Japanese Publisher and be left blank.

Like with Toppan, OpenRefine was utilized in some early work with Monash's data. Aside from some minor statistics gathering, OpenRefine was used to separate data found in the Japanese Title column. For the Japanese works in Monash's database, the Japanese Title column gives the title in its original Japanese syllabary, along with the English transliteration of the title in brackets. For example, for the records about the manga Dragon Quest, the Japanese Title column reads "ドラゴンクエスト (doragon kuesuto)." This transliteration information is not usually important, but the separation of the Japanese from its transliteration is useful for two reasons. First, the isolation of the Japanese language title makes working with the data easier, e.g. matching the values to Toppan's various Title columns, none of which would contain an English language transliteration. Secondly, while the transliteration is generally not useful in aggregating data, it can occasionally be used to match data. Referring to Figure 3, one can see that Stanford has included the transliteration of "Burīchi" in a title property. Rarely, then, the transliterations may be of use to match Monash data to another data source that has chosen to record the same data. The separation process in OpenRefine can be seen in the figures below. Figure 8 shows the sample Monash data in its original form,

Figure 9 shows use of the “Split column” function, and Figure 10 shows the final result. Not that the “Split column” function will remove the separating open parenthesis, but keep the close parenthesis. This was removed through the use of OpenRefine’s Google Refine Express Language (GREL) expression `chomp(value, " ")`.

Show as: **rows** records

Show: 5 10 25 50 rows


| ▼ All | ▼ ID | ▼ Title | ▼ JapaneseTitle | ▼ Vol | ▼ Author |
|--|------|--------------------|----------------------------|-------|--------------|
| ☆  | 1. | 10931 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 1 | Sanjou, Riku |
| ☆  | 2. | 10932 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 2 | Sanjou, Riku |
| ☆  | 3. | 10935 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 3 | Sanjou, Riku |
| ☆  | 4. | 10937 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 4 | Sanjou, Riku |
| ☆  | 5. | 10939 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 5 | Sanjou, Riku |
| ☆  | 6. | 10940 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 6 | Sanjou, Riku |
| ☆  | 7. | 10942 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 7 | Sanjou, Riku |
| ☆  | 8. | 10943 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 8 | Sanjou, Riku |
| ☆  | 9. | 10944 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 9 | Sanjou, Riku |
| ☆  | 10. | 10947 Dragon Quest | ドラゴンクエスト (doragon kuesuto) | 10 | Sanjou, Riku |

Figure 8: Original Monash JapaneseTitle column

Split column JapaneseTitle into several columns

How to Split Column

☒ by separator

Separator ☐ regular expression

Split into columns at most (leave blank for no limit)

☐ by field lengths

List of integers separated by commas, e.g., 5, 7, 15

After Splitting

☒ Guess cell type

☒ Remove this column

OK Cancel

Figure 9: Split column function inside OpenRefine

Show as: **rows** records

Show: 5 10 25 50 rows





















| <input type="checkbox"/> All | <input type="checkbox"/> ID | <input type="checkbox"/> Title | <input type="checkbox"/> JapaneseTitle 1 | <input type="checkbox"/> Japanese Title R | <input type="checkbox"/> Vol | <input type="checkbox"/> Author |
|---|-----------------------------|--------------------------------|--|---|------------------------------|---------------------------------|
|   1. | 10931 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 1 | Sanjou, Riku |
|   2. | 10932 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 2 | Sanjou, Riku |
|   3. | 10935 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 3 | Sanjou, Riku |
|   4. | 10937 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 4 | Sanjou, Riku |
|   5. | 10939 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 5 | Sanjou, Riku |
|   6. | 10940 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 6 | Sanjou, Riku |
|   7. | 10942 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 7 | Sanjou, Riku |
|   8. | 10943 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 8 | Sanjou, Riku |
|   9. | 10944 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 9 | Sanjou, Riku |
|   10. | 10947 | Dragon Quest | ドラゴンクエスト | doragon kuesuto | 10 | Sanjou, Riku |

Figure 10: Data separated into the Japanese title and its reading

Though not addressed in this thesis outside of a recommendation in Section 5.3, there is also the future possibility of a manga-focused description ontology that may feature a property allowing for the recording of a transliteration, so that non-Japanese users may read the Japanese title.

Monash data, then, is similar to Toppan's. It describes standard bibliographic descriptive properties, and may be suitable to share a description model with both Toppan and other institutions like libraries. The data is not as granular as Toppan's, though it is still describing manga at the volume, or FRBR Item level, because of the inclusion of the "Volume" property. Unlike Toppan's data, however, there are not many properties that are bound to the FRBR Item level, such as physical properties or specific Volume titles, so there is possibly room to use Monash's data at the FRBR Work level for properties such as author or series title. The combination of English and Japanese information may prove useful in identifying related works open to aggregation, which will be explored in Section 4.3.

3.2.3 US Academic Library Analysis

Though being analyzed in this thesis after Toppan and Monash, US academic library data was the first to be examined during the author's study of this topic. Aside from libraries being the most obvious source of bibliographic data – prior to seeking collection of Toppan

and Monash data – library data was important to examine because of the interest in using the BIBFRAME model, created mainly for library data, as a way to describe manga.

As mentioned previously, data from libraries was gathered manually on an individual record basis and came in the form of either record display webpages from an online catalogue, or MARC data accessed through said webpage. While this meant that a large amount of data was not gathered, a positive was that in this early phase, data several different libraries could be examined. Bibliographic records from Cornell, Duke, Illinois, Ohio State, and Stanford universities were examined to find what bibliographic description properties were common throughout all the institutions. These properties are as follows: English title(s), Japanese title(s), Author name (in English), Author name (in Japanese), Collaborators (e.g. translators), Publisher (in English), Publisher (in Japanese), Date first published, Series, Subject, and Language.

Records for both English and Japanese items were examined, and while the level of granularity described was consistent, most of the records translated manga in English left out any Japanese language information. The records for original Japanese items, however, feature English translations of titles, authors, and publication information, while putting the original Japanese information in MARC field 880, which records alternate script information (see <http://www.loc.gov/marc/authority/ad880.html>). This makes the Japanese language records in these libraries similar to Monash's data in their recording of both English and Japanese data, at least for the records describing Japanese materials.

As mentioned in Section 2.2 and by Fee (2008), serialized manga may be catalogued in different ways. A single serial record may be created, which contains all of the holdings information on a series in a single catalogue record. This is the practice for all of the aforementioned libraries in this section. While this is easier on cataloguing staff and may be a sufficient description level depending on patron type, it also limits the amount of information that a record can contain. Stanford's record for Bleach simply has individual volumes in their holdings as added MARC field 999 entries. Similarly, Cornell's Dragon Ball Z record is for the entire series, with individual volume ISBNs having separate MARC 020 entries. If any deeper level of granularity is to be had, such as volume summaries / abstracts, storylines, changing artists, etc., then analytic cataloguing would have to replace the current serial records. As contact with library staff has revealed, staff time and workload may be the main obstacle in pursuing this, so making the data available via Linked Data sources would be useful for those institutions interested.

The final portion of analyzing the contents of library records was focused on Linked Data. As the use of BIBFRAME was being pursued, it was of interest to see what portions of

manga library records would be suitable, model allowing, for Linked Data. A visual representation of the information commonly found in library manga records was made, shown in Figure 11, with properties that have a possible relevant Linked Data source being labeled as a “Linked Data Opportunity.”

This identification of these “Linked Data opportunities” proved interesting. It appears as though much of a single record can be linked to an external source for data. Many of these are simply links to existing authorities, such as linking creator information to LOC’s Name Authority or Virtual Internet Authority File (VIAF). More unique opportunities exist with other properties, such as obtaining summary or abstract information from Wikipedia/DBpedia, or cover art, reviews, etc., from various sources, then applying them using BIBFRAME’s Annotation entity. This step also revealed areas where manga-specific information is lacking inside library records. Adhering to the use of LCSH, the “Subject” field in manga records is typically useless. While even the use of a “manga” subject would be minimally sufficient, LCSH lacks this term, so manga are typically attached the subject of “Comic books, strips, etc. -- Japan -- Translations into English,” as seen in the Dragon Ball Z records at Ohio State and Cornell libraries. Though some libraries add local subject entries, these are usually the series name itself rather than a relevant subject. A local manga subject tag, however, may be possibly supplemented through Wikipedia’s “Genre” field, for another Linked Data opportunity. Though it is outside the scope of this thesis, it is clear that a Linked Data-capable manga authority resource, at the very least for subjects and genres, would be useful in providing more adequate descriptions for manga records.

In summary, libraries tend to catalogue manga at a fairly shallow level of bibliographic description. The lack of analytic cataloguing and any manga-specific features and descriptive terms mean the records are of little use outside of holdings information. This lack of analytic cataloguing also means that despite cataloguing items at the FRBR Item level, exemplified through the use of volume numbers, publisher information, volume ISBNs, etc., the fact that there is a single record for an entire manga series makes the record itself describe the manga at the FRBR Work level. Despite the shallow level of descriptive data, however, there many opportunities for these records to be supplemented with Linked Data, provided libraries can open up their traditional silos of closed data to the LOD space. This will be discussed further in the BIBFRAME discussion in Section 3.3.2.

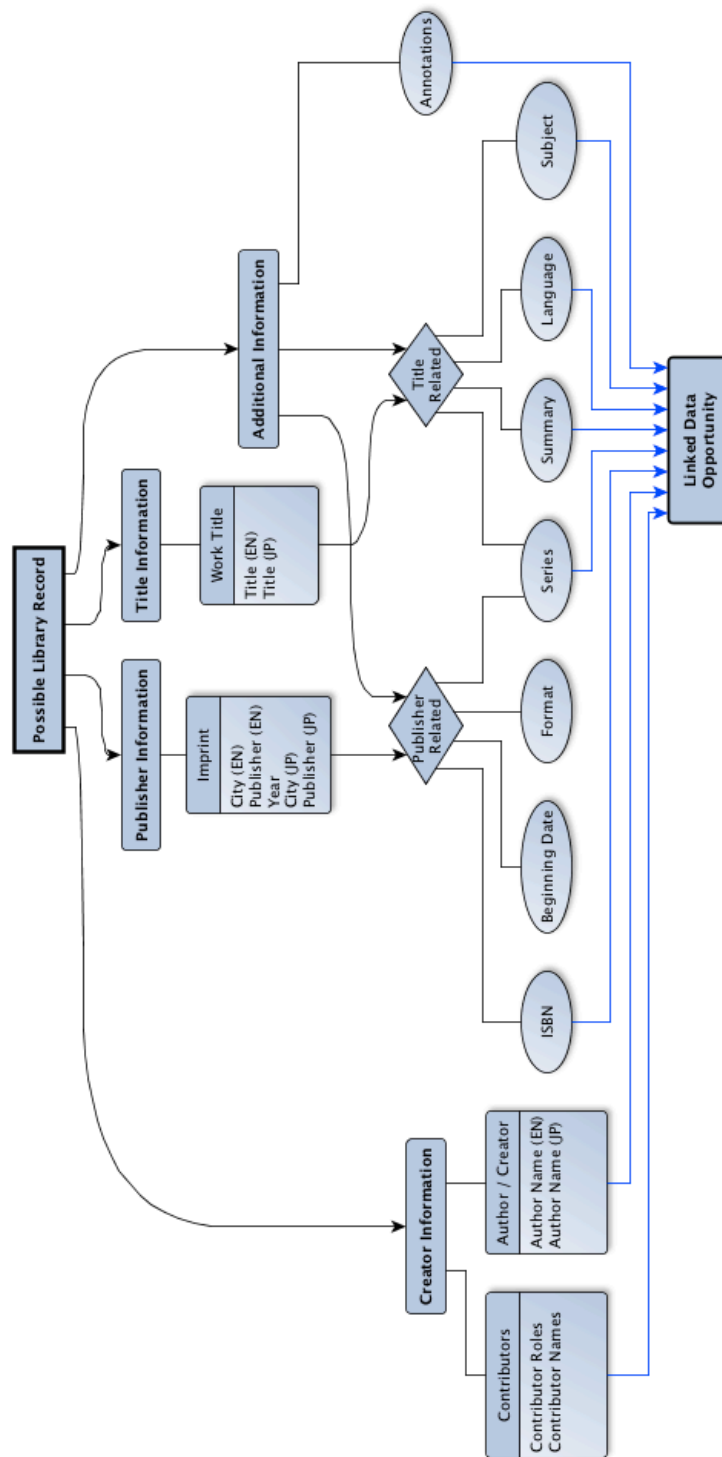


Figure 11: Possible Linked Data opportunities found inside library records for manga

3.2.4 Web Data Analysis

As Web data was not gathered in a significant amount, the analysis that took place was fairly shallow, and was performed mainly to see how granular the information was, particularly information available as Linked Data. This meant that the main source of Web data that was examined came from Wikipedia and DBpedia.

The Wikipedia information available for manga is the most granular of all the analyzed sources. Wikipedia entries for manga contain detailed information on setting, storylines, characters, genres, authors, illustrators, publishers, and other bibliographic data, with most of this data being available as Linked Data via DBpedia. DBpedia's main source of data comes from the sidebar of a Wikipedia page containing specific properties and their values. The more detailed textual information in the Wikipedia articles themselves are generally not available, though are in some cases. Nevertheless, the granularity of the "sidebar" data is typically deeper than the other English sources, and matches or exceeds Toppan's data. An example of DBpedia's information display can be seen in Figure 12.

| | |
|----------------------------|---|
| dbpprop:author | ▪ dbpedia: Eiichiro_Oda |
| dbpprop:caption | ▪ 0001-02-04 (xsd:date) |
| dbpprop:demographic | ▪ Shōnen |
| dbpprop:director | ▪ dbpedia: Gorō_Taniguchi ▪ dbpedia: Kōnosuke_Uda ▪ Munehisa Sakai ▪ Hiroaki Miyamoto ▪ Katsumi Tokoro ▪ Naoyuki Ito |
| dbpprop:episodeList | ▪ List of One Piece episodes |
| dbpprop:episodes | ▪ 641 (xsd:integer) |
| dbpprop:first | ▪ 1997-08-04 (xsd:date) ▪ 1999-10-20 (xsd:date) |
| dbpprop:genre | ▪ dbpedia: Adventure_fiction ▪ dbpedia: Action_(fiction) ▪ dbpedia: Comedy-drama |
| dbpprop:hasPhotoCollection | ▪ http://wifo5-03.informatik.uni-mannheim.de/flickrwrappr/photos/One_Piece |
| dbpprop:jaKanji | ▪ ONE PIECE (ワンピース) |
| dbpprop:jaRomaji | ▪ Wan Pīsu |
| dbpprop:magazine | ▪ dbpedia: Weekly_Shōnen_Jump |

Figure 12: A portion of DBpedia properties for the Wikipedia entry for "One Piece"

Another important feature of Web data is the varied FRBR levels of data that different articles describe. The main Wikipedia entry for One Piece, found at http://en.wikipedia.org/wiki/One_Piece, contains information not only about the manga, but the anime series, theatrical releases, video games, and so on. The article, therefore, acts as a FRBR Work entity, being home to all of the Expressions of the intellectual property. This works well for using the article's data to describe the manga at the FRBR Work level, but initially, it appeared less suitable for BIBFRAME. This is due to the fact that the BIBFRAME Work level is not as conceptually high as FRBRs – that is, while One Piece as a FRBR Work can represent the

manga, anime, and other mediums under a single umbrella, these would be separate BIBFRAME works with relationships established between them. After further examination, however, it was decided that the data for the Wikipedia articles could still be used at the more manga-specific BIBFRAME Work level, because Wikipedia views the manga expression as the central entity in the articles. The entries for One Piece, Bleach, Dragon Ball, et al., open with the phrase “is a Japanese manga series,” so despite housing information on other media formats, the bibliographic data is still applicable to the BIBFRAME Work level for manga.

While the main Wikipedia entry for a represents the Work level, related articles represent others. Continuing the One Piece example, the main Wikipedia entry links to the “List of *One Piece* manga volumes” entry as the “Main article” for the manga (see http://en.wikipedia.org/wiki/List_of_One_Piece_manga_volumes). This article contains data that is closer to the FRBR Manifestation and Item levels, such as specific volume titles in English and Japanese, release dates, and ISBNs. Further, collections of 20 volumes have their own “Main article” that describes an even deeper level of detail. The “List of *One Piece* manga volumes” entry, for example, features a link to another article for volumes 1-20, title “List of *One Piece* chapters (1–186)” (see [http://en.wikipedia.org/wiki/List_of_One_Piece_chapters_\(1-186\)](http://en.wikipedia.org/wiki/List_of_One_Piece_chapters_(1-186))). While these chapter pages do not contain a large amount of data, they contain information that is unique when compared to the other sources, such as chapter titles and volume summaries. An example of a volume summary from the aforementioned chapters page can be found in Figure 13.

| No. | Title | Japanese release | English release |
|--|--|--|--|
| 1 | <i>Romance Dawn</i> <i>Romance Dawn: Bōken no Yoake</i> (Romance Dawn —冒険の夜明け—) | December 24, 1997 ^[1] ISBN 4-08-872509-3 | June 30, 2003 ^[4] ISBN 1-56931-901-4 |
| <div> <div> 1. "Romance Dawn" (Romance Dawn —冒険の夜明け— "<i>Romance Dawn: Bōken no Yoake</i>"^(?)) 2. "They Call Him 'Straw Hat Luffy'" (その男「麦わらのルフィ」 "<i>Sono Otoko 'Mugiwara no Ruffi</i>"^(?)) 3. "Enter Zoro: Pirate Hunter" (「海賊狩りロロノア・ゾロ」登場 "<i>Kaizoku-Gari Roronoa Zoro' Tōjō</i>"^(?)) </div> <div> 4. "The Great Captain Morgan" (海軍大佐「斧手のモーガン」 "<i>Kaigun Taisa 'Onote no Mōgan</i>"^(?)) 5. "The King of the Pirates and the Master Swordsman" (海賊王と大剣豪 "<i>Kaizoku Ō to Daikengō</i>"^(?)) 6. "Number One" (一人目 "<i>Hitorime</i>"^(?)) 7. "Friends" (友達 "<i>Tomodachi</i>"^(?)) 8. "Nami" (ナミ登場 "<i>Nami Tōjō</i>"^(?)) </div> </div> <p>The seven-year-old boy Monkey D. Luffy tries to join the pirate crew of "Red-Haired" Shanks, but is rejected for being too young. Accidentally, he eats a devil fruit that causes his body to gain the properties of rubber but also makes him permanently unable to swim. After an ordeal with mountain bandits, Luffy gives up on joining Shanks' crew, and instead, vows to one day surpass Shanks, build up a crew of his own, and become the next King of the Pirates. Ten years later, Luffy sets out to sea, where he soon frees a young boy, Coby, from a slave's life in the pirate crew of Alvida, and saves the three-swords-wielding bounty hunter Roronoa Zoro from being executed by the Navy. With Zoro as Luffy's first crewman, they set sail for the Grand Line, the sea where the One Piece, the treasure of the last King of the Pirates, is supposedly hidden, and make the acquaintance of the thief and expert navigator Nami.</p> | | | |

Figure 13: Dual language chapter titles and a Volume 1 summary, from [http://en.wikipedia.org/wiki/List_of_One_Piece_chapters_\(1-186\)](http://en.wikipedia.org/wiki/List_of_One_Piece_chapters_(1-186))

While Wikipedia contains a lot of interesting information, unfortunately not all of it is accessible through DBpedia. Most of the general information available in the main Work level entry can be accessed, but much of the unique data, such as the volume summaries or chapter titles shown in Figure 13, are not in DBpedia in a way that allows for easy Linked Data access. While the data can still be access in other ways, such as HTML scraping, information such as chapter titles or volume summaries are not given their own accessible properties in DBpedia. Still, the information that is available makes DBpedia perhaps the most authoritative source for manga data on the Web, as shown by He et al. (2013), and while unique information such as volume summaries can not be accessed by standard Linked Data methods, the information is present and the possibility of its use by other institutions remains.

3.3 Methods of Bibliographic Description and Aggregation

This section will move away from specifics of the data properties and their contents, and onto general description methods, both those that are currently being used and those that this thesis will pursue. Section 3.3.1 will examine how the data from the sources in Section 3.1 are modeled, 3.3.2 will examine the BIBFRAME model and how it can be applied, and Section 3.3.3 will focus on EDM as a method of data aggregation.

3.3.1 Existing methods of Bibliographic Description

The existing methods for the bibliographic description of manga vary depending on the institution or other source in charge of the data. Bibliographic data from Toppan and Monash follow no formal bibliographic description model. They each employ their own number of descriptive properties, which form the column headers in their tabular data. On the other hand, the library data examined adheres to fairly strict cataloguing rules, and is modeled as a serial according to MARC, AACR2, and RDA guidelines. Each of these methods has benefits and drawbacks. Referring to Toppan's data in Figure 6, one can see that a lack of strict rules applied to an entire format, such as a monograph or serial, mean a variety of attributes can be described. The fact that manga is technically a serial and thus gets catalogued as one means that it, by default, must undergo analytic cataloguing to describe any granular detail deeper than the most basic of bibliographic descriptive properties. Manga, despite being a type of serial, is not well served by standard cataloguing rules for serials. That said, there are benefits to having a formal structure to the data and rules for cataloguing. Strict cataloguing rules for titles, for example, ensure levels of uniformity for bibliographic records. More relaxed rules for institutions mean that some may be tempted to catalogue titles under common nicknames that patrons will presumably search for, but this can lead to inconsistent data. Monash's data

features some examples of this. The records for the Japanese manga こちら葛飾区亀有公園前派出所 are catalogued under two different English names: the full transliteration, “Kochira Katsushikaku Kameari Kouenmae Hashutsujo,” and the shortened transliteration of the Japanese nickname, “Kochikame.” If one is unaware of this distinction, or does not compare the “JapaneseTitle” property contents for both, it may appear as though volumes for this manga are missing. The cataloguing of popular yet unofficial titles and other information is certainly useful for patron information and more access points, and is typically not done in traditional library records, but it may also lead to inconsistent records.

An important discovery made during the analysis of data in the previous sections was that despite having different levels of granularity, the data sources were typically describing the same properties of manga. This may simply be the nature of bibliographic description, but it is important to keep in mind, as it means there is an opportunity for different types of institutions to use the same model of bibliographic description, whether they are traditional libraries or corporate databases. Currently, however, this has not taken place, because institutions outside of libraries do not use the backbones of bibliographic description in libraries, such as MARC, AACR2, Z39.50, etc. The BIBFRAME initiative seeks to move away from some of these technologies, so it is worth investigating how suitable BIBFRAME would be for use outside of libraries for institutions that are doing bibliographic description in a way that is close to libraries. The next section will investigate this.

3.3.2 Bibliographic Description Using BIBFRAME

As library data was the earliest examined bibliographic data for manga used in this study, and because there was an initial interest in Linked Data for manga, BIBFRAME was an obvious model to investigate. This section will discuss some of the fundamentals of the BIBFRAME model and vocabulary. More information on BIBFRAME can be found at the BIBFRAME LOC homepage at <http://www.loc.gov/bibframe/>, along with the full vocabulary at <http://bibframe.org/vocab/>.

BIBFRAME is a new conceptual model and vocabulary for bibliographic description. Conceptually, BIBFRAME has some similarities to FRBR. These similarities include the ability to differentiate between intellectual content and its physical manifestations, and exposing relationships between entities (Library of Congress, 2015a). Figure 14 shows a visualized portion of the BIBFRAME model, showing off three of the four main entities, Work, Instance, and Authority. While the initiative is headed by Library of Congress and is aimed primarily at libraries, BIBFRAME’s goal of replacing MARC and the technologies behind it being standard

Web / Linked Data technologies mean there is an opportunity for other non-library institutions interested in bibliographic description to use BIBFRAME, or at least, use library data in a way they were not able to before.

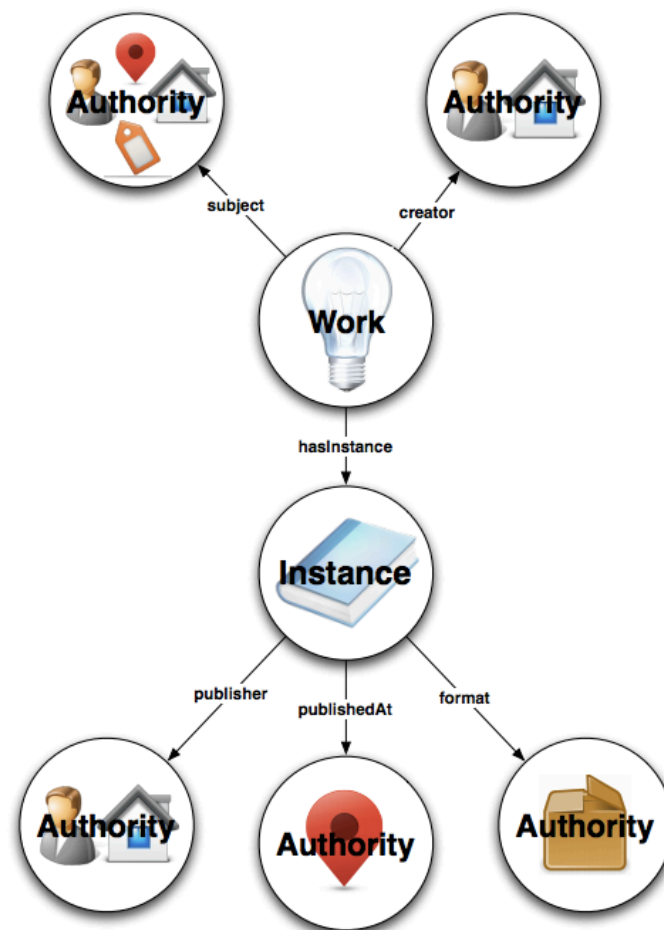


Figure 14: BIBFRAME model showing a Work to Instance relationship, and authority resources (Miller, et al. 2012)

The four main BIBFRAME entities are Work, Instance, Authority, and Annotation (Library of Congress, n.d.). Work, short for Creative Work, is a resource representing the conceptual essence of the material. While this is similar to FRBR's Work entity, there is a difference in how high the conceptual level can go. As mentioned in Section 3.2.4, the FRBR Work entity can be conceptual enough to handle different media formats, provided they share the same main intellectual property. The example given was that the FRBR Work for "One Piece" acts as a kind of "super work," representing the main conceptual content for the One Piece manga, anime, theatrical releases, and so on. In BIBFRAME however, the Work entity does not reach this "super work" level. Each media format would represent its own BIBFRAME

Work, with relationships being established between them, rather than all of the resources being under a single Work umbrella. This is quite different from FRBR, but emphasizes the importance of relationships between entities in BIBFRAME, with Work-to-Work relationships being one of the main types. Despite the differences, however, BIBFRAME and FRBR's Work entities are similar, and BIBFRAME's Work can be said to be a combination of FRBR's Work and Expression entities, as shown in Figure 15.

Instance is the second main BIBFRAME entity. Instance represents a material (not necessarily physical) embodiment of the Work to which it is an embodiment of. As Figure 15 illustrates, Instance can be said to be a combination of FRBR's Manifestation and Item entities. If the intellectual concept of the *One Piece* manga is represented by BIBFRAME's Work, then Instance would represent a specific publication of said Work. Work to Instance relationships are the other main relationship type in BIBFRAME. Figure 14 illustrates this relationship. The high level conceptual Work has information attached that is agnostic to the publication Instances, such as subject and creator, where the Instance has specific publication information attached. The attached properties are examples of the next BIBFRAME entity, Authority.

BIBFRAME Authorities are reusable resources that have specific roles associated with a BIBFRAME Work or Instance. While not designed to replace existing library authorities, BIBFRAME Authorities represent similar types of data, such as people, places, organizations, or topics. As they are not designed to replace authorities, they simply represent their data, an important Authority concept is the domain, or who is in charge of the authority information. The BIBFRAME Authorities draft specification (Library of Congress, 2014a) provides some examples of authority data in RDF, shown in Figure 16. Lines 7 and 8 here feature the properties `bf:authorizedAccessPoint` and `bf:hasAuthority`. The `bf:authorizedAccessPoint` property features a controlled vocabulary term for the name "Bartoluzzi, Bruno," and the `bf:hasAuthority` property identifies the authoritative origin of the data, which in this case is Library of Congress' Linked Data Name Authority File.

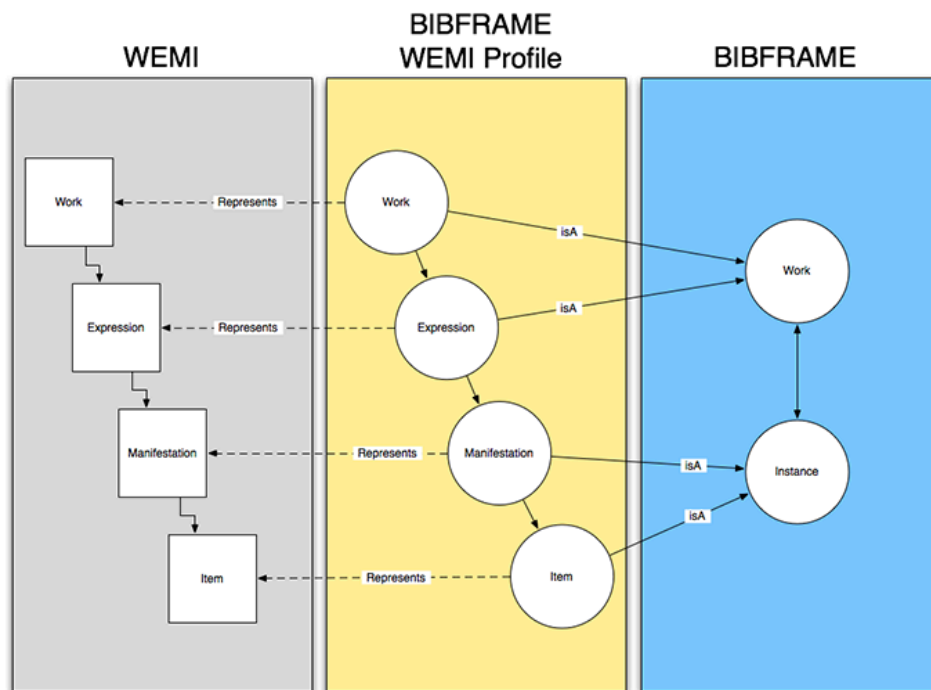


Figure 15: A mapping of FRBR entities to BIBFRAME's Work and Instance (Library of Congress, 2014b)

```

1 <bf:Work>
2   <bf:title>Collage</bf:title>
3   <bf:creator rdf:nodeID=" bnode-x24z100" />
4 </bf:Work>
5 <!-- BIBFRAME Authority -->
6 <bf:Person rdf:nodeID=" bnode-x24z100" >
7   <bf:authorizedAccessPoint>Bartolozzi, Bruno</bf:authorizedAccessPoint>
8   <bf:hasAuthority rdf:resource=" http://id.loc.gov/authorities/names/n80103954" />
9 </bf:Person>

```

Figure 16: An RDF example of BIBFRAME Authority properties (Library of Congress, 2014)

Authority data may also have multiple links to multiple authorities. In Figure 16, in addition to the `bf:hasAuthority` linking to the LOC Name Authority File, the property `bf:referenceAuthority` can be used to link to secondary authorities, such as VIAF. While the authority source is important, it is not a requirement in the BIBFRAME model or vocabulary. Nevertheless, the ability to make authoritative statements about reusable resources such as author names, subjects, or organizations, and link to external sources that claim the authoritativeness of this data, is an important feature of BIBFRAME's ability to link

library data with non-library sources – something that was not possible with MARC and key for Linked Data for libraries.

The final main entity in BIBFRAME is the Annotation. Annotations are meant to attach supplemental information to a resource. While the scope of information that can be used is currently unclear, the BIBFRAME vocabulary page for Annotation (see <http://bibframe.org/vocab/Annotation.html>) does feature some specific Annotation types, such as cover art, summaries, and reviews. Figure 17 features a visualizing of the Annotation entity.

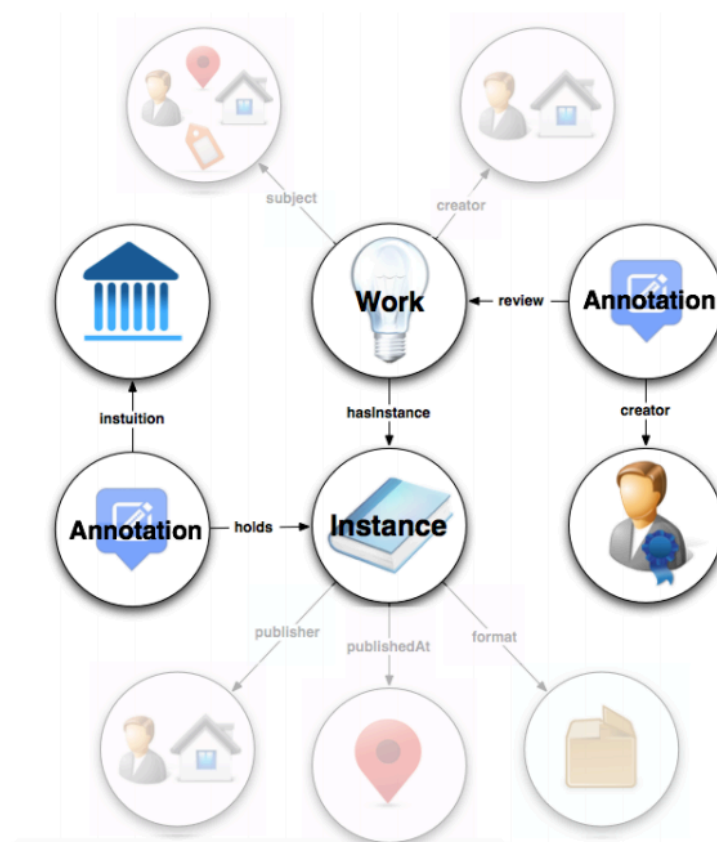


Figure 17: Holding information and a review attached to a BIBFRAME Instance and Work, respectively (Miller, et al. 2012)

While the possibility of supplementing a resource with external data seems endless, the current ambiguity about Annotations limits the usefulness of conjecture. The most common usable example for Annotations is to attach library holdings information to a resource, as shown in Figure 17.

As a conceptual model, BIBFRAME appears able to be utilized by non-library institutions. The focus on intellectual entities and their relationships, while using a vocabulary based on RDF, an open and popular Web standard, means that bibliographic description using BIBFRAME is not bound to library technologies or traditional cataloguing rules. The conceptual model and RDF vocabulary, combined with BIBFRAME Profiles (Library of congress, 2014b), which are syntactic constraints used to serve unique community's descriptive needs (i.e. a Profile paired with usage guidelines, etc., may be created by a community specifically for the bibliographic description of manga), mean there is room for BIBFRAME to be a way to model and describe bibliographic data for manga, both in and outside libraries, regardless of what FRBR entity the data aligns closest to.

3.3.3 Aggregating Data using EDM

As introduced in Section 2.1.3, EDM, or the Europeana Data Model, is the underlying model that serves the Europeana cultural heritage portal. While EDM is quite a complex data model meant to serve many functions, this section will focus on the core of the data model, as well specifics of aggregating data with EDM.

EDM consists of three core properties. `edm:ProvidedCHO` represents a “cultural heritage object,” such as painting, book, sculpture, and so on. `edm:WebResource` represents accessible digital representations of the cultural heritage object. Typically, this is a photo of the object in question. Finally, the `ore:Aggregation` property collects data from one source, including one or more WebResources, and aggregates the data for a single ProvidedCHO. A visualization of these three core classes can be seen in Figure 18.

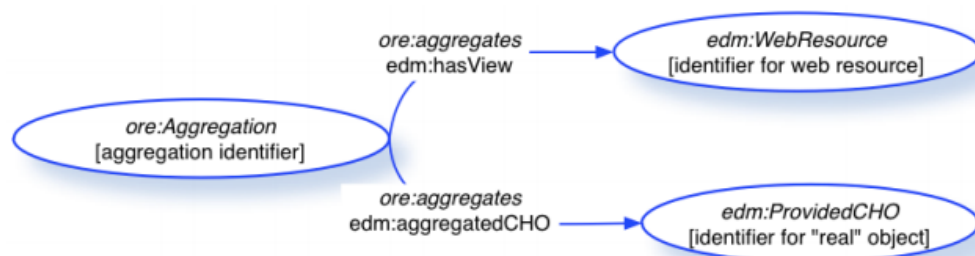


Figure 18: The three core classes in EDM and the properties connecting them (Isaac, 2013)

Typically, data about the object in question, such as creator information, gets attached to the ProvidedCHO itself. The Aggregation property holds information about the organization responsible for the descriptive data, named using the `edm:dataProvider` property.

Depending on the object in question, attaching data directly to the ProvidedCHO is suitable. When using EDM to take data from many different providers and describe the same object in different ways, however, this method has problems, such as multiple institutions attempting to attach creator names information, which may be slightly different, to the same ProvidedCHO. To alleviate this, EDM utilizes the `ore:Proxy` property from the Object Re-Use and Exchange model and vocabulary. This slightly alters the core three-property model and introduces a fourth. The `ore:Aggregation` property still collects digital representations and descriptive data about a single ProvidedCHO, but that descriptive data is attached to an instance of `ore:Proxy`, which acts as a representation of the real world object. This means that many different data providers can describe the same object, or `edm:ProvidedCHO`, as long as the data is attached to an `ore:Proxy`. A visualization of this is shown in Figure 19.

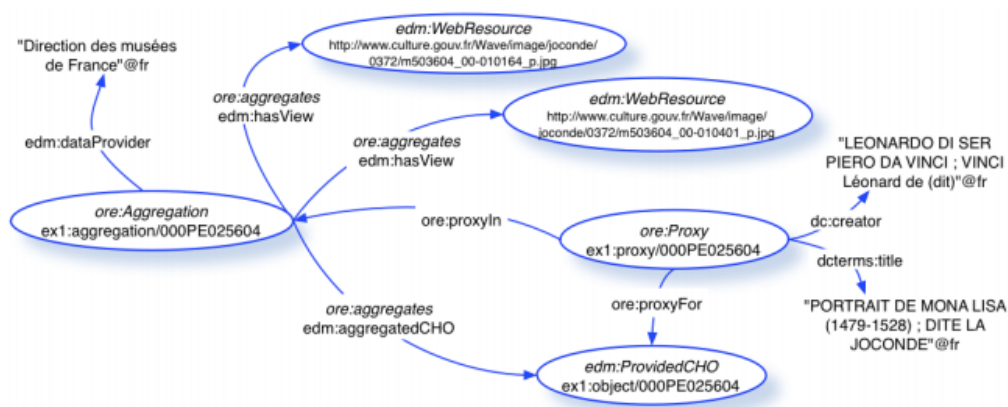


Figure 19: A single provider's aggregation, with descriptive data attached to the `ore:Proxy` (Isaac, 2013)

This aggregation method, allowing multiple institutions to describe the same real world object in different ways, is vital in combining bibliographic data from different institutions for the same real world manga. The use of EDM for manga aggregation will be shown in Section 4.1.

4. MODELING BIBLIOGRAPHIC DATA FOR MANGA & ITS AGGREGATION

4.1 Aggregation Model for Manga Metadata

This section will introduce and describe a bibliographic data aggregation model that utilizes EDM and BIBFRAME. While this section will feature the overall model, including its bibliographic description aspects, the focus will be on the aggregation methods. Specifics about bibliographic description will be discussed in Section 4.2. Due to its size, the model is difficult to display in full in a readable manner. Therefore, while a small visualization is shown in Figure 20, this section will discuss the model in several individual groupings. The model in full can be found in Appendix A, with a generalized version in Appendix B.

This aggregation model establishes relationships amongst multiple levels of manga resources, while still allowing for their differing levels of bibliographic description. This model uses a similar BIBFRAME to EDM path proposed by Zapounidou et al. (2014b), but differs in the way it uses `edm:InformationResource`. It also incorporates the use of the `edm:PhysicalThing` property and a method of connecting ordered manga volumes based on an EDM library metadata alignment report (Angeli, et al., 2012).

The data aggregation shown in this model consists of two main types. The first is the aggregation of data from both the Work and Item FRBR entities, which gathers data about manga in general and its specific volume instances, with Work data being provided by Web resources, and Item data by other institutions. The second is the aggregation of data for the Item level, or data about specific volumes of manga coming from various institutions. The data from multiple providers was described according to the model (see Section 4.2), and identical manga resources from these different providers were found in order to identify data that was available to be aggregated; for example, Monash and Toppan would both have bibliographic data about the same edition of *One Piece* Volume 4, which was then identified and aggregated in the model. Section 4.3 discusses the identification of these related resources.

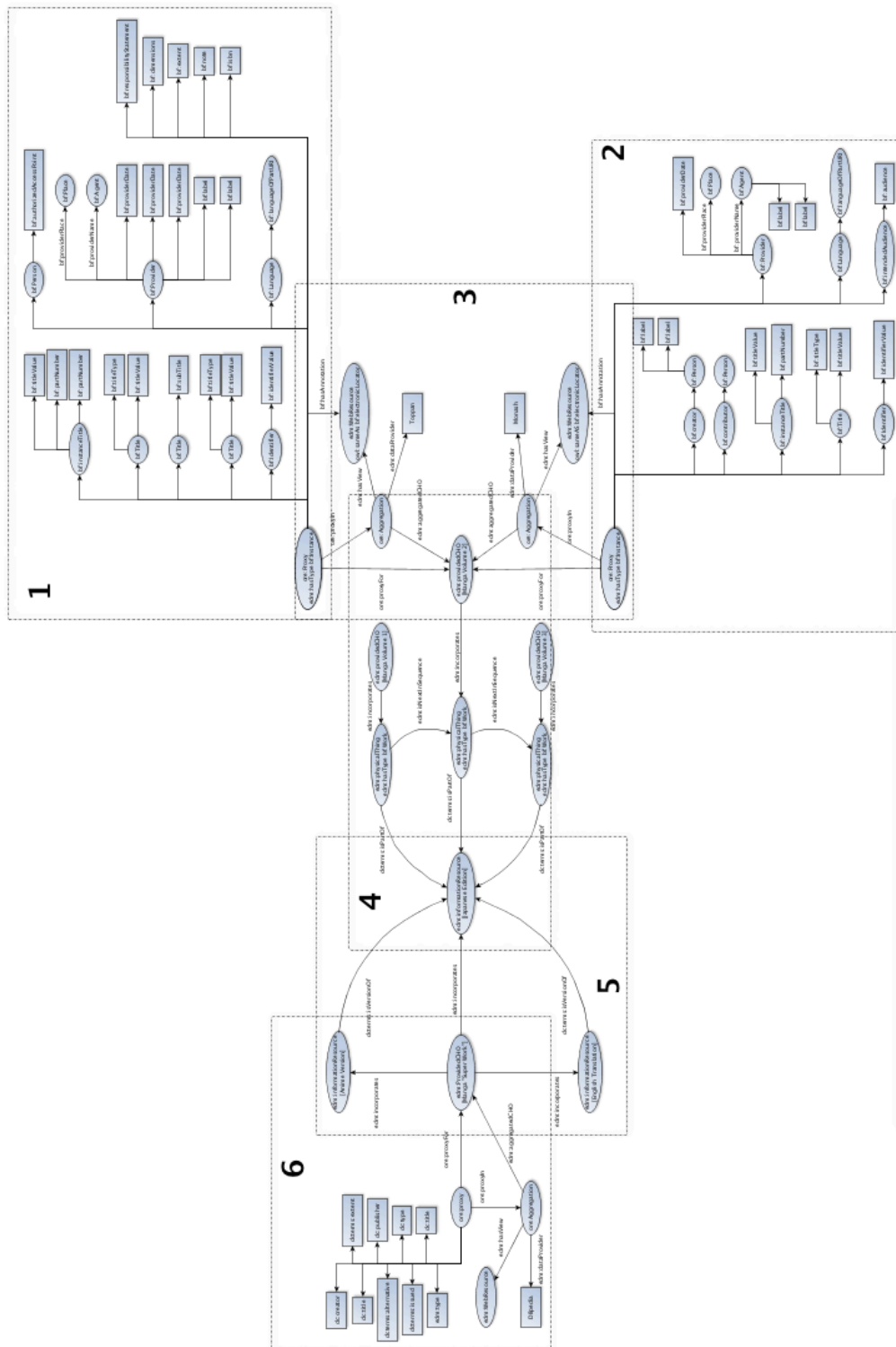


Figure 20: Overview of the full aggregation model

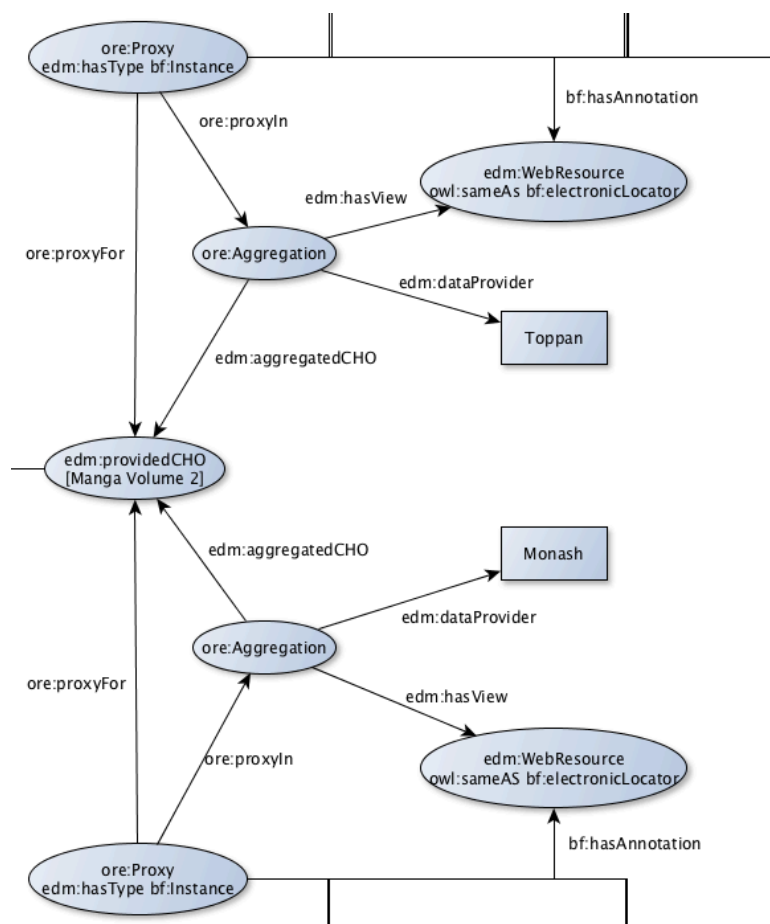


Figure 21: The aggregation of data from multiple sources for a single manga volume

As stated previously in this thesis, the aggregation model sought to serve various FRBR entities, or “levels” of bibliographic description, from the broader Work level to the more specific Item level. At the core of the Item level portion of the model is the aggregation of data from multiple sources for a single volume of manga. In Figure 20, this portion is identified in the Group labeled 3, with a focused view available in Figure 21, and a full-page view at Appendix E. This view features all of the core classes of the EDM model, along with the necessary `ore:Proxy` property required for maintaining unique data provided by different institutions. Here, `edm:ProvidedCHO` represents a single volume of manga, the second in a series, labeled “Manga Volume 2.” `ore:Aggregation` provides information on the separate data providers, in this case Toppan and Monash, as well pointing to the required `edm:WebResource` property. Their unique bibliographic data is described in Figure 20’s Groups 1 and 2, and are connected to their own respective `ore:Proxy` properties. This allows

for Monash and Toppan to describe the same volume of manga, or ProvidedCHO, from their respective point of views. While Toppan and Monash are used as examples, there is no limitation on how many institutions can provide data for a single volume.

Giving the `ore:Proxy` property the RDF type `bf:hasInstance` may appear an unorthodox decision, but it is suitable, as all of the bibliographic data is attached to the proxy, and not the `ore:Aggregation` or `edm:ProvidedCHO`. The use of the `bf:hasAnnotation` property, which labels the `edm:WebResource` as a `bf:electronicLocator` via the `owl:sameAs` property, is necessary, because of EDM's mandatory requirement of `edm:WebResource`. While examples of EDM commonly use a digital photo of the ProvidedCHO as its WebResource, a similar digital representation for a textual object is more difficult to produce. As URI representing the ProvidedCHO is an allowed form of a WebResource, it was decided that the use of the `bf:electronicLocator` property, which represents an electronic location from which the resource is available (see <http://bibframe.org/vocab/electronicLocator.html>), would be the most consistently available WebResource to use. Zapounidou et al. (2014a, 2014b) used this property in a similar role.

This portion of the model is the main area of data aggregation, and focuses on individual volumes of manga that reflect resources close to the FRBR Item level. The specifics of each institutions bibliographic description will be shown in Section 4.2, but this aggregate portion is required to bring together multiple institutional data for a single resource, while maintaining the unique point of view and content that each institution provides. The next portion of the model to be examined is shown in Figure 20 as the Group labeled 4, with a more focused view available in Figure 22, and full-page view at Appendix F.

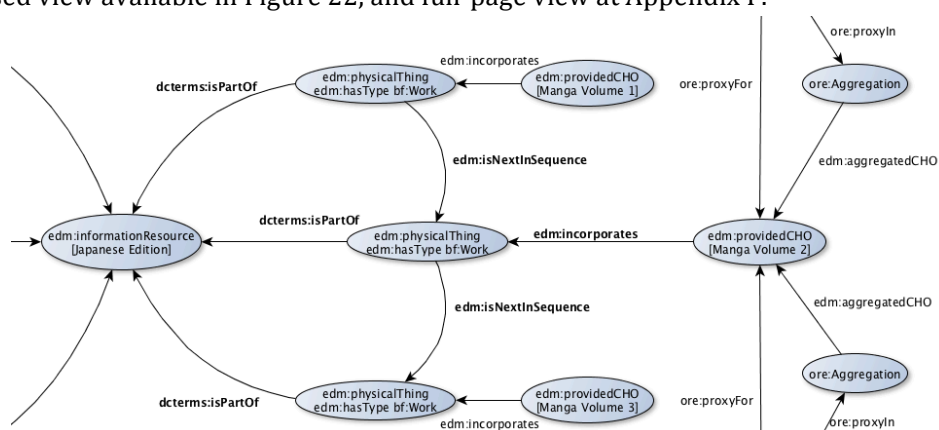


Figure 22: Connecting successive volumes of manga in the aggregation model

Figure 22 focuses on the portions of the aggregation model that allows for the serial nature of manga, that is, a sequence of volumes sharing a successive relationship, to be incorporated. Information useful in forming this portion of the model was found in the EDM

library metadata alignment report (Angjeli, A., et al., 2012). Section 6 of the report examined the use of EDM for serials, with a focus on journals and their articles, and modeled two types of serial relationships. The first type modeled hierarchical relationships, aimed at representing the relationships between serials and their contents using the Dublin Core connecting property `dcterms:isPartOf`. While useful for describing the relationships between articles and the journals they reside in, it is less useful for a resource like manga. The second type of relationship described was more suitable for manga and modeled a successive relationship type using the property `edm:isNextInSequence`. This property logically connects one resource with its successor, and the use of this in aggregation model can be seen in Figure 22, where three `edm:physicalThing` entities, each representing a different `edm:ProvidedCHO` (unique manga volume) are connected in succession using `edm:isNextInSequence`.

The use of `edm:PhysicalThing` in this model is unique and was not used a similar way, either by Zapounidou et al. (2014a, 2014b), or (Angjeli, A., et al., 2012). The author felt the use of `edm:PhysicalThing` was necessary to distinguish between the different levels of FRBR entities. Instead of using the broad class, `edm:InformationResource` (see <http://onto.dm2e.eu/edm#InformationResource>), `edm:PhysicalThing` is used to represent the manga at the volume, or Item level. This not only is a more accurate representation of the manga volume resource, but also frees the `edm:InformationResource` property for use with FRBR Work level descriptions. This is shown in Figure 22, where each `edm:PhysicalThing` entity is linked to a single `edm:InformationResource` entity, representing the manga at the FRBR Work level, using the `dcterms:isPartOf` property. Therefore, the portions of the model examined thus far allow for the aggregation of data for specific manga volumes from different institutions, while also allowing for multiple relationships to be made: relationships between successive volumes of manga, and relationships from individual manga volumes to the greater conceptual Work.

The portion of the aggregation model that features the Work level is shown in Figure 23, and in the Group labeled 5 in Figure 20, with a full-page view at Appendix G.

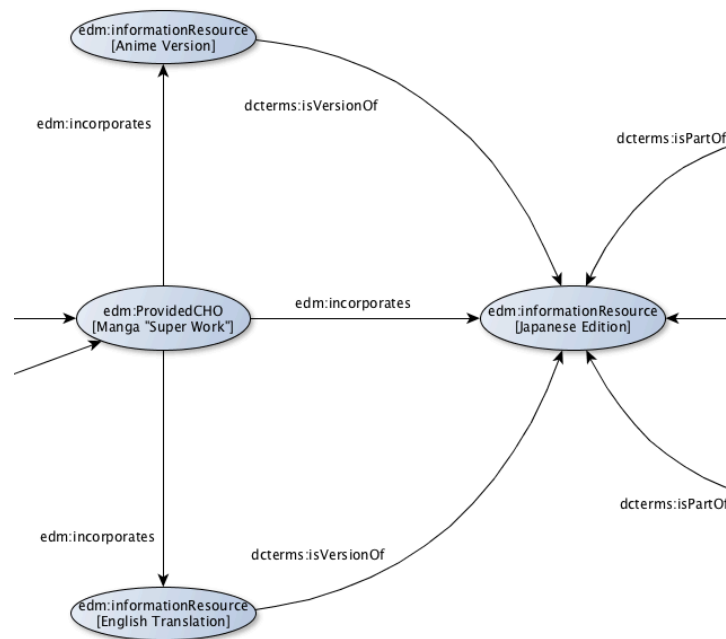


Figure 23: Portion of the aggregation model showing the Work level entities

The `edm:InformationResource` property labeled [Japanese Edition] represents the Japanese manga, and therefore, like Wikipedia entries, the main conceptual Work of the resource. In order to attach bibliographic data to the Work level, an instance of `edm:ProvidedCHO` is attached. Its label, “Super Work,” has no real semantic meaning, but is meant to distinguish between the BIBFRAME Work properties used on the Volume side of the model. This “Super Work” represents the conceptual Work to a similar level as a Wikipedia entry. Thus, this is where bibliographic data at the Work level is described in the model. While this will be examined in Section 4.2, the descriptive data for this level can be seen in the Group labeled 6 in Figure 20.

While the aggregated manga and volume relationship data shown in Figures 21 and 22 were about the Japanese manga resource, shown in Figure 22 as the `edm:InformationResource` labeled “Japanese Edition,” Figure 23 also shows how related resources can fit into the model. The examples used are an English translation of the manga, and an anime adaptation. While it may seem like a different media format, such as the anime adaptation, would not be directly linked to the `ProvidedCHO` that represents the manga, as shown in the discussion on Work-level Web data in Section 3.2.4, Work-level Web data, such as a Wikipedia entry, represent both the main entry for the manga and the main umbrella entry for other media formats. The aggregation model accommodates this by treating the separate

mediums as unique instances of the `edm:InformationResource` property, linked to the “super work” `ProvidedCHO` using `edm:incorporates`.

This concludes discussion on the non-descriptive portions of the model. The segments of the model described here enable the aggregation model to perform several different tasks. First, and central to the idea of aggregation, is the aggregating of bibliographic data for manga from multiple sources. This allows a single manga resource to be described from different points of view and of different levels of granularity, while identifying where the data has come from. Next, the model allows for the successive volume-to-volume relationships of manga to be modeled in EDM, while still allowing for the aggregation of data for each volume, and also containing the volumes conceptually in a single resource, representing the full manga series. The final portion discussed shows how the more conceptual Work level for manga is represented in the model, as well as how other media formats may be represented and linked to manga resources. The next section will examine the portions of the model focused on bibliographic description.

4.2 Bibliographic Data for Manga

This section will discuss portions of the model that are responsible for bibliographic description for manga. After the data examination from Section 3.1 and 3.2, developing bibliographic description methods that fit with the aggregation model was relatively straightforward. The decision was made to describe the two different “sides” of the model, that is, the portions of the model describing the conceptual Work level, and the portions describing specific volumes of manga, in different ways; the volume portion utilizes BIBFRAME and will be discussed in Section 4.2.1, while the Work level is described using standard EDM Dublin Core properties, which will be shown in Section 4.2.2.

4.2.1 Manga Volumes Described in BIBFRAME

After the examination of BIBFRAME, discussed in Section 3.3.2, it was determined that model and vocabulary was suitable for use in the aggregation model. Not only did the properties of the BIBFRAME vocabulary allow for a high level of granularity, which is required at the volume or Item level, but also the lack of dependence on traditional library technologies meant that BIBFRAME was useful for non-library institutions that also wanted to describe granular levels of bibliographic data.

To illustrate the use of BIBFRAME with non-library sources, the model uses Toppan and Monash as data providers for a single manga volume. Their respective data is described using the BIBFRAME vocabulary, and applied to the same volume of manga, as seen in the

aggregation model discussion in Section 4.1. While the examination of library bibliographic data for manga was important in the formulation of the model, it is not featured in this aggregation example. The inclusion of library data, along with data from Toppan and Monash, is certainly possible with this aggregation model, though libraries stand more to gain from the aggregation of manga bibliographic data, rather than be providers of a large amount of unique manga data with which they can contribute.

Figure 24 (full-page at Appendix C) illustrates how BIBFRAME properties are used to apply Toppan's data to the total aggregation model, while Table 3 shows the mapping between Toppan's data and the BIBFRAME vocabulary. Of note are four properties that contained Sufficient data levels in Table 2, but were omitted in the final data application: 著者典拠 ID (Author Authority ID), レーベル番号 (Label Volume ID), レーベル典拠 ID (Label Authority ID), and 出版者典拠 ID (Publisher Authority ID). These properties were omitted, as the author was unable to determine the semantic meaning of both the properties and their respective values.

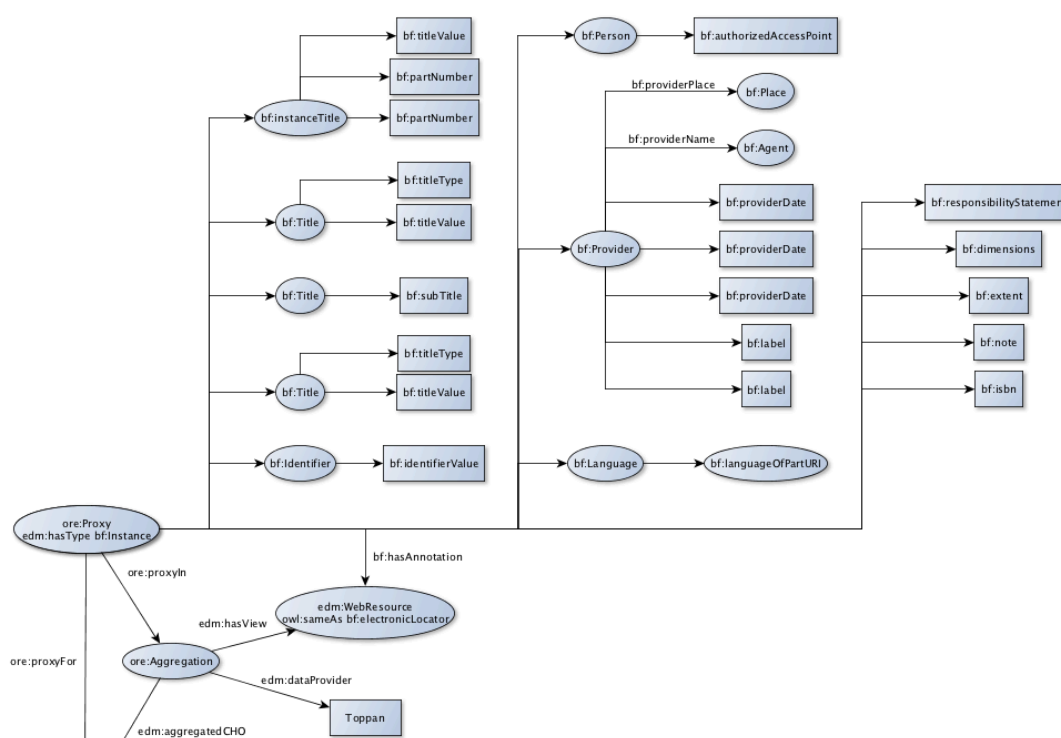


Figure 24: Toppan data described using BIBFRAME properties

Table 3: Used Toppan headers and their mapped BIBFRAME properties

| Original Japanese Header Name | English Translation | BIBFRAME Properties | |
|-------------------------------|---------------------------------------|---------------------------|-------------------------------|
| マンガ単行本 ID | Manga Volume ID | bf:Identifier | bf:identifierValue |
| マンガ単行本名 | Manga Volume Title | bf:InstanceTitle | bf:titleValue |
| 巻 | Volume | | bf:partNumber |
| 巻ソート | Volume Order | | bf:partNumber |
| マンガ単行本名 ヨミ | Manga Volume Title, Reading | bf:Title | bf:titleType bf:titleValue |
| マンガ単行本名 追記 | Manga Volume Title Supplement | bf:Title | bf:subTitle |
| マンガ単行本名 追記ヨミ | Manga Volume Title Supplement Reading | bf:Title | bf:titleType bf:titleValue |
| 責任表示 | Statement of Responsibility | bf:resonsibilityStatement | |
| 標目 | Author Access Point / Entry Heading | bf:Person | bf:authorizedAccessPoint |
| 初版発行年(西暦) | Year First Issued | bf:Provider | bf:providerDate |
| 初版発行月 | Month First Issued | | bf:providerDate |
| 初版発行日 | Day First Issued | | bf:providerDate |
| 単行本レーベル (サブレーベル) | Publisher Label First Issued In | | bf:label |
| 単行本レーベル ヨミ | Publisher Label First Issued In | | bf:label |

| | | | | |
|-----------|---------------------------------------|---------------|----------------------|----------|
| | Reading | | | |
| 出版者名 | Publisher | | bf:providerName | bf:Agent |
| 出版地 | Place of Publication | | bf:providerPlace | bf:Place |
| ページ数 | Number of Pages | bf:extent | | |
| 縦の長さ_横の長さ | Physical Dimensions | bf:dimensions | | |
| ISBN | ISBN | bf:isbn | | |
| 全国書誌番号 | Japanese National Bibliography Number | bf:Identifier | bf:identifierValue | |
| 言語区分 | Language | bf:Language | bf:languageOfPartURI | |
| 分類 | Nippon Decimal Classification (NDC) | bf:Identifier | bf:identifierValue | |
| メモ | Memo | bf:note | | |

Apart from the four aforementioned properties, all of the “Sufficient” level Toppan data has been expressed in BIBFRAME. While some property usage is not optimal, such as the use of multiple `bf:titleType` and `bf:titleValue` pairings to describe the “reading” of a title, the vocabulary allowed for all of the required bibliographic data to be recorded without a loss of granularity from Toppan’s original data. A similar display of information for Monash’s data can be seen in Figure 25 (full-page at Appendix D) and Table 4.

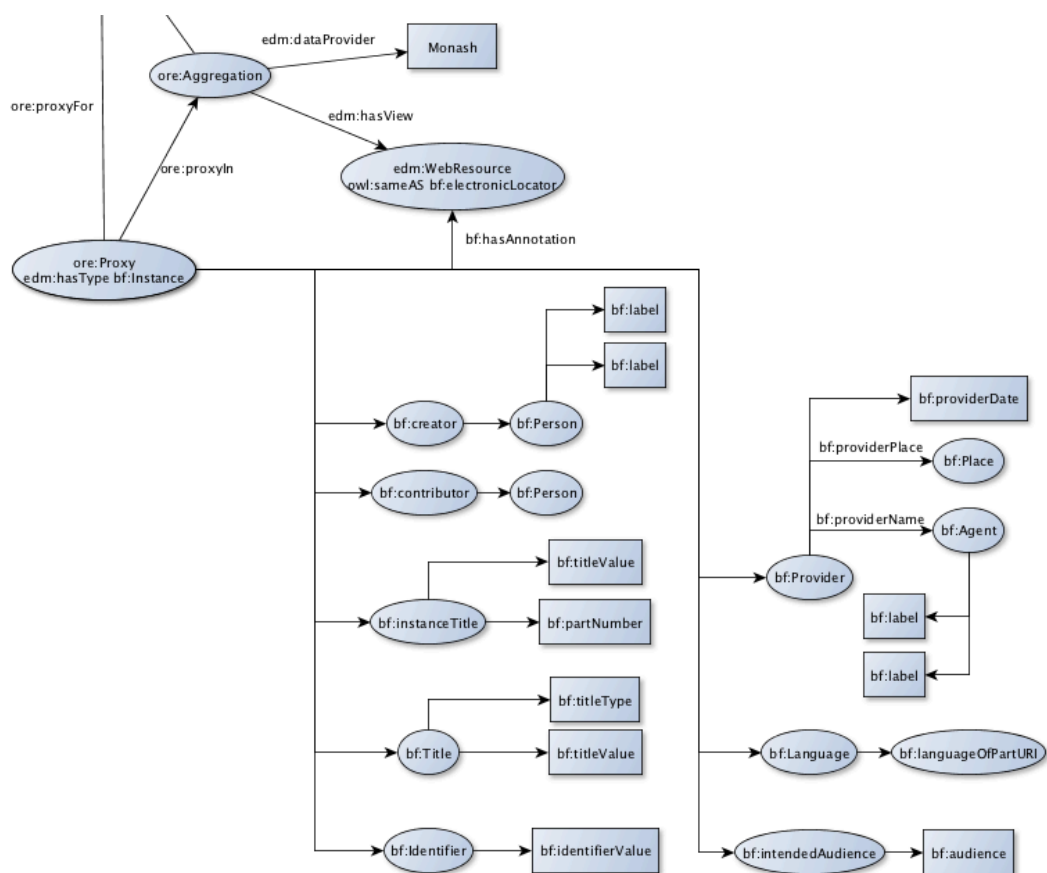


Figure 25: Monash data described using BIBFRAME properties

Table 4: Monash headers and their BIBFRAME properties

| Monash Header Name | BIBFRAME Properties | | |
|--------------------|---------------------|-------------------------------|----------|
| ID | bf:Identifier | bf:identifierValue | |
| Title | bf:instanceTitle | bf:titleValue | |
| Vol (Volume) | | bf:partNumber | |
| JapaneseTitle | bf:title | bf:titleType bf:titleValue | |
| Author | bf:creator | bf:Person | bf:label |
| JapaneseAuthor | | | bf:label |

| | | | |
|-------------------|---------------------|----------------------|----------|
| Translator | bf:contributor | bf:Person | |
| Publisher | bf:Provider | bf:providerName | bf:label |
| JapanesePublisher | | | bf:label |
| PublishedIn | bf:Provider | bf:providerDate | |
| Rating | bf:intendedAudience | bf:audience | |
| Language | bf:Language | bf:languagePartOfURI | |

Similar to Toppan's data, Monash's is described accurately using BIBFRAME, without a loss in granularity. In fact, using BIBFRAME, there is room for more granularity in Monash's data without the need of supplemental information. While the general `bf:contributor` role is used to describe the Translator, one can also use the Relator entity, along with the property `bf:relatorRole`, to specify that the role is that of a translator. Another property that can add additional information is `bf:audienceAssigner` to specify that Monash created the rating system. In a hypothetical Linked Data scenario, this property would be useful if Monash had a resource such as a Manga content rating scale available for other institutions to use; making closed institutional data available in the Linked Data space would allow for the creation of this and other interesting manga Web resources and authorities.

4.2.2 Manga Works Described in EDM and DC

While bibliographic data describing specific volumes of manga is performed using the BIBFRAME model and vocabulary, data that describes the conceptual Work level is more straightforward, and utilizes the common Dublin Core terms, found in the EDM library metadata alignment report (Isaac, 2013), available for use with EDM. There are two main reasons for the use of DC here, rather than BIBFRAME, as was done with manga volumes. First, as bibliographic data for the Work level in this study is coming from the Web, using a common Web ontology like DC is more suitable – and realistic – for Web data providers than BIBFRAME.

While BIBFRAME was shown to be suitable for institutions outside of libraries, these data providers perform levels of bibliographic description similar to that of libraries, so arguing for the use of BIBFRAME is rational. Despite being suitable for use in non-library communities, it is unlikely BIBFRAME would be adopted over DC by institutions producing primarily Web data. The second reason is that the DC terms used in EDM can be used to describe a level of granularity that is suitable at the Work level. Section 3.2.4 discussed how Web data contained high levels of granularity, this information, such as chapter lists and

volume summaries, would be applied at the manga volume or Item level. The information applied to the Work level, such as creator and series title information, is typically not very granular, and is able to be described properly using DC terms.

The model still allows for granular data about the Work to be described, even using BIBFRAME if one desired, but the example uses Web data and DC, presuming this scenario to be the norm. A simple mapping of DBpedia properties to DC properties used in EDM was performed, with an example for the DBpedia page for the Wikipedia entry on the manga Astro Boy (see http://dbpedia.org/page/Astro_Boy) shown in Table 5, with a visualization of the data within the aggregation model shown in Figure 26, with a full-page view at Appendix H.

Table 5: DBpedia properties and their respective EDM DC properties, along with the literal values taken from the “Astro Boy” DBpedia pages

| DBpedia Property | EDM DC Property | Value from DBPedia |
|---|---------------------|--------------------|
| dbpedia-owl:author | dc:creator | Osamu Tezuka |
| dbpedia-owl:numberOfVolumes | dcterms:extent | 23 |
| dbpedia-owl:publisher (first published by) | dc:publisher | Kobunsha; Kodansha |
| dbpedia-owl:type | dc:type | manga |
| dbpprop:name | dc:title | Astro Boy |
| dbpprop:jaKanji | dc:title | 鉄腕アトム |
| dbpprop:jaRomaji | dcterms:alternative | Tetsuwan Atomu |
| dbpprop:first | dcterms:issued | April 1952 |
| dbpedia-owl:type | edm:type | manga (TEXT) |

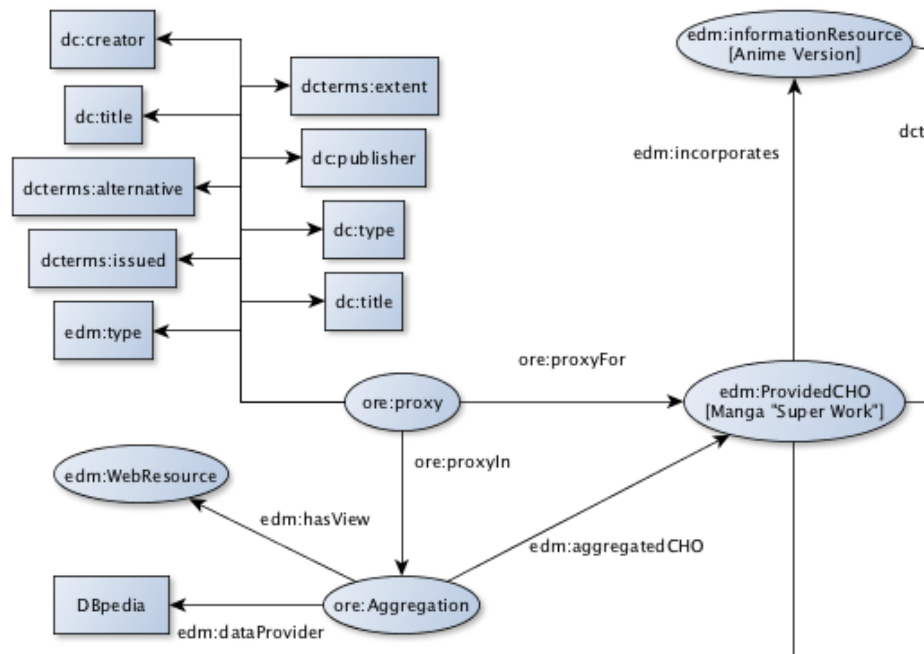


Figure 26: DBpedia's data attached to the Work level resource

Note for the use of `edm:type` in this table: `edm:type` is used to classify an object as TEXT, IMAGE, SOUND, VIDEO, or 3D (Europeana, 2013). While DBpedia fails to provide a property specifying the resource type at this level, one can use the `dbpedia-owl:type` property, which specifies the resource as a manga, to infer that the resource is a TEXT, and label it accordingly. This property is mandatory for use in the EDM portal, so its omission is also a possibility.

As the table shows, the information described is relatively basic in terms of its bibliographic granularity, but it is suitable as it is being applied to the level of the conceptual Work. Of note here is the `dc:type` property being given the value of "manga". While DCMI recommends the use of a controlled vocabulary for this term (see <http://dublincore.org/documents/dcmi-terms/#terms-type>), it is not required. In EDM, one must also select one of four mandatory properties to use: `dc:subject` or `dc:coverage` or `dc:type` or `dcterms:spatial` (Angjeli et al., 2012). Using `dc:type` to specify the Work as a manga appears to describe the most relevant piece of bibliographic data when compared to the other mandatory choices. As He et al. (2013) demonstrated, DBpedia is useful as an authority for

manga due to lack of traditional alternatives, so using it for manga-related controlled vocabulary terms is also a possibility.

Aside from the different vocabulary used, Web data describing the conceptual Work is fairly similar to institutional data describing the Item, or volumes. Unlike the use of the BIBFRAME property `bf:electronicLocator` to identify holdings URI for volume data, Web data can use more orthodox values for the `edm:WebResource` property. In the given example, the Wikipedia article URL itself can be the WebResource representation of the ProvidedCHO representing the total manga Work.

While this study used one source of data, Wikipedia, for the Work level, the aggregation of multiple sources would work similarly to the volume level. During the seeking and gathering of bibliographic data for manga, Wikipedia was deemed to be a sufficient source on its own for providing general, Work level data. General Work data, such as main title and creator names, could be aggregated from multiple sources, though typically different data providers do not provide unique or conflicting data for these main properties. More granular data that varies between institutions or Web data sources is typically aimed at specific volumes of manga. Still, the model allows for aggregation at this level and provides a method for the useful aggregation for volumes of manga. The next section will discuss the identifying of resources for aggregation.

4.3 Identifying Related Manga Data

This section will discuss methods of identifying bibliographic manga data that comes from different sources, yet describes the same resource. These methods differ depending on whether the aggregation of data is for the conceptual manga Work, or individual volumes of manga. While an automated method of identifying related data was not investigated, the use of OpenRefine software, also used in Section 3, was also used in aiding with identifying related bibliographic data for manga.

As the Work level of the aggregation model is of a conceptual level that is similar to the FBRB Work, aggregating data at this level entails the identification of the conceptual Work for which the resource in question is a part of. Aspects of a resource such as language, publisher, edition, and even media format, do not necessarily need to be taken into consideration when aggregating data at this level. As the model utilizes general Web data to represent the conceptual Work, namely Wikipedia articles, which, as discussed earlier, act as umbrella entries for the entire intellectual property of a manga series, any bibliographic data describing the resource at the Work level can be aggregated if the resources match a Wikipedia entry. OpenRefine was used in order to perform this data-to-Wikipedia matching.

OpenRefine features a Reconciliation function to link and extend data with various webservices (OpenRefine, n.d.). The Reconciliation function matches a column from tabular data inside the software against a Web service of the user's choosing. These services are typically Linked Data sources with an accessible SPARQL endpoint, with the option to add additional services through a URL. The main services used in this study were DBpedia and VIAF, access to which is provided by <http://refine.codefork.com/>. Figure 27 shows sample Monash data inside OpenRefine, with the Title column selected for Reconciliation.

| Show as: rows records | | Show: 5 10 25 50 rows | | | | | |
|------------------------------|------|-----------------------|----------------------------------|------------|----------|------------------|------|
| ▼ All | ▼ ID | ▼ Title | ▼ JapaneseTitle | ▼ Vol | ▼ Author | ▼ JapaneseAuthor | |
| ☆ | 1. | 10107 | Facet | ナン (konan) | 6 | Aoyama, Goushou | 青山剛昌 |
| ☆ | 2. | 10109 | Text filter | ナン (konan) | 5 | Aoyama, Goushou | 青山剛昌 |
| ☆ | 3. | 10110 | Edit cells | ナン (konan) | 4 | Aoyama, Goushou | 青山剛昌 |
| ☆ | 4. | 10112 | Edit column | ナン (konan) | 3 | Aoyama, Goushou | 青山剛昌 |
| ☆ | 5. | 10113 | Transpose | ナン (konan) | 3 | Aoyama, Goushou | 青山剛昌 |
| ☆ | 6. | 10115 | Sort... | ナン (konan) | 3 | Aoyama, Goushou | 青山剛昌 |
| ☆ | 7. | 10118 | View | ナン (konan) | | | 青山剛昌 |
| ☆ | 8. | 10119 | Reconcile | | | | 青山剛昌 |
| ☆ | 9. | 10122 | Start reconciling... | | | | 青山剛昌 |
| ☆ | 10. | 10124 | Facets | | | | 青山剛昌 |
| | | | QA facets | | | | 青山剛昌 |
| | | | Actions | | | | 青山剛昌 |
| | | | Copy reconciliation data... | | | | 青山剛昌 |
| | | | Discover related RDF datasets... | | | | 青山剛昌 |

Figure 27: Reconciling the Title column of Monash's data

Freebase Query-based Reconciliation

Sindice

DBpedia

DBP

DBPJP

LOC

Virtual International Authority File

Reconcile each cell to an entity of one of these types:

☐ owl:Thing

<http://www.w3.org/2002/07/owl#Thing>

☐ schema:CreativeWork

<http://schema.org/CreativeWork>

☐ http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#InformationEntity

☐ dbo:Work

<http://dbpedia.org/ontology/Work>

☐ dbo:Comics

<http://dbpedia.org/ontology/Comics>

☒ dbo:Manga

<http://dbpedia.org/ontology/Manga>

☐ dbo:WrittenWork

<http://dbpedia.org/ontology/WrittenWork>

☐ yago:Abstraction100002137

<http://dbpedia.org/class/yago/Abstraction100002137>

☐ yago:Act100030358

<http://dbpedia.org/class/yago/Act100030358>

☐ Reconcile against type:

☐ Reconcile against no particular type

☒ Auto-match candidates with high confidence

Figure 28: Recommended DBpedia properties inside of the Reconciliation menu based off of the selected column data

After selecting a column for Reconciliation, the menu in Figure 28 appears, allowing the user to select a service for which to match the data against. A short analysis is performed by the software, which then suggests recommended properties from the service of choice. In Figure 28, relevant recommended properties from the DBpedia ontology, such as Manga, Comics, and WrittenWork are shown. Even if no relevant properties are found for which to automatically match data against, Reconciliation can be performed on the column and the user can manually select an entity that matches a particular set of data. Figure 29 illustrates this option.

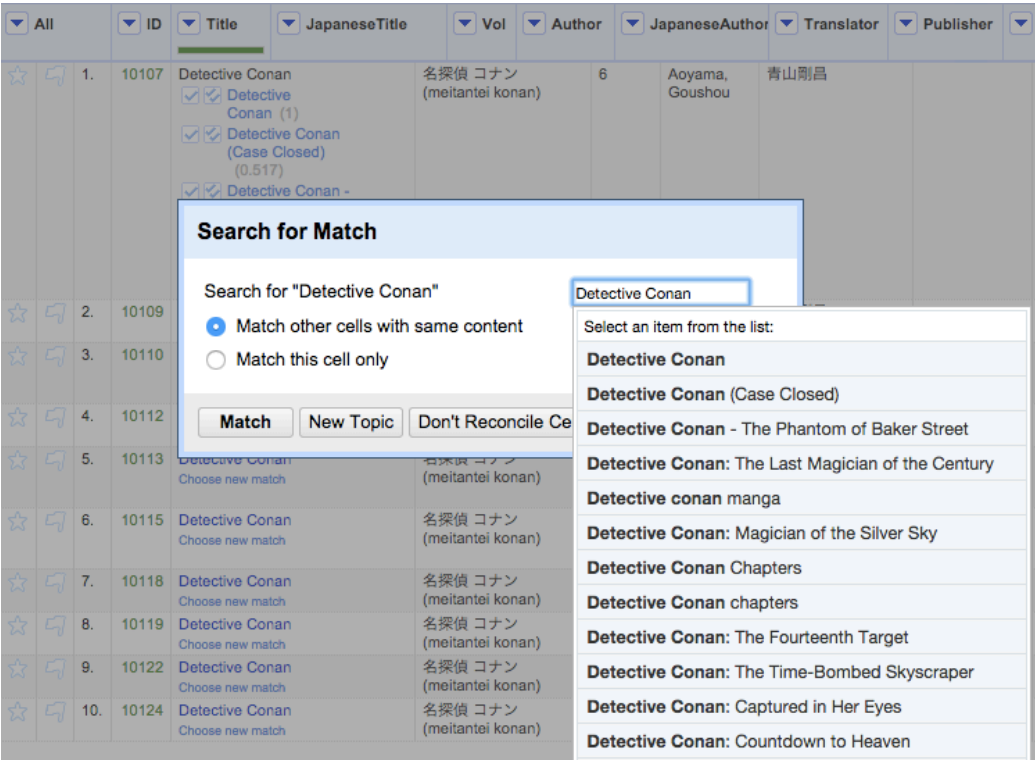


Figure 29: A list of possible matching DBpedia articles based on the data value “Detective Conan”

Once a match is found, the cell will have an embedded hyperlink to the relevant Web resource, which in these examples, is a link to the relevant DBpedia page. In order to display the relevant URL, the “Add column based on this column” function is used on the reconciled column, along with the GREL expression `cell.recon.match.id`. This produces another column in the data based on the reconciled column’s URL. Figure 30 shows the end result of this function, with a new column titled DBpedia URL added. Note that in this figure, the Title column values have been changed from “Detective Conan” to “Case Closed.” Case Closed is the

official English release name of the manga 名探偵コナン, which was translated as Detective Conan, but changed to Case Closed for legal reasons. While the values read Case Closed in OpenRefine based on the title of the reconciled URL, the original data is unchanged, so exporting the data would still yield the original Detective Conan values.

Show as: **rows** records

Show: 5 10 25 50 rows










| <input type="checkbox"/> All | <input type="checkbox"/> ID | <input type="checkbox"/> Title | <input type="checkbox"/> DBpedia URL | <input type="checkbox"/> Japanese Title | | |
|---|---|--------------------------------|--------------------------------------|---|---|----------------------------|
|  |  | 1. | 10107 | Case Closed Choose new match | http://dbpedia.org/resource/Case_Closed | 名探偵コナン Choose new match |
|  |  | 2. | 10109 | Case Closed Choose new match | http://dbpedia.org/resource/Case_Closed | 名探偵コナン Choose new match |
|  |  | 3. | 10110 | Case Closed Choose new match | http://dbpedia.org/resource/Case_Closed | 名探偵コナン Choose new match |
|  |  | 4. | 10112 | Case Closed Choose new match | http://dbpedia.org/resource/Case_Closed | 名探偵コナン Choose new match |
|  |  | 5. | 10113 | Case Closed Choose new match | http://dbpedia.org/resource/Case_Closed | 名探偵コナン Choose new match |

Figure 30: A DBpedia URL column added based on the reconciled URLs from the Title column

This reconciled URL signifies the same Wikipedia entry that represents the Work level `edm:ProvidedCHO` for which data is aggregated against in Figure 26. As the Wikipedia entries for manga series’ are umbrella resources for many different entities, the same URL is found based on the title and agnostic of other data. Importantly, this reconciling works across languages thanks to DBpedia properties such as `dbpprop:jaKanji`, as seen in Table 5, allowing for the aggregation of data not only from different resources, but of different languages. In other words, performing Reconciliation with DBpedia on Toppan’s data with the title 名探偵コナン will produce the same DBpedia URL as Monash data titled “Detective Conan.”

While using Reconciliation to obtain a URL to aggregate data around is suitable for the Work level, the Item level is less straightforward. This is due to the fact that when aggregating data at the Item level, one cannot ignore details regarding specific volumes of manga, such as publisher, edition, or volume number. Aggregation at this level is possible if one manually examines the data to identify whether the data matches fully, but a more efficient way of

aggregating data at this level has yet to be determined. OpenRefine can make a portion of this task easier, however, using the `cell.cross` GREL function. This function allows a user to compare data across multiple datasets in order to find related items. Figure 31 illustrates the usage of this function. In this figure, the DBpedia URL column from the Monash data is being matched against the DBpedia URL column from the Toppan data ("toppanastro" in the GREL expression), and is returning with the value of the unique Toppan identifier column マンガ単行本 ID. The result of this, shown in Figure 32, is a new column in the Monash data labeled Toppan ID containing the Toppan identifier value for data that reconciled the same DBpedia URL as the Monash data. The full GREL expression is as follows:

```
if (value!='null',cell.cross("toppanastro", "DBpedia
URL").cells["マンガ単行本 ID"].value[0], '')
```

Add column based on column DBpedia URL

New column name:

On error: ☒ set to blank ☐ store error ☐ copy value from original column

Expression: Language: No syntax error.

Preview History Starred Help

| row | value | if (value!='null',cell.cross("toppanastro", "DBpedia URL").cells["マンガ単行本 ID"].value[0],") |
|-----|---------------------------------------|---|
| 1. | http://dbpedia.org/resource/Astro_Boy | MMM000057706 |
| 2. | http://dbpedia.org/resource/Astro_Boy | MMM000057706 |
| 3. | http://dbpedia.org/resource/Astro_Boy | MMM000057706 |
| 4. | http://dbpedia.org/resource/Astro_Boy | MMM000057706 |
| 5. | http://dbpedia.org/resource/Astro_Boy | MMM000057706 |
| 6. | http://dbpedia.org/resource/Astro_Boy | MMM000057706 |

OK Cancel

Figure 31: Use of the `cell.cross` function in OpenRefine

| ID | Title | DBpedia URL | Toppan ID | JapaneseTitle | JapaneseTitle R | Vol |
|-------|-----------|---|--------------|---|-----------------|-----|
| 22097 | Astro Boy | http://dbpedia.org/resource/Astro_Boy | MMM000057706 | 鉄腕アトム <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic Search for match | tetsuwan atomu | 1 |
| 22098 | Astro Boy | http://dbpedia.org/resource/Astro_Boy | MMM000057706 | 鉄腕アトム <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic Search for match | tetsuwan atomu | 2 |

Figure 32: Retrieved ID value from Toppan's data that matches Monash's reconciled DBpedia URL

This function is not restricted to DBpedia URLs and can be performed using other values such as titles themselves, or VIAF URLs from author reconciliation. As matching similar data like publisher, published date, and volume number is needed to identify resources available for aggregation at the volume level, it would be valuable to match this data as well. Unfortunately, this values for this data such as “2003” or “Volume 2” are not limited by the manga series, and would return results from an entire dataset. The ability to use the `cell.cross` function with multiple columns simultaneously would make the identification of the same manga volume across resource possible inside OpenRefine, even across datasets of different languages. One could, for example, create an expression that matched reconciled DBpedia URLs for a title, VIAF URLs for an author, publisher name and year, and volume number, in order to match exact data for manga volumes. This does not appear to be possible currently, however, so an efficient way of aggregating manga data at the volume level remains to be achieved in this study; this limitation is discussed in Section 5.2. The current use of reconciliation was nevertheless useful, however, in reducing the workload for manually identifying matching data across different datasets, particularly in finding matching manga resources across languages. due to the ability to reconcile the same DBpedia URL for both English and Japanese resources.

5. SUMMARY AND CONCLUSIONS

This study developed a conceptual aggregation model that enables the collection of bibliographic data for related manga resources obtained from multiple institutions. The model allows for the aggregation of bibliographic data for manga at both broad conceptual “Work” level, and the more specific volume or “Item” level. The Europeana Data Model-based (EDM) model connects these levels together, while also preserving the successive volume-to-volume relationship of serialized manga. The model also utilizes the new BIBFRAME model and vocabulary for use with various types of institutions creating metadata for manga. This allows institutions to share their bibliographic data, no matter the level of granularity, as Linked Data for use on the Semantic Web. The following sections detail the significance of the study, its limitations, recommendations, and suggestions for further study.

5.1 Significance

The result of the study presented in this thesis is an aggregation model allowing for an increased level of manga bibliographic data from various providers to be available in the Linked Data space for use by other institutions, libraries, Web services, and others. The model relies on the use of EDM for the roles of bibliographic data aggregation from multiple providers for individual manga resources, and also the portrayal of relationships between them, be it a relationship is between successive volumes of manga, or a single volume and the intellectual work to which it belongs. For bibliographic description of data, the model uses EDM’s Dublin Core (DC) properties for data that is describing the manga at the conceptual “Work” level, typically sourced from Web resources, and for more granular data about individual volumes of manga, coming from institutions such as libraries and corporate databases, the new BIBFRAME model and vocabulary are used. The model, therefore, enables the aggregation of bibliographic data for manga at multiple conceptual levels, while maintaining the desired granularity level at each, establishing relationships between them, and making the data available on the Semantic Web.

The creation of the model relied on the analysis of bibliographic data for manga from various types of institutions. Data from academic libraries, a manga-focused special library, and a corporation handling the manufacturing of manga, was examined with the goal of understanding how different institution types handle bibliographic data for manga. It was discovered that multiple different institution types all are concerned with specific volumes of manga, thus describing manga at a level near the FRBR Item entity. Different institution types described different levels of information, but they were all concerned with data that dealt with specific volumes of manga rather than the series as a whole. Bibliographic data describing the

conceptual work was found on the Web, with a Linked Data source already available via DBpedia's offering of information found in Wikipedia articles. This Web data also contained more granular volume level data, though this data was not as accessible as the Work level data (see Section 3.2.4).

EDM was used as a basis for creating the model, as it allowed for the aggregation of data, sourced by multiple providers, to be brought together to describe the same resource from different points of view and of differing levels of granularity. A method of describing serials in EDM was employed to ensure that data could be aggregated about specific volumes of manga while at the same time connecting manga volumes via their natural successive relationships. The relationship types described in the model extend to the linking of manga volumes with the conceptual Work with which they are logically contained.

For the bibliographic description of manga resources, the study utilized the BIBFRAME model and vocabulary for individual volume of manga, and EDM-DC properties for data describing the conceptual Work. BIBFRAME allows for a level of granularity for bibliographic data that is not available with the EDM-DC properties, and as institutional data for manga volumes was shown to be more granular than the Work level, BIBFRAME was utilized to allow for a proper level of description. A method of describing serials using BIBFRAME while allowing it to work within an EDM-based model was also developed. As an RDF-based vocabulary and Linked Data model, BIBFRAME also enabled the closed "silos" of bibliographic data to be opened for use on the Semantic Web. Work level data, typically less granular and containing more generic bibliographic descriptive properties, was described using the core DC properties available with EDM. This provided a sufficient amount of granularity for data common at this level, while at the same time using a popular Web vocabulary, allowing for easier use of existing Web data for resources.

The final result, therefore, is a conceptual model that provides the basis for aggregating bibliographic manga data from different sources and making it available as Linked Data for use by any interested parties for tasks such as Semantic Web applications for manga, or for the addition of greater levels of descriptive detail by institutions that deal with bibliographic data for manga.

5.2 Limitations

While the work in this study produced aggregation and description models founded on other established models and vocabularies, a practical implementation of the final aggregation model was not investigated. Individual portions of the overall aggregation model are valid in their accordance with the rules and frameworks from which they are based, e.g. the

bibliographic description of manga volumes adhering to the rules of BIBFRAME model, mapping the use of BIBFRAME entities to EDM properties similar to Zapounidou et al. (2014a, 2014b), and the use of EDM in linking successive serial volumes based on the library metadata alignment report (Angjeli, A., et al. 2012). The possibility of the practical implementation of the sum of these parts, which make up the complete aggregation model, however, is unknown.

An automated method of matching related manga data for which to aggregate remains to be found. Data matching was done manually, and while the automation of matching some data was investigated using OpenRefine (see Section 4.3), a more reliable way to find related or identical resources is needed. If the available data existed in data stores accessible through a method such as a SPARQL endpoint, this would be feasible, though all of the data examined would need to be both transformed into and made available as Linked Data, as recommended in Section 5.3.

Validation of the model was not also addressed in this study. In its current conceptual stage, one can examine the total amount of bibliographic data being aggregated for a single resource and argue that its availability for use on the Semantic Web is a subjective improvement over the current situation, where the data is inaccessible. A method of technical validation for this, however, was not examined. A possibility is to perform technical validation on RDF data produced based on the model, though it is debatable that this would have significant meaning, as shown by Baker et al. (2014) discussing the problems with RDF validation.

As long as the various providers of bibliographic data are contributing data that differs from one another, then the granularity of data is improving. Over the course of the data analysis, however, some of the more interesting data was inaccessible with the methods used in the presented model. Conversations with librarians from academic libraries in the US revealed specific kinds of data that they would like to include in their records, such as summaries for specific volumes of manga. While this data is available, is it not accessible via the same Linked Data methods as other, more standard bibliographic data (see Section 3.2.4). This is representative of the lack of manga data on the Semantic Web, and lack of manga-centric authorities on the Web in general. The aggregating of data examined in this study, along with its transformation into Linked Data formats, would improve this situation, though making available the most unique data requires changes outside of the scope of this study, such as changes to how DBpedia makes certain Wikipedia data available, or the accessibility to descriptive data from manga fan sites like <http://mangaupdates.com>.

While the method of aggregation used in this study was based on EDM, alternative methods were not investigated. EDM is the foremost aggregation method (Peroni, Tomasi, &

Vitali, 2013), though its use has led to some issues based on the model being primarily aimed at serving the Europeana portal (see Section 4.2.2). Another possibility is to model the data according to more standard Linked Data practices and utilize RDF “named graphs” in order to identify and aggregate data from different providers, though the effectiveness of this remains to be investigated, as the named graph model lacks an effective way of distinguishing between graphs once published on the Web (Wikipedia, 2015).

5.3 Recommendations

As Fee has pointed out (2013), minutiae matters for fans of resources such as manga. It would be beneficial for users, then, for institutions to describe as much bibliographic detail that they can reasonably produce. While needs of the patron that an institution serves will always take precedence in deciding what level of data to describe, as more institutions make their data available as Linked Data, the amount of people interested in their data can increase greatly. It was somewhat surprising to find that no data outside that found in hobbyist resources catalogued data such as volume summaries or other similar data. Increasing granularity with aggregation is beneficial for patrons, but more institutions describing the actual content of manga would be of interest. For specialist institutions without format cataloguing guidelines to follow, a possibility is to choose less-orthodox properties, such as summaries or chapter titles, to record. For libraries that have more strict bibliographic description rules, the use of analytic cataloguing (see Section 2.2, 3.2.3, & Fee (2008)) is recommended.

Similarly, while the recording of more data is recommended, its publishing as Linked Data should be performed as well. The model presented in this study outlines how bibliographic data for manga can be described using Linked Data ready vocabularies, but the publishing of data has to be done by the parties responsible. Though publishing Linked Data is a recommendation that can be applied to almost any data provider, the differences in the data produced by institutions like libraries versus more hobbyist-oriented sources means that fans of manga have a lot to gain from the sharing of data between these different provider types, particularly if the prior recommendation comes to fruition and manga data providers being to record more interesting types of data. As Gonzales (2014) discussed, the combination of authority data and user-generated Web data creates a network of reliable, rich, and far-reaching resources that better fulfill the needs of users. Toppan’s data was published on the Web late in the course of this study, located at <http://mediaarts-db.jp>, but it is not available in any Linked Data format. Had it been, it would serve the Web and Linked Data communities,

both as a source of general bibliographic data for manga, and a strong manga authority, of which few exist.

5.4 Future Work

As EDM made up an important part of the aggregation model presented in this study, the implementation of a Europeana-like “portal” for manga would be interesting. This would resolve with some issues that arise when using EDM solely as a model for aggregation rather than a harvesting method, and would also provide an interesting source of manga data on the Web, both as the home to the data examined in this study, and to future manga data sourced from other data providers. Had Toppan’s Web database at <http://mediaarts-db.jp> been published as Linked Data, this could have perhaps served as the starting point for which aggregated manga data could be brought together. Still, the establishment of a manga metadata portal on the Web could provide a home to manga metadata on the Web for any granularity level, from fan-generated volume reviews, to institution-produced authority data. As mentioned in Angjeli et al. (2012), the use of the property `edm:isNextInSequence` in the model would also enable the sequential browsing of a resource within the portal, establishing functionality beyond simply searching through resources.

Regarding granularity, depending on the type of bibliographic data that can be produced and made available as Linked Data, there may be an opportunity, or need, for the creation of a manga-focused ontology. For the types manga bibliographic data currently produced by institutions, DC and BIBFRAME were shown to be successful at accurately describing the information. Depending on how granular the data becomes, however, these general vocabularies may be inadequate. Petiya (2014) developed an interesting ontology aimed primarily at, but not limited to, comic books, featuring properties such as “character” and “storyArc” to describe fairly specific content within comic books. The importance of minutiae mentioned throughout this thesis means that properties used to describe similar contents of a manga would be beneficial as well. Since Petiya focused mainly on comic books, analyzing how the ontology works with manga in practice would reveal whether it is suitable, or if the need for a manga-specific ontology would be of use.

As BIBFRAME is still in development, its true use both inside and outside of libraries should be investigated once the model and vocabulary mature further. Much of the discussion and testing surrounding BIBFRAME is, understandably, focused on libraries. BIBFRAME Profiles, discussed in Section 3.3.2, are document(s) that set local cataloguing guidelines within the context of “functional requirements, domain models, guidelines on syntax and usage, and possibly data formats” and “contain the description, statements and constraints specific to

a community”(Library of Congress, 2014b). While the recommendation suggests the investigation of a manga-focused ontology, it is worth examining whether the creation of a BIBFRAME Profile focused on manga would be valuable, either to set manga bibliographic description guidelines to data creators both in and outside of libraries, or aimed specifically at libraries as a way to encourage analytic cataloguing of manga resources.

5.5 Conclusion

This study presented in this thesis resulted in the creation of a conceptual model meant to aggregate bibliographic data for manga, of different entity levels, and from a variety of data providers, with the goal of collecting unique, granular manga information, and enabling its use on the Semantic Web as Linked Data. The model sought to address the “bibliographic data gap” between different institution types, namely sources that provide authoritative data such as libraries, and more specialist or hobbyist sources that provide more granular data. The methods for aggregating data, expressing the successive relationships between serialized manga, and description at the conceptual Work level were based on the Europeana Data Model, while the new BIBFRAME model and vocabulary were utilized to describe more granular data at the Item level, which focuses on individual volumes of manga. The bibliographic description methods were shown to accurately describe the amount of granularity for data available at their respective levels, while the aggregation method brought together data from multiple sources to describe a single resource from different viewpoints and languages.

The popularity of manga, both recreationally and academic, is constantly increasing. The amount of data available for this unique resource, particular at the academic and institutional levels, has not kept pace. To better serve all types of manga fans, improved descriptive data is required, but in a time of shrinking library budgets, one cannot expect traditional bibliographic data providers to handle this task unaided. The sharing and reuse of bibliographic data for manga is therefore invaluable in improving the landscape for fans of the medium, both through the contribution of decades of cataloguing experience and authoritative data from professional institutions, and through the enthusiasm and knowledge of fans and specialists.

ACKNOWLEDGEMENTS

A special thank you to my advisors, Dr. Shigeo Sugimoto and Dr. Mitsuharu Nagamori for their advice and direction over the course of my degree. Their assistance, both with my studies and with extra-curricular matters, has been invaluable. I am also indebted to all of the members of the various graduate labs at Tsukuba University who have listened to proposals and updates and given their excellent feedback throughout my study. A special thanks to both Toppan Printing Co. and Tadgh Dinnage from Monash's Manga Library for allowing the use of their data in this research. And finally, to my family and friends, both in Japan and abroad, for their understanding, encouragement, and support over the past two years.

REFERENCES

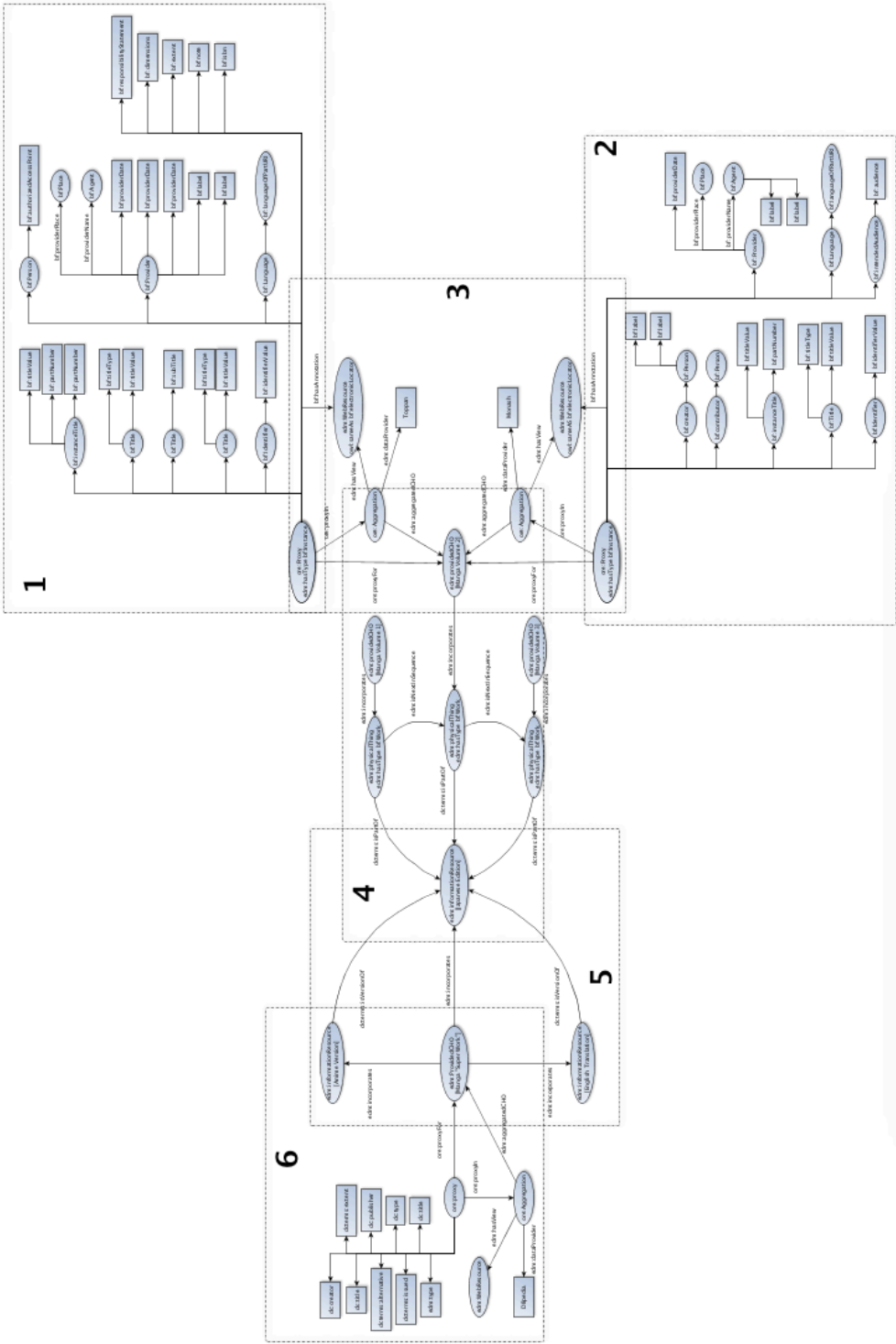
- Agenjo, X., Hernández, F., & Viedma, A. (2012). Data Aggregation and Dissemination of Authority Records through Linked Open Data in a European Context. *Cataloging & Classification Quarterly*, 50(8), 803–829. doi:10.1080/01639374.2012.711441
- Angjeli, A., et al. (2012, December 19). *D5.1 Report on the alignment of library metadata with the Europeana Data Model (EDM)*. Retrieved from http://www.theeuropeanlibrary.org/confluence/download/attachments/12091395/D5.1_EDM_for_libraries_v2.0.pdf
- Baker, T., Coyle, K., & Petiya, S. (2014). Multi-entity models of resource description in the Semantic Web. *Library Hi Tech*, 32(4), 562–582. doi:10.1108/lht-08-2014-0081
- Berners-Lee, T. (2009). Linked Data. Retrieved from <http://www.w3.org/DesignIssues/LinkedData.html>
- Cole, T. W., Han, M.-J., Weathers, W. F., & Joyner, E. (2013). Library Marc Records Into Linked Open Data: Challenges and Opportunities. *Journal of Library Metadata*, 13(2-3), 163–196. doi:10.1080/19386389.2013.826074
- Coyle, K. (2004). Future considerations: the functional library systems record. *Library Hi Tech*, 22(2), 166–174. doi:10.1108/07378830410524594
- Dilevko, J., & Gottlieb, L. (2004). Selection and Cataloging of Adult Pornography Web Sites for Academic Libraries. *The Journal of Academic Librarianship*, 30(1), 36–50. doi:10.1016/j.jal.2003.11.006
- Europeana. (2013). *Definition of the Europeana Data Model, version 5.2.4*. Retrieved from <http://pro.europeana.eu/edm-documentation>
- Europeana. (2014a, November 18). EDM Factsheet. *Europeana Data Model Documentation*. Retrieved from <http://pro.europeana.eu/share-your-data/data-guidelines/edm-documentation>
- Europeana. (2014b). Data structure. *Europeana Labs*. Retrieved from <http://labs.europeana.eu/api/linked-open-data/data-structure/>
- Fallgren, N., Lauruhn, M., Reynolds, R. R., & Kaplan, L. (2014). The Missing Link: The Evolving Current State of Linked Data for Serials. *The Serials Librarian*, 66(1-4), 123–138. doi:10.1080/0361526x.2014.879690
- Fee, W. T. (2008). Do You Have Any Ditko?: *Serials Review*, 34(3), 175–189. doi:10.1016/j.serrev.2008.06.003
- Fee, W. T. (2013). Where Is the Justice... League?: Graphic Novel Cataloging and Classification. *Serials Review*, 39(1), 37–46. doi:10.1016/j.serrev.2013.02.004
- Gonzales, B. M. (2014). Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record. *ITAL*, 33(4). doi:10.6017/ital.v33i4.5631
- Gruber, T. (2008). Collective knowledge systems: Where the Social Web meets the Semantic

- Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(1), 4–13. doi:10.1016/j.websem.2007.11.011
- He, W., Mihara, T., Nagamori, M., & Sugimoto, S. (2013). Identification of works of manga using LOD resources. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries - JCDL '13*. doi:10.1145/2467696.2467731
- Hooland, S., & Verborgh, R. (2014). *Linked data for libraries, archives and museums: How to clean, link and publish your metadata*. London: Facet Publishing.
- IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). Functional requirements for bibliographic records. *IFLA Series on Bibliographic Control*, 19. Retrieved from <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>
- Isaac, A. (2013). Europeana data model primer. Retrieved from http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf
- Jiji. (2014). Manga library opens at Peking University. Retrieved from <http://www.japantimes.co.jp/news/2014/11/29/national/manga-library-opens-peking-university/>
- Konstantinou, N., Houssos, N., & Manta, A. (2014). Exposing Bibliographic Information as Linked Open Data Using Standards-based Mappings: Methodology and Results. *Procedia - Social and Behavioral Sciences*, 147, 260–267. doi:10.1016/j.sbspro.2014.07.169
- Kroeger, A. (2013). The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future. *Cataloging & Classification Quarterly*, 51(8), 873–890. doi:10.1080/01639374.2013.823584
- Library of Congress (2014a, April 28). *BIBFRAME Authorities*. Retrieved from <http://www.loc.gov/bibframe/docs/bibframe-authorities.html>
- Library of Congress (2014b, May 5). *BIBFRAME Profiles: Introduction and Specification*. Retrieved from <http://www.loc.gov/bibframe/docs/bibframe-profiles.html>
- Library of Congress (2015a). *BIBFRAME Frequently Asked Questions*. Retrieved from <http://www.loc.gov/bibframe/faqs/>
- Library of Congress (n.d.). *Overview of the BIBFRAME Model*. Retrieved from <http://www.loc.gov/bibframe/docs/model.html>
- Markham, G. W. (2009). Cataloging the Publications of Dark Horse Comics: One Publisher in an Academic Catalog. *The Journal of Academic Librarianship*, 35(2), 162–169. doi:10.1016/j.acalib.2009.01.008
- Miller, E., Ogbuji, U., Mueller, V., & MacDougall, K. (2012). Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services. In *Washington, DC: Library of Congress*.

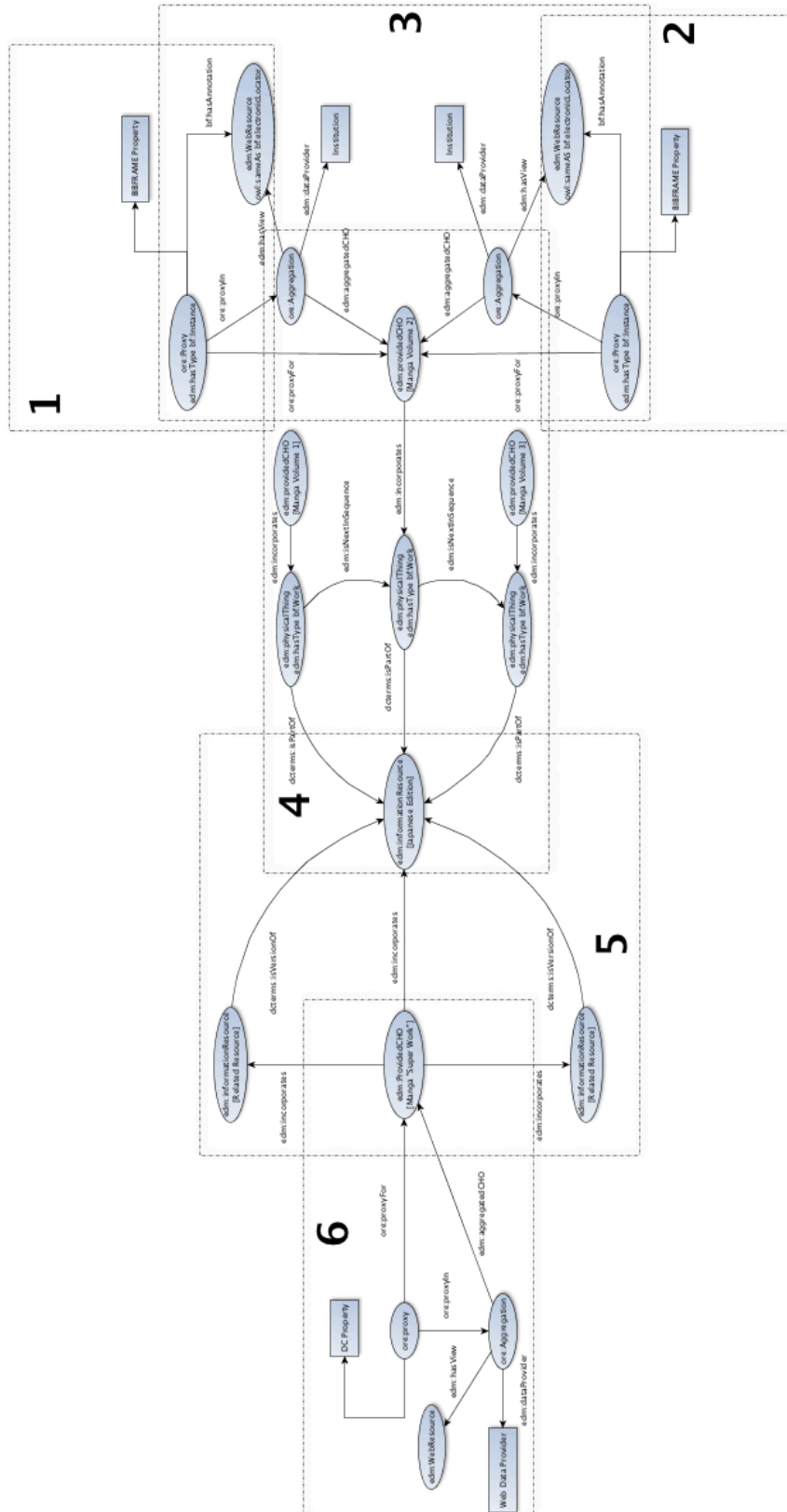
- Moulaison, H. L., & Million, A. J. (2014). The Disruptive Qualities of Linked Data in the Library Environment: Analysis and Recommendations. *Cataloging & Classification Quarterly*, 52(4), 367–387. doi:10.1080/01639374.2014.880981
- Morozumi, A., Nomura, S., Nagamori, M., & Sugimoto, S. (2009). Metadata framework for manga: A multi-paradigm metadata description framework for digital comics. In *Proceedings from International Conference on Dublin Core and Metadata Applications* (p. 61-70). Seoul, KR: Dublin Core Metadata Initiative.
- O’Nale, R. (2010). Manga. In *Encyclopedia of Comic Books and Graphic Novels*. (Vol. 2, p. 378-387). Santa Barbara, CA: Greenwood Press.
- OpenRefine.org. (n.d.). *OpenRefine: A free, open source tool for working with messy data*. Retrieved from <http://openrefine.org/>
- Peroni, S., Tomasi, F., & Vitali, F. (2013). The aggregation of heterogeneous metadata in web based cultural heritage collections: a case study. *International Journal of Web Engineering and Technology*, 8(4), 412. doi:10.1504/ijwet.2013.059107
- Petiya, S. (2014). *Building a Semantic Web of Comics: Publishing Linked Data in HTML/RDFa Using a Comic Book Ontology and Metadata Application Profiles*. (Electronic Thesis or Dissertation). Retrieved from <https://etd.ohiolink.edu/>
- Southwick, S. B. (2015). A Guide for Transforming Digital Collections Metadata into Linked Data Using Open Source Technologies. *Journal of Library Metadata*, 15(1), 1–35. doi:10.1080/19386389.2015.1007009
- Tennant, R. (2002). MARC Must Die!. *Library Journal*, 127 (17), 26–28.
- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2014a). Integrating library and cultural heritage data models. Proceedings of the 18th Panhellenic Conference on Informatics - PCI '14. doi:10.1145/2645791.2645805
- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2014b). Library Data Integration: Towards BIBFRAME Mapping to EDM. *Communications in Computer and Information Science*, 262–273. doi:10.1007/978-3-319-13674-5_25
- W3C (2015a). *Semantic Web*. Retrieved from <http://www.w3.org/standards/semanticweb/>
- W3C (2015b). *Linked Data*. Retrieved from <http://www.w3.org/standards/semanticweb/data/>
- Wikipedia. (2015, February 4). *Named Graph*. Retrieved from http://en.wikipedia.org/wiki/Named_graph

APPENDICES

A. Full Aggregation Model

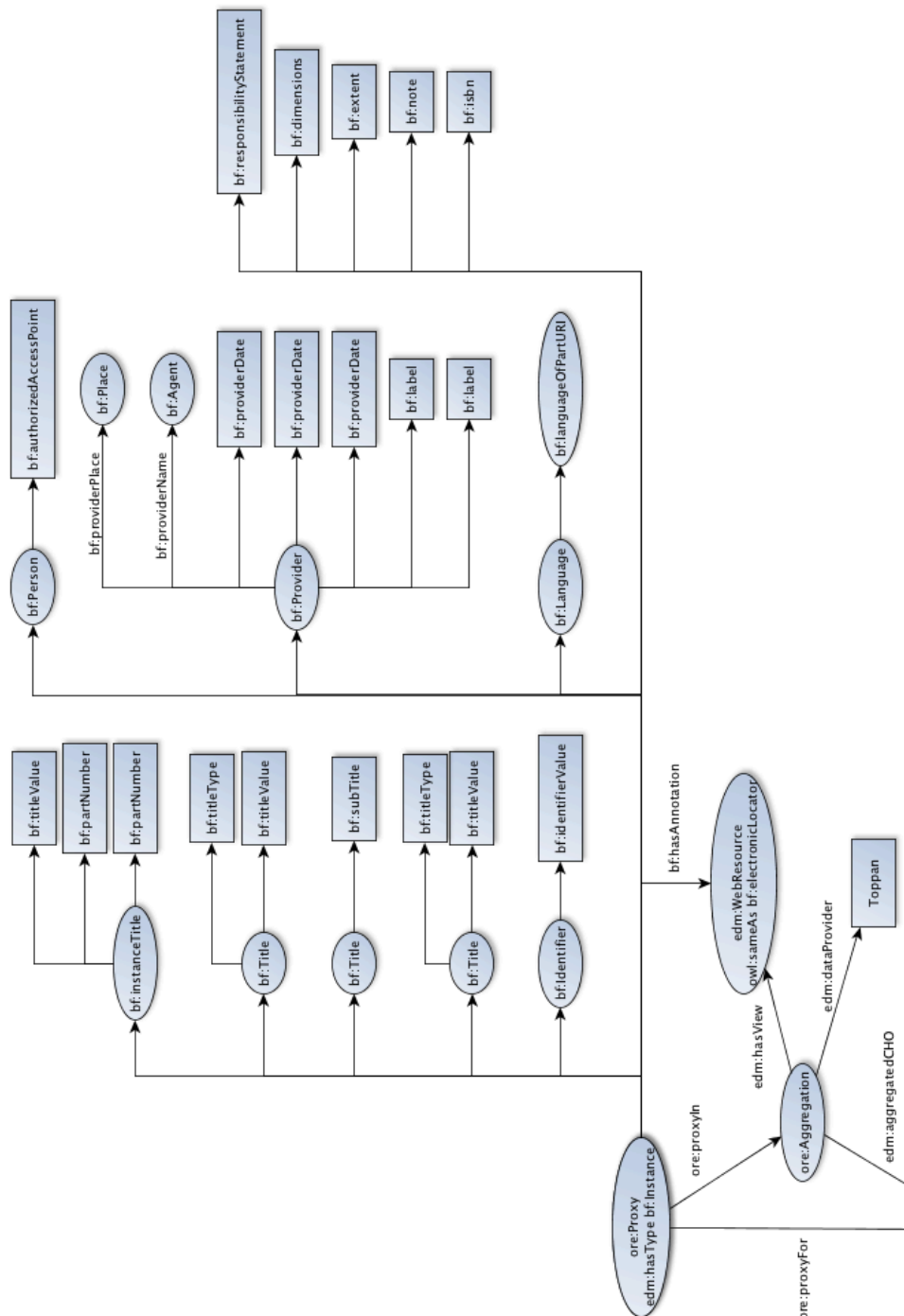


B. Full Aggregation Model - Generalized



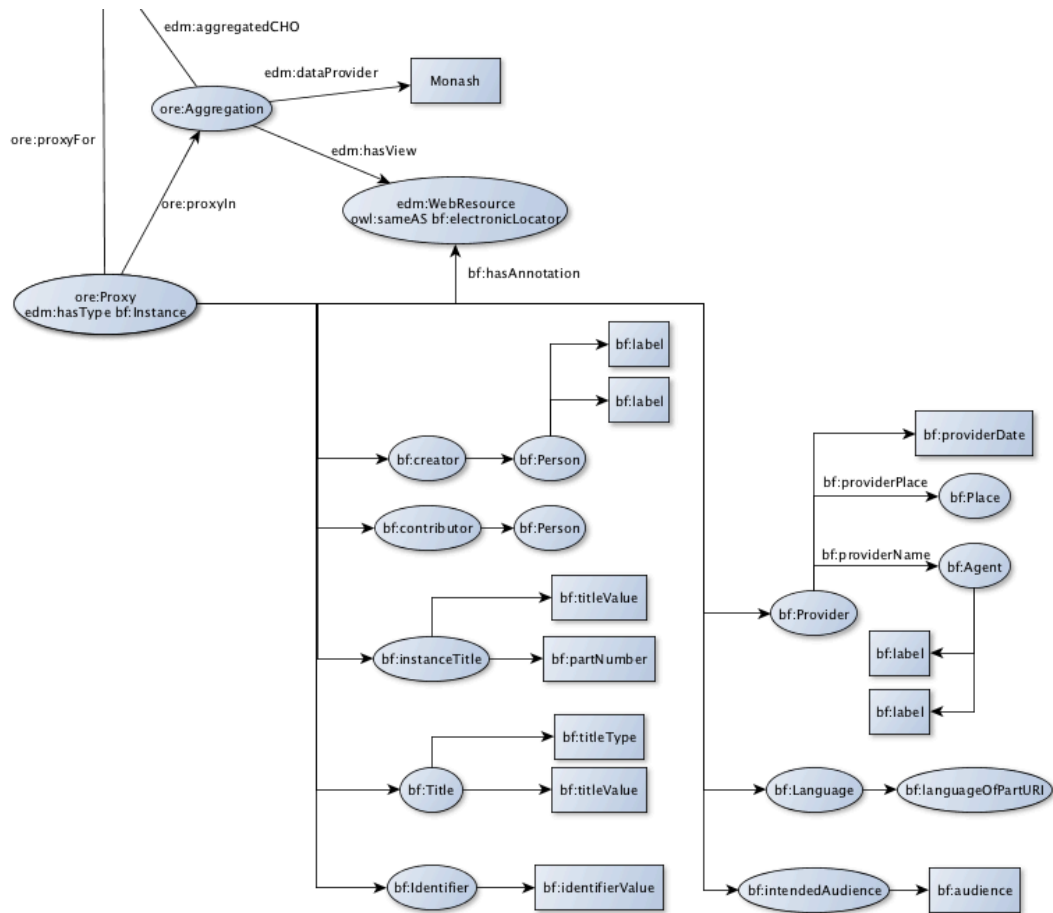
Appendices A and B show the aggregation model in its entirety, with Appendix A having specific properties from data providers used in this study included in the model, and Appendix B showing the generalized version without specific data. Both have the same structure and grouping numbers, which refer to the subsequent appendices. Group 1 and 2 represent data sourced from two providers. In the case of this study, Toppan and Monash, and in the case of the generalized model in Appendix B, two unnamed institutions. These groups are describing the manga at the Item, or volume level, and so use BIBFRAME properties as the method of description. Group 3 is the portion of the model that aggregates data at the Item or volume level. Group 4 shows how the successive relationship of manga (volume-to-volume) is represented within the model. Group 5 shows how the central Work level resource for the manga series is represented, and how other resource types may be included and connected. In Appendix A, the examples given are an English translation of the original Japanese manga, and an anime based on the manga. Finally, Group 6 is similar to Groups 1 and 2 in that it represents descriptive data for the manga, but this is attached to the higher Work level, and utilizes the Dublin Core vocabulary rather than BIBFRAME.

C. Aggregation Model – Group 1



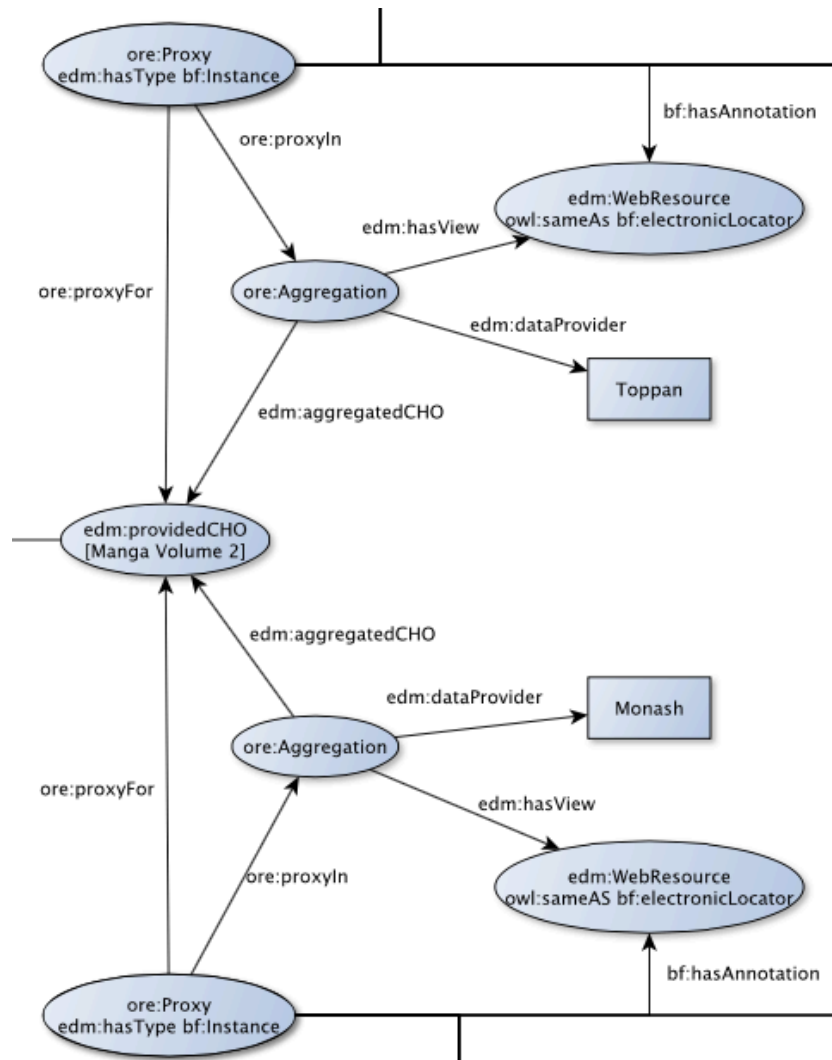
Group 1 represents data from one institution, in this case Toppan, for a single manga volume. BIBFRAME is used here as the bibliographic description vocabulary to accommodate the more granular levels of detail that are required at the volume or Item level. The BIBFRAME properties used here can be seen in Table 3.

D. Aggregation Model – Group 2



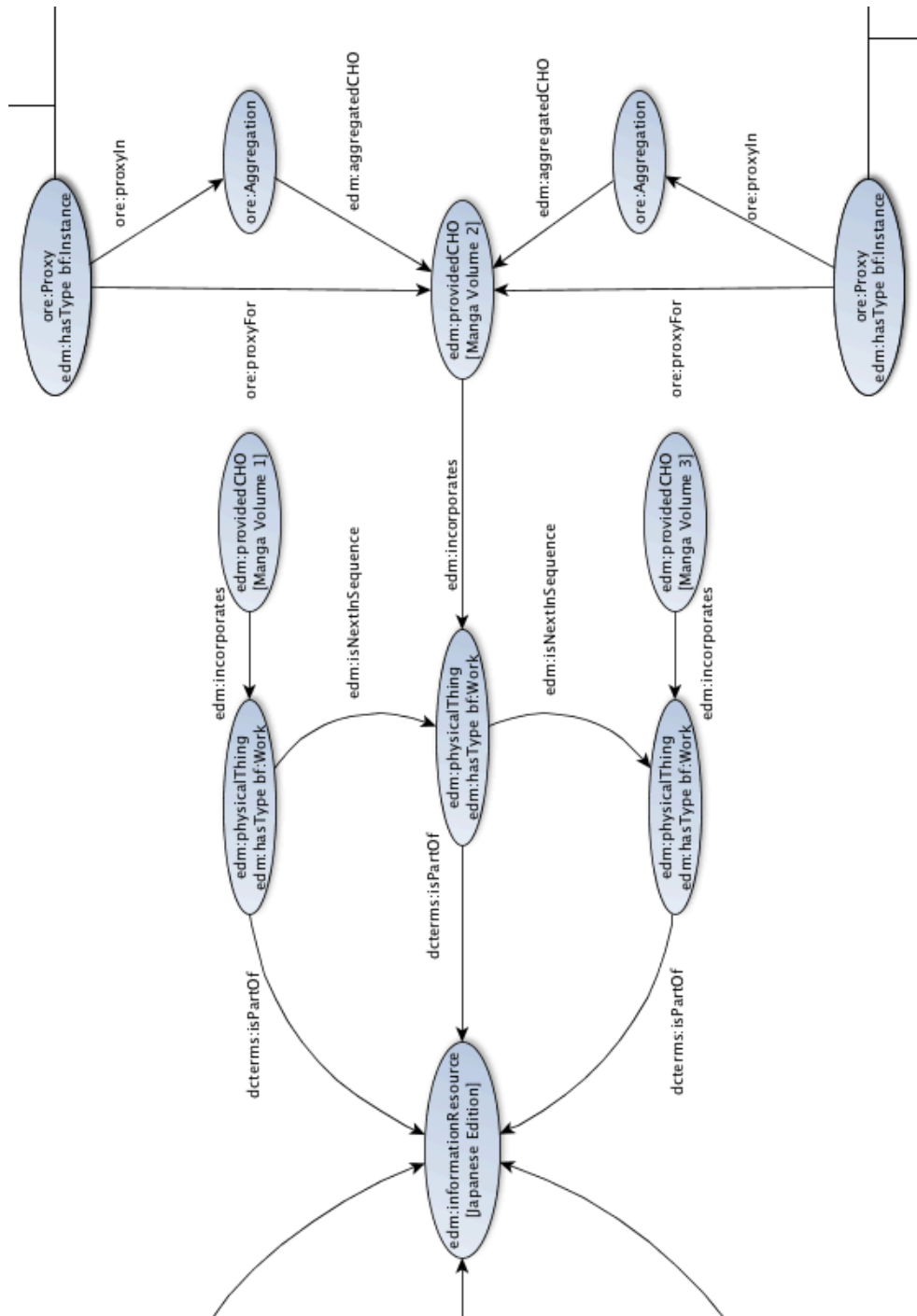
Group 2 is similar to Group 1 in structure, but the data in this example is sourced from Monash rather than Toppan. Table 4 contains the properties used in this Group.

E. Aggregation Model – Group 3



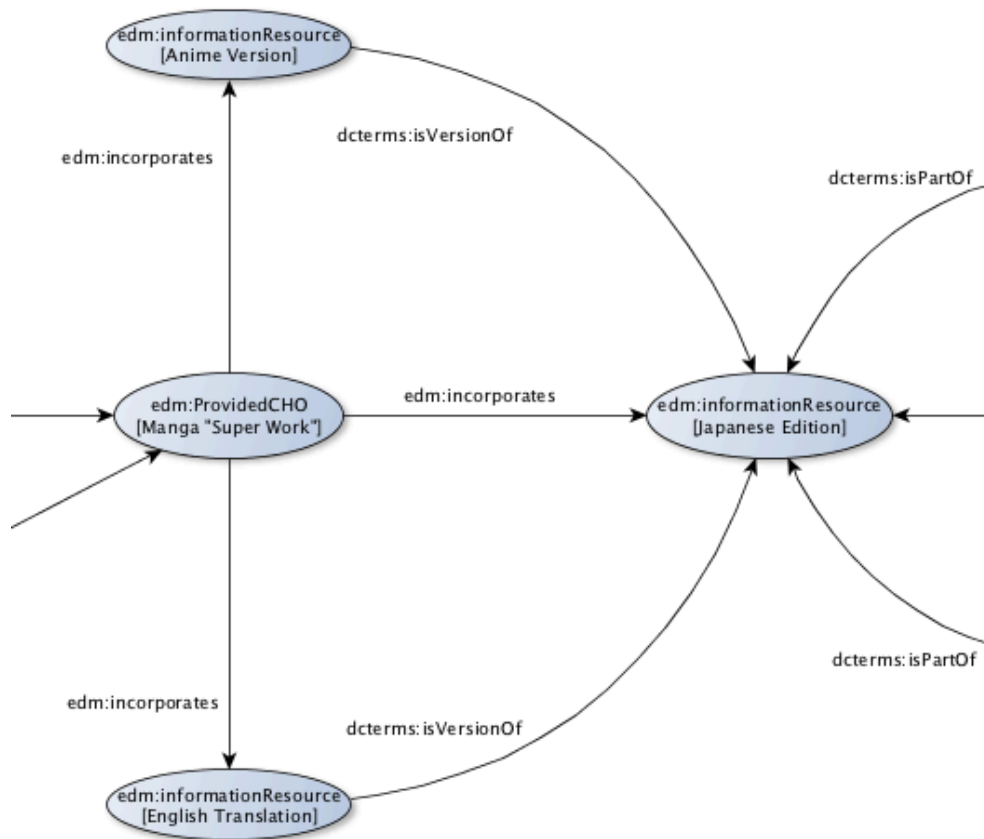
Group 3 illustrates how data from different institutions is aggregated according to the model. Each institutions data is attached to a unique **ore:Proxy** property and is connected to the manga volume, represented by the **edm:ProvidedCHO**, along with the **ore:Aggregation** and **edm:WebResource** properties.

F. Aggregation Model – Group 4



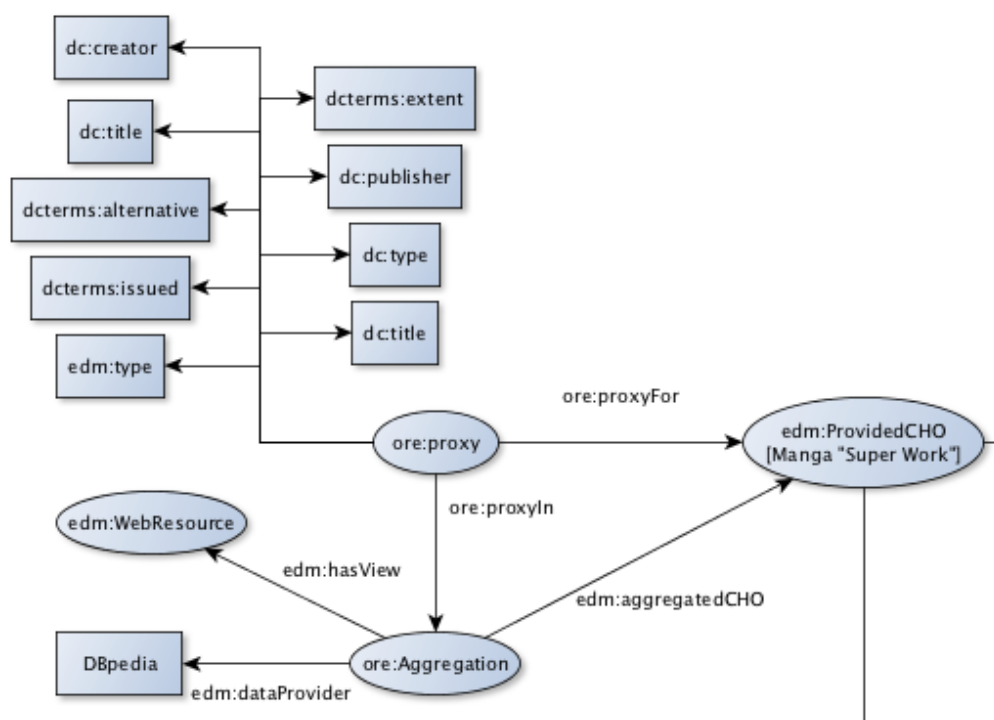
Group 4 demonstrates the use of the `edm:isNextInSequence` and `dcterms:isPartOf` properties in order to allow for the modeling of two manga relationship types – the successive, volume-to-volume relationship, connecting volumes to one another in their proper order, and the Item to Work relationship, which connects individual volumes to the conceptual Work they are a part of.

G. Aggregation Model – Group 5



Group 5 features the **edm:ProvidedCHO** that represents the conceptual manga Work level. As this represents not only the original manga itself but also other media forms of the same series, the use of example English Translation and Anime Version resources are included to illustrate how these other resources may be connected to the broad Work level.

H. Aggregation Model – Group 6



Group 6 is similar to Groups 1 and 2, but describes the manga at the Work level rather than the Item or volume level. The description is also performed using Dublin Core rather than BIBFRAME.