

A Study on Improving Bibliographic Descriptions for Objects
of Popular Culture Through Multimedia Franchise
Representation, Hierarchical Modeling, and Metadata
Aggregation

SENAN KIRYAKOS

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

February 2019

A Study on Improving Bibliographic Descriptions for Objects of Popular Culture Through Multimedia Franchise Representation, Hierarchical Modeling, and Metadata Aggregation

Abstract

Advances in technology that have created new easily accessible and popular forms of media, combined with commercial and creative incentives to develop large singular universes consisting of connected and shared narratives, characters, and worlds across these mediums, has resulted in a significant growth in the phenomenon of multimedia franchises. These are collections of related and derivative creative works, often with a single progenitor, that are connected to one another in a variety of ways. Though this has resulted in vast engaging fictional worlds spread across a number of media types for individual users and communities to explore and consume, this cross-media or transmedia phenomenon has also presented several issues for creators, fans, and those responsible for bibliographic description seeking to adequately meet the unique needs of the users of these resources. These issues include an insufficient representation of resources compared to the way they are viewed and understood by users, low levels of descriptive granularity with regards to bibliographic data and the vocabularies used for descriptions, and an overall lack of the sophisticated descriptive and relational semantics which these resources require in order to sufficiently meet the needs of users for tasks such as search and retrieval. Because many resource types within these multimedia franchises, such as manga, anime, and video games, have traditionally been seen as niche by agents normally responsible for the creation of descriptive data like libraries and other institutions, some of these issues appear when attempting to apply the paradigms of existing bibliographic concepts, models, and vocabularies to these mediums. The comparatively poor treatment of certain multimedia object types has resulted in countless hobbyists taking it upon themselves to portray and describe these resources to a level which more adequately meet their needs. Their efforts can be seen on the web as fan pages, dedicated franchise wikis, or Wikipedia itself, all of which demonstrate a surprisingly thorough and complex understanding of these resources by non-institutional agents such as casual fans or collectors. As these web resources are essentially crowdsourced pools of bibliographic data, the level of descriptive granularity exceeds any amount that a traditional institution could hope to achieve. Because the data and the structure of the data is created by users, however, it can also be used as guidance for institutions to better model and describe their data for the benefit of their respective audiences, or even as a way to attempt to align their data with this hobbyist data, enabling the inclusion of useful data which institutions have produced, such as bibliographic authority files created by libraries. In an attempt to explore this possibility, this dissertation presents research conducted as an exploration and experimentation of the role of user-created hobbyist data on the web and how it can be used to help in the creation of new entities, models, and paradigms for the purpose of improving the bibliographic description of multimedia objects.

To achieve this, this dissertation presents the culmination of several past research studies conducted in an attempt to address the various concerns named thus far. These studies have focused on

distinct yet related areas of research that have mainly examined the portrayal of popular culture entities in bibliographic models and on the web, and how to improve upon description of these entities in ways that address user needs. Based on an examination of bibliographic data for various multimedia objects sourced from both traditional institutional and newer hobbyist data providers, the research explores the creation of a franchise-level bibliographic entity called the Superwork. A bibliographic entity that represent the concept of an entire franchise is not clearly supported by existing bibliographic models, and the creation of such an entity both conforms to the way in which users collocate data themselves on the web and enables the establishment of a variety of resources and relationships that are otherwise difficult to describe. The research then outlines the creation of a new bibliographic hierarchy, which includes the Superwork entity, that is meant to better conform to the nature of multimedia works, extending from the overall franchise concept down to the singular object. Again using hobbyist providers and the ways in which they portray multimedia objects as an exemplar, four core entities – the Superwork, Series, Work, and Object – are used to create a bibliographic hierarchy that divides franchises and their objects into logical levels of conceptuality that are able to portray and describe multimedia objects according to a current understanding of descriptive data for these resources. Following this, a metadata aggregation model using the Open Archives Initiative Object Reuse and Exchange (OAI-ORE) as its basis is used to build the aforementioned core entities, connect them to establish the defined hierarchy, provide descriptive bibliographic data through metadata aggregation of various providers, and enable a variety of semantic relationships between entities. The simplicity and openness of the model is demonstrated as a key enabler in support of the evolving, dynamic resources that form the core of this study, and as a way to support as many community and audience needs as possible, be they hobbyist or institutional in nature.

In addition to the presentation of conducted studies and their results, this dissertation expands upon broad theoretical topics that guided each research area and were central to their goals and future work considerations. These topics include the importance of enabling relationships between related resources and the implications this has for bibliographic description and meeting user needs, and the role of user-generated content on the web, such as Wikipedia, in bibliographic description and as examples of collective knowledge. These topics are not only central to many of the ideas of the research presented here but have broad implications in other related research areas as well, both within and outside of the field of popular culture. In conjunction with three main research areas outlined previously, these topics are attempts to answer broad questions about how data for certain resources, such as manga, anime, and video games, can be better represented for the benefits of multiple user groups, and this dissertation presents the results, discussions, and recommendations that the research and discussions that have surrounded the research has produced. With this dissertation, the author aims to present the lessons learned in their attempts to improve bibliographic description of multimedia objects, such as the importance of acknowledging the sophisticated concepts and constructs that users are capable of contributing to the areas of bibliographic modeling and description, and the need for bibliographic

models that are intended for use with niche resources to be as open and inclusive as possible so as to not restrict the levels of semantic information that can be supported. Through the context of pop culture materials and multimedia franchises, the author hopes that this dissertation will move others towards thinking about bibliographic modeling and description in new ways, while also contributing novel concepts to these areas.

Table of Contents

Abstract	II
Table of Contents	V
List of Figures	VII
List of Tables	IX
Chapter 1: Introduction	1
Research Novelty.....	3
Relevant Terminology.....	4
<i>Pop culture materials and MAG objects</i>	4
<i>Multimedia franchise, transmedia, and Superwork</i>	5
Methodology & Structure of this Dissertation.....	7
Chapter 2: Related Literature	9
Superwork-like Entities and Hierarchical Representations.....	9
Aggregations and the OAI-ORE Model.....	12
Chapter 3: The Portrayal of Bibliographic Data for MAG Objects	15
Bibliographic Data for MAG Objects as Portrayed by Libraries and Similar Institutions.....	15
<i>The Media Arts Database</i>	17
Bibliographic Data for MAG Objects as Portrayed by Wikis and Other Fan-sites.....	21
Chapter 4: The Representation of Multimedia Franchises using the Superwork Entity	27
Encountering the Need for a Franchise-level Bibliographic Entity.....	27
Defining the Superwork as Distinct from Related Entities in Existing Models.....	28
<i>Superwork and the FRBR Work entity</i>	28
<i>Superwork and the FRBR_{OO} Complex Work Entity</i>	29
<i>Superwork and Subjects</i>	32
Existing Web Entities Resembling Superwork-like Concepts	33
Applicable Mediums and Boundaries of the Superwork Entity.....	37
Chapter 5: The Creation of a Bibliographic Hierarchy for MAG Objects	38
A Bibliographic Hierarchy for MAG Objects.....	38
<i>Entity boundary intentions and hierarchy summary</i>	45
<i>Relationships between core entities</i>	48
<i>Intra-layer relationships between core entities</i>	48
Existing Semantic Relationships Found among Hobbyist Data Providers	50
Chapter 6: The Aggregation of Descriptive Metadata from Institutional and Hobbyist Data Providers	57
Overview of the OAI-ORE Aggregation Model	57
<i>Core entities within OAI-ORE</i>	57
<i>Nesting aggregations within OAI-ORE</i>	59

<i>The role of OAI-ORE's Proxy resource</i>	60
Determining Bibliographic Entities Based on Web Data.....	61
<i>Identification of related resources across data providers</i>	62
Portraying Existing Web Resources as a Bibliographic Hierarchy.....	63
Providing Increased Granularity for Bibliographic Descriptions using Metadata	
Aggregation.....	64
Supporting Semantic Relationships Described by Hobbyist Providers.....	65
<i>Inter-franchise relationships</i>	68
The Role of Authoritative Data from Institutions.....	69
Summarizing the Use of the OAI-ORE Model.....	70
Chapter 7: Discussion on Findings, Implications, Limitations, and Challenges	71
Lessons Learned Throughout this Dissertation and its Related Research.....	71
<i>Distinguishing between the carrier and the content</i>	71
<i>The importance of user behavior</i>	71
<i>Supporting flexibility within models</i>	74
Implications and Limitations of the Research and Results.....	74
<i>Implications and limitations for the Superwork entity</i>	75
<i>Implications and limitations for the bibliographic hierarchy</i>	75
<i>Implications and limitations in the use of the OAI-ORE model</i>	76
<i>Methodology limitations</i>	77
Other Encountered Issues.....	78
Chapter 8: Conclusions and Future Research Opportunities	80
Summarizing the Research Findings Presented in this Dissertation.....	80
Possibilities for Future Research	81
Acknowledgements	84
References	85
List of Publications	91
Appendices	92

List of Figures

Figure 3.1. Screenshot of the MADB entry for the <i>One Piece</i> manga series.	18
Figure 3.2. Screenshot of the MADB entry for the <i>Ghost in the Shell</i> anime series.	19
Figure 3.3 Screenshot of the MADB entry for first volume of the <i>One Piece</i> manga.....	19
Figure 3.4 Reproduced figure from Tallerås et al., (2018) showing a user-creating central node.....	20
Figure 3.5. AniDB page for the <i>One Piece</i> anime illustrating the multilingual data available for the title property.....	24
Figure 3.6. A sample of websites that form the ecosystem of various types of hobby data providers found on the web.	25
Figure 4.1. Euler diagram showing a simplified relationship between the conceptual Superwork, Complex Work, and Work entities.....	31
Figure 4.2 Abstract overview of entities and relationships between resources mentioned in the <i>Ghost in the Shell</i> franchise Wikipedia article.....	34
Figure 4.3 Reproduced figure from Tallerås et al., (2018) showing a central node with indirect relationships...35	35
Figure 4.4. Excerpt of Figure 4.2 showing relevant entities and relationships mentioned in the <i>Ghost in the Shell</i> (1995 film) Wikipedia article.....	36
Figure 5.1. A previous iteration of the manga-centric bibliographic hierarchy.....	38
Figure 5.2. A generalized hierarchy for multimedia objects showing different entity levels required for adequate portrayal, based primarily on descriptive data from the web.....	39
Figure 5.3. Property inheritance illustration for hierarchical entities.....	40
Figure 5.4. The hierarchical model with entities from the <i>Gundam</i> franchise.....	42
Figure 5.5. A hierarchy of a portion of the <i>Metal Gear</i> franchise illustrating the use of the Series entity to denote canonicity.....	42
Figure 5.6. Hierarchy demonstrating a single entry in the <i>Final Fantasy</i> Mainline series, <i>Final Fantasy VII</i> , as its own Series entity.....	43
Figure 5.7. The <i>Final Fantasy</i> hierarchy from Figure 5.6, with bolded entities illustrating a Series-to-Series relationship	49
Figure 5.8. Sequential Object entities illustrated (bold), based on the <i>Gundam</i> franchise hierarchy shown in Figure 5.4.....	50
Figure 5.9. Infobox screen capture for the <i>Ghost in the Shell</i> Wikipedia franchise article, listing various media instances.....	52
Figure 5.10. Screen captures of the related work sections for the <i>Ghost in the Shell</i> manga and anime pages from My Anime List.....	53
Figure 5.11. Screen capture of the AniDB entry for the 1995 <i>Ghost in the Shell</i> film showing Directly and Indirectly Related Anime	54
Figure 5.12. Screen capture of the AniDB Relation Graph for <i>Ghost in the Shell</i> listing various entities and their relationship type.	56
Figure 6.1. The three core entities of the OAI-ORE aggregation model (Resource map, Aggregation, and Aggregated Resource) and their relational properties	57

Figure 6.2. The role of Aggregations (a# entities) and Aggregated Resources (ar# entities) in the portrayal of the bibliographic hierarchy	59
Figure 6.3. A depiction of nested aggregations as part of the One Piece hierarchy	60
Figure 6.4. Use of the Proxy resource and a sample <i>manga:nextVolume</i> property to assert a sequential relationship between two Aggregation entities	61
Figure 6.5 The formation of a bibliographic hierarchy for a portion of the Gundam franchise, with entities formed based on the descriptions of various data providers	63
Figure 6.6. Bibliographic data available from a single source (top) along with an aggregated second source (bottom) for the <i>One Piece</i> manga Work entity	65
Figure 6.7. A franchise hierarchy with sample relational properties, both vertical and horizontal, to enable increase semantic descriptions and relationships between resources	67
Figure 6.8. A depiction showing relevant relationships made between different franchise hierarchies, both of which feature the <i>Dynasty Warriors: Gundam</i> video game	68
Figure 7.1. Paragraph describing the One Piece manga on the main franchise page, with links to more granular entries	72
Figure 7.2. The more granular layer which lists single volumes in the One Piece manga	73
Figure 7.3. The most granular layer describing single manga volumes, including chapter lists	73

List of Tables

Table 3.1. Manga page styles and their unique properties within the MADB	21
Table 3.2. Granular properties available from various data providers.....	24
Table 3.3. Hobbyist data providers displayed in Figure 3.5.....	26
Table 3.4. General summary of data provider types and their attributes	26
Table 6.1 Property types requiring analysis in order to determine entity being described (column 2) and identical resources (column 3)	62

Chapter 1: Introduction

Increasingly, previously niche forms of media such as manga, anime, and video games have grown in production, consumption, and cultural acceptance and appreciation, and now make up a significant sector of modern popular culture (pop culture) (e.g. see Asahi Shimbun, 2017; ESA, 2016; Pineda, 2017; Reid, 2016). Due to the relatively young age of these mediums, their uniqueness when compared to traditional objects of bibliographic description such as monographs or serials, and the late acceptance by traditional memory institutions in acknowledging them as worthy of sincere effort with regards to their collection, description, and preservation, the state of data availability and quality for these resources varies and is, in many cases, poor. The data situation outside of institutions like libraries and museums, however, is quite different. Particularly on the web, one can find data for resources like manga, anime, and video games in both a greater quantity and granularity than could be expected by institutions traditionally responsible for the creation and maintenance of bibliographic data. Created mainly by fans of the mediums, this web data is better able to address the needs of most users, be they collectors, fans, researchers, or others who may want access to semantically rich information.

Apart from objects such as manga, anime, and video games being unique objects in isolation, another feature common among them is that they are often a part of a network of related creative works – referred to as a multimedia franchise – that share contents such as worlds, characters, concepts, etc. across a number of different media formats. Because there are ever increasing commercial incentives that come from the generation of these relationship networks (see Langford, 2010, p. 207) this means that “every successful new film or videogame is the potential progenitor of a vast cross-media franchise” (Lemke, 2004, p. 4). Thus, successful pieces of media will now often have other media format adaptations or may even continue stories or tell related ones in alternative mediums, encouraging users to seek out and consume these as well for a more coherent understanding of the world, story, or characters in question. This means that to better serve users that may be interested in seeking out these other works or better understanding a single work in the context of its greater franchise, members of these multimedia franchises should ideally be connected to one another and have these relationships described in semantically informative ways. Similar to the issue of descriptive data, these semantic connections are often not found at traditional memory institutions but are present on the web at sites like Wikipedia, where fans are able to contribute data themselves and establish relationships between various types of related media objects.

That is not to say the situation of pop culture data representation on the web is perfect. While a limited domain knowledge or structural constraints found at traditional data providers can lead to a lessened amount or granularity of data for pop culture resources, there are benefits to this more formally created information. For example, using abstract data models such as the Functional Requirements for Bibliographic Records (FRBR) as the foundation for modeling and describing data can help better understand how a single resource exists among the various conceptualizations of the same or related creative works. Similarly, the division of a single creative work into multiple distinct yet related entities

(in FRBR's case, the Work, Manifestation, Expression, and Item, or WEMI entities) can also improve how resources are described, as bibliographic data describing an entity which represents a specific level of abstraction can be applied to that entity without confusion, as singular entities will not represent multiple conceptual layers of a single creative work. The formal rules which define data creation at memory institutions also tends to provide some level of consistency, and this has resulted in the creation of bibliographic authorities, such as those provided by the Library of Congress¹, National Diet Library², and others.

Though observant readers may now question whether the way to move forward is a straightforward one – that is, simply apply existing bibliographic models to these niche mediums – attempting to do this reveals several inadequacies. The cause of most of these inadequacies stems from the core issue this dissertation examines; many objects of popular culture, and the multimedia franchises come into existence based on their relationships, are not adequately serviced by existing bibliographic description paradigms, particularly when one seeks to properly meet the needs of users. The issues can be described as such:

1. **Limited support for common entities and relationships.** Though bibliographic models such as FRBR are not strictly limited to the modeling of monographs, their limitations when applied to more dynamic resources such as manga or anime, or resources with numerous derivatives and ambiguous manifestation levels like video games become apparent. The entities and their relationships between entities that such models have predefined not only limit (or encourage vague interpretations of) the types of entities that can be defined, but crucially may limit the types of relationships between resources that can be expressed. Again, this has large implications in search and retrieval for users, particularly on the web and particularly when seeking to establish relationships across media types or across distinct multimedia franchises.
2. **Poor descriptive granularity.** Perhaps more than most other resource types, fans of pop culture materials are highly interested in extremely granular data (Fee, 2013). The amount of granular data found among web resources with data created by hobbyists, such as Wikipedia or the countless fan pages that exist for various pop culture franchises, provide no shortage of examples of this. Traditional library cataloguing rules, vocabularies, institutional constraints, and commonly staff expertise, are often not able to adequately describe specific themes, genres, character archetypes, art styles, relationship types, and other granular data that potential users of this data would like to access. While some may view this level of granularity as superfluous for certain memory institutions whose main goal may be to facilitate access to their collection rather than thoroughly describe its contents, studies have shown that such niche data is important for users in aiding in tasks like search and retrieval (Lee *et al.*, 2013), encouraging a

¹ <https://authorities.loc.gov/>

² <https://id.ndl.go.jp/auth/ndla/>

more granular level of description even for such institutions. Even if more granular data was available, however, like point (1), models and their defined ontologies may not be able to apply this data in a way that provides added semantic context for users.

Luckily for the author, these issues are quite intertwined, and again can be traced to the inability for these unique resources to conform to traditional bibliographic models and concepts. The dissertation presented here is the summary of both theoretical and practical research that has been performed in seeking to explore these issues and their possible solutions. In short, the central argument of this dissertation is that multimedia objects, particular those that are a part of large franchises, are sufficiently distinct from resources which have traditionally been the focus of descriptive modeling and that this distinction warrants new paradigms in order to sufficiently portray them, establish relationships between related objects, and describe them to a level that adequately meets user needs. Although descriptive models and entity creation are typically performed by professionals and imposed downwards onto users, this dissertation takes the view that the way that the intelligent and complex ways in which hobbyists currently portray and describe pop culture objects on the web should be used as guidance in the formation of new descriptive models and methods of description. This study stipulates that the effort put forth by hobbyists in portraying multimedia objects on the web can and should be used to not only to formulate new and better ways of modeling and connecting related pop culture materials, but that the existing descriptive data they have created for these objects can be used, through methods like metadata aggregation, to provide them with granular descriptive data. While the research goal of “improving pop culture data representation” is rather broad, the betterment relates directly to what new functions an improved portrayal and description of pop culture materials can enable, and how users can be better served, be they casual or hardcore fans on the web, or researchers and other individuals with different needs. Thus, this dissertation seeks to improve on the portrayal and description of various types of multimedia resources, not only for the benefit of users, but building on the concepts and descriptive efforts they themselves have made.

Research Novelty

Though some research has previously examined the phenomena of user interaction and new media inclusion for multimedia franchises, transmedia, or pop culture objects generally, the role of these concepts within the larger bibliographic landscape and specifically the role of the franchise as a singular conceptual entity has received little attention. Similarly, while some attention has been paid to how features such as additional ontologies can enable a more thorough description of pop culture resources, these possible sources of this granular data has not been well explored. Novel aspects of this research address these two areas by examining the bibliographic description of a multimedia franchise at its highest conceptual level, the Superwork, and by exploring the use of user-created bibliographic metadata as worthy of inclusion into more formal descriptive practices. Additionally, this dissertation

highlights the importance of enabling semantically informative relationships between entities and analyzes how users currently portray such relationships themselves on the web. The roles of open vocabularies and flexible models that support the establishment of relationships are central to this research, both to connect entities into a franchise hierarchy, but also to allow for the connection of whatever entities diverse user groups wish to connect as a way of bringing semantic context to a variety of resources that would otherwise lack such context.

The Open Archives Initiative Object Reuse and Exchange (OAI-ORE) aggregation model is designed primarily for the description and aggregation of web resources. The function of aggregating existing web resources was used throughout this research in order to create single resources representing for various entities within multimedia franchises, with existing web resources that describe the same entity being used both as representatives of the resource in question and as sources of descriptive metadata. Additionally, OAI-ORE includes features such as the ability to create recursive, nested aggregations, and the ability to both describe and connect aggregation entities using vocabularies of the users choosing. In utilizing these features in unique ways meant to address central research questions, another novel aspect of this research involved the creation of bibliographic hierarchies using OAI-ORE and existing web resources as a way of modeling multimedia franchises. Aggregated entities were created based on existing web resources, and relationships between these entities were established in order to create said hierarchies. This not only results in a web-friendly version of a FRBR WEMI-style bibliographic hierarchy, but the use of open vocabularies allows the relationships between resources that form a hierarchy to be described with relational terms that enable greater semantic context such as prequels or sequels, spin-offs, derivations, adaptations, and so on. In addition, the act of creating multiple entities within a franchise hierarchy and the allowance of connections between any entities meant that relationships could be made across franchises, allowing for relationships such as spiritual sequels and crossovers to be described. Summarizing, this dissertation aims to contribute novel research ideas surrounding bibliographic modeling, entity and relationship portrayal, and bibliographic descriptions of diverse pop culture objects and multimedia franchises, and how user-created concepts can positively impact each of these areas.

Relevant Terminology

Though more specifics on chapter-relevant terms will be discussed where appropriate, a clarification of some major terminology that is present throughout the paper can be helpful prior to introducing later ideas that frequently reference them or rely on their understanding.

Pop culture materials and MAG objects

In referencing pop culture materials, this paper is referring to a specific group of resources, namely manga, anime, and video games. Though discussion will occasionally touch on related resources, these three resource types and the relationships between them are the core mediums that the majority of the

research presented here is concerned with. With regards to specific resource types or mediums, the most common referred to throughout this dissertation will be manga, which is a style of Japanese comic. Though typically appearing first in serialized weekly magazines as individual chapters, these magazines are rarely collected and catalogued as whole units. Individual chapters from magazines are eventually collected into and distributed as volumes, and, unless stated otherwise, this volume format will be the focus of most manga discussion. Japanese animation (anime) and to a lesser extent, video games (games) will occasionally be discussed, but rarely individually and more often as a group of resources – manga, anime, games (MAG) – that frequently have important relationships to one another and hold special relevance in this dissertation when discussing multimedia franchises. The majority of the data collection was done primarily with a focus on manga and occasionally anime, though video game data was not thoroughly gathered or analyzed throughout any of the research presented here.

Multimedia franchise and transmedia

Throughout this dissertation, the use of the term Superwork is meant to evoke a conceptual entity that represents a given multimedia franchise at its highest conceptual level. *Franchise* may carry specific commercial meanings, though in the context of this dissertation, it is used to refer to a group of related media instances and creative works that are connected via relationships such as prequels, sequels, derivations, and spinoffs across several different mediums not limited to the aforementioned MAG resources (though these will remain the most relevant in most discussions). While the prevalence of the multimedia phenomenon is difficult to determine, the commercial incentives that come with adapting a popular novel, game, film, etc. into other formats is clear and consistently exploited. In the Wikipedia article for the *List of highest-grossing films*,³ 24 out of the top 30 films belong to multimedia franchises, with others (e.g. *Titanic*, *Frozen*) belonging to major brands or franchise-like entities. Though a universal definition is difficult to obtain, there have been attempts at providing guidance as to when a creative work and its related materials are eligible to be referred to as a *multimedia franchise*. For example, one of the best resources for finding examples of multimedia franchise is the *List of multimedia franchises* Wikipedia article⁴, and this article does attempt to both define the term and provide guidelines for inclusion into the article:

A multimedia franchise is a media franchise for which installments exist in multiple forms of media, such as books, comic books, films, television series, and video games. Multimedia franchises usually develop due to the popularization of an original creative work, and then its expansion to other media through licensing agreements, with respect to intellectual property in the franchise's characters and settings, although the trend later developed wherein franchises would be launched in multiple forms of media simultaneously.

In order to qualify for this list, a franchise must have works in at least three forms of media, and must have two or more separate works in at least two of those forms of media (a television series or comic book series is considered a single work for purposes of this list; multiple spin-

³ https://en.wikipedia.org/wiki/List_of_highest-grossing_films

⁴ https://en.wikipedia.org/wiki/List_of_multimedia_franchises

off series or reboots of a previously ended series are considered multiple works). For example, a television series that spawned one film and one novelization would not qualify; a television series that had a spin-off series, or was remade as a new series, and which spawned two films and one novelization does qualify. This list does not include public domain works from which adaptations have been made in multiple media, but which do not involve licensing or other means by which an author or owner controls the franchise. A franchise may be included if it obtained multimedia franchise status prior to works within the collection entering the public domain. (Wikipedia, 2018a)

The definition given here – a franchise with installments in multiple media formats that usually have a single progenitor – is one that does align with colloquial usage of the term and does match the concept of the Superwork as used throughout this dissertation. In other words, any further mention of a *multimedia franchise* or the Superwork should be understood by the reader in these terms. Additionally, the use of the term will retain a broad meaning of a series of related works containing instances in various media formats, and a narrow definition of what exact formats and how many installments in each are required to be considered a multimedia franchise will not be carried out here.

The concept of the multimedia franchise has also been referred to by numerous other terms. In Japan, for example, the phenomenon of a connected series of works spanning multiple media formats is referred to as メディアミックス, or *media mix*. Similar to the definition for a multimedia franchise given above, a *media mix* is a “cross-media serialization and circulation of entertainment franchises” (Steinberg, 2012, pp. viii) and the prevalence of manga, anime, games, character merchandise, etc. in Japan has meant that creative works more often than not become part of, or develop into, larger media mixes or franchises (see Picard and Pelletier-Gagnon, 2015; Nozawa, 2013). The prevalence of the extension of creative works into other mediums can further be seen by the existence of specific terms denoting specific medium adaptations (Wikipedia, 2018b), such as 映画化 or ゲーム化 (*eiga-ka* or *ge-mu-ka*) indicating a work has been recast into a film or game, respectively.

The term *transmedia*, which may be defined as narrative and production techniques that tell a single story or portray a single imaginary world across multiple media platforms (Vukadin, 2014), is perhaps the most used in related research into the multimedia franchise phenomenon. While numerous related terms exist, such as the previously mentioned cross-media (Lemke, 2004; Steinberg, 2012), distributed narratives (Walker, 2004), networked narrative environments (Zapp, 2004) and distributed storytelling (Davenport, 1998), the term *transmedia* can be used to imply the same ideas in most cases; *transmedia* also more prevalent in recent literature and related research, and so these other terms will not be expanded upon. Though common, the usage of the term is not always in reference to the same phenomenon. In particular, there is often some ambiguity regarding whether or not *transmedia* refers to a multimedia franchise-like practice, or if the term is to be used specifically when a single narrative has been divided and spread across multiple mediums; while the former implies that individual works across mediums are related yet distinct, the latter suggests that for a total understanding of a story, one must access multiple or all of the different media instantiations. Vukadin (2014) writes that though

adaptations use each medium to create a unique experience and different interpretation of a work, they “mostly transfer the same content from one medium to another” whereas transmedia “delivers unique pieces of content in each of the media channels, seeking to synchronize them and integrate them into a complex, yet coherent narrative structure” (p. 298). Branch *et al.* says transmedia storytelling uses “elements of a media property that advance the narrative” and that this differs from “elements that are simply brand expressions” (2017, p. 2772), which may be regarded as more aligned with the definition of a multimedia franchise. Giovagnoli (2011) states that transmedia is a narrative form that shares “the same elements (plots, characters, atmospheres...) but that change depending on the publishing platform through which they are released” (p. 12). Referring to Jenkins, Dena (2009) states that he “argues that practitioners should provide unique story information in each media. But when this happens fans of a fictional world are splintered into groups according to whether they are literate in the various media” (p. 160), while Jenkins’ writing claims that each medium in transmedia storytelling “does what it does best” and that “each franchise entry needs to be self-contained enough to enable autonomous consumption. That is, you don’t need to have seen the film to enjoy the game and vice-versa” (2003, p. 4), suggesting that each medium exploits its strengths to tell an adapted story rather than a prescription that each should provide unique story information. That said, Jenkins also states “consumers must assume the role of hunters and gatherers, chasing down bits of the story across media channels” to fully experience a fictional world (2006a, p. 21). Regarding adaptation, Dena states it is a central part of transmedia and Vukadin writes the TV series *Game of Thrones* “can be viewed as a new franchise expanding strategy rather than a mere TV adaptation of George R. R. Martin’s novels.” (2014, p. 298). The differences between *transmedia* and *transmedia storytelling* phenomenon are interesting and hold important implications in determining what group of works constitutes a Superwork, but an attempted clarification of the definitions of each term will not be presented here. Thus, for the remainder of this dissertation, the term Superwork will be used to represent an entity that evokes similar concepts as those which other sources may refer to as *multimedia franchise*, *transmedia*, or various other related terms, with *transmedia storytelling* being reserved for the use of a single narrative spread across multiple media types that requires one to consume these multiple formats for a thorough understanding of the narrative, sometimes also referred to as *media convergence*. The distinction between the two terms in this way can be seen elsewhere⁵ should allow for the expression of either idea with more clarity. A more thorough discussion of what the Superwork entity represents, and how it relates to similar, but medium-specific terms, such as a *bibliographic family*, will be carried out in later chapters.

Methodology & Structure of this Dissertation

This paper presents the explanation, exploration, and proposal of solutions to the previously outlined issues surrounding the portrayal of bibliographic data for pop culture resources on the web and in

⁵ https://en.wikipedia.org/wiki/Transmedia_storytelling

traditional memory institutions. To do this, the main body of the dissertation will detail three main related research issues, each of which includes work conducted and presented as several past publications. Note that some of the content included in this dissertation is found verbatim in these publications.

The chapter following this introduction will detail literature on related topics such as franchise-level bibliographic entities, the modeling of pop culture objects, and metadata aggregation models. Next, chapter 3 will present an overview of the types of bibliographic data provider that much of this research was focused on, either as targets of aggregation or as the basis for building new bibliographic models. The first main research focus is detailed in chapter 4 is that of the Superwork entity, which is a conceptual entity meant to represent a multimedia franchise at the highest level. The analysis of existing bibliographic data on the web made it apparent that a conceptual entity at this level was necessary. After an examination of existing bibliographic models and the entities that may fit this requirement, the establishment of the Superwork entity is described. This chapter also briefly discusses the difference between the Superwork as a representation of a multimedia franchise, and broader concepts such as topics or subjects. With this entity established, chapter 5 details the second main research issue and examines the ability (or inability) of those same bibliographic models (i.e. FRBR and FRBR_{OO}) to describe the totality of bibliographic data for various pop culture objects as they exist on the web. Again, using the way users currently portray and relate materials on the web, a new bibliographic hierarchical model is proposed. Discussed here is the ability of the entities within this model to portray data as it exists among various web providers and their placement into a hierarchy to establish a complete multimedia franchise, including the conceptual Superwork level. The flexibility of the model is emphasized as a way to support a diverse range of resource types, description methods, and hobbyist communities. This chapter also outlines how the model has changed throughout the course of previously performed research. The final primary research issue of metadata aggregation is the subject of chapter 6. Here, the basics of the OAI-ORE aggregation model are briefly outlined before moving on to its role in this research. Using existing web data providers as aggregation targets, OAI-ORE was used in a novel way to create core entities, including the Superwork, by identifying web pages describing the same core entity and aggregating them into a single resource; these aggregations were then linked together to form the bibliographic hierarchy as it was defined earlier. This chapter also discussed the role of OAI-ORE in aggregating descriptive metadata from web resources to provide the hierarchical entities with granular descriptive metadata. Like the hierarchy itself, the flexibility of OAI-ORE is emphasized as a way to enable a variety of descriptive properties, semantic relationship terms, and resource types that may differ from community to community. Chapter 7 follows with a discussion on the lessons learned throughout the dissertation, the limitations that the proposed methods and models are faced with, and the overarching implications that the research has on various relevant areas. Chapter 8 concludes with ideas for future work and closing remarks.

Chapter 2: Related Literature

This chapter details related literature focusing on the entities similar to the Superwork, both within institutions and on the web, the hierarchical representation of bibliographic objects, and related OAI-ORE and other aggregation works. Though rare, related work on hierarchical representations and aggregations that deal with multimedia franchise and pop culture objects are also discussed, as are more general aggregation studies that have been relevant throughout studies conducted as a part of this dissertation.

Superwork-like Entities and Hierarchical Representations

The idea of a conceptual entity representing multiple related works has been examined in the past using various names and in various contexts, particular that of library collections and novel ways to organize, catalogue, and display their contents. This library-focused research has discussed an entity resembling the Superwork in various contexts, such as research into the development of a superwork-like entity (Carlyle, 1996; 1999; Smiraglia, 2007; Smiraglia *et al.*, 2014; Svenonius, 2000), studies on whole-part relationships between different bibliographic entities (Niu, 2013), the establishment of bibliographic families and instantiation networks based on related works (Smiraglia and Leazer, 1999; Wilson, 1968), and visualizations of identity networks (Leazer and Furner, 1999). The concepts discussed by these authors using terms such as bibliographic families or instantiation networks differ in their specifics, but all discuss entities meant to represent a series of related, though separate, creative works. Other studies have been performed that attempted to implement this concept of bibliographic families or instantiation networks based on FRBR principles into navigational systems. Ercegovac (2006) examined the effects of such systems on the user experience and found that creating links between applicable related entities led to a better grouping of those entities and enhanced search result navigation, and resource discovery and use. Merčun *et al.* (2012; 2017) created the FrbrVis prototype to create a hierarchical visualization of complex work sets and found that visual displays of these networks aided in the exploration of hierarchical and horizontal relationships that were present in these networks. Though this research was influential in understanding how similar conceptual entities and entity networks were thought of, much of the applicability was limited (outside of a specific manga context) for MAG objects due to the focus on the library catalogue and literary materials. For example, using a random sampling of 996 WorldCat records, Bennett *et al.* (2003) found that only 25% of works existed as multiple manifestations, contrary to the multimedia nature of the MAG objects which are the focus of this dissertation. That said, Zapounidou *et al.* (2017) states that “the identification of bibliographic families and the clustering of all related entities are extremely important and one of the main functions that library catalogs need to deliver” (p. 16).

While much core research has been performed in a library context, there has been other related research either focused on communities outside of the library or focused on multimedia works more broadly. Related multimedia research exists as investigations into the transmedia phenomenon. Branch,

et al. (2017) developed an ontological model for describing transmedia fictional worlds for better knowledge organization and search and retrieval by end users. Though the model and its findings focused mainly on the contents of the transmedia franchises they examined, their domain analysis revealed that in attempting to include fan-fiction and other non-licensed works, which are quite popular among most large multimedia franchises, “establishing relationships from these variants to the licensed elements” was a challenge (p. 2777). These fan works pose a similar challenge in the creation of a Superwork entity, particularly when attempting to define what types of fan works are included – if they are to be included at all – in the range of works that a Superwork represents.

Additional past transmedia works (Jenkins, 2003; Dena, 2009; Giovagnoli, 2011) have thoroughly defined and explored the phenomenon, and have been vital in helping shape what the boundaries of a franchise may be, and thus what exact characteristics make a work suitable for inclusion in the Superwork umbrella. Though non-transmedia works have also discussed this issue in trying to define the superwork or a similar entity, e.g. Svenonius’ notion that the superwork would include commentaries, references, and derivatives (2000), the multimedia nature of the transmedia phenomenon is an important feature that should be examined separately from literature-centric research.

Ana Vukadin’s work *Bits and Pieces of Information: Bibliographic Modeling of Transmedia* (2014), looked at modeling works of transmedia and, like the research presented here, examined the suitability of the FRBR Work and FRBR_{OO} (FRBR Object-Oriented) Complex Work entities in portraying relationships found in transmedia works. Vukadin does not argue for a singular franchise-level Superwork entity, but instead looks at the ability for the Work and Complex Work to portray various relationships found in transmedia. Her findings state that though the Work entity can be conceptual enough to encompass the overarching concept (IFLA, 2006, p. 17, 26-27) that individual works connected to a Superwork may share, the fundamental whole/part relationships found in transmedia works are not able to be properly represented. Instead, “FRBR_{OO} seems to offer more precise definitions for this purpose” using the Complex Work entity (Vukadin, 2014, p. 296). She also argues that the FRBR_{OO} Expression Fragment subclass may be used describing shared assets such as characters, settings, and footage across transmedia works.

Important to the development of the Superwork in this research was Tallerås *et al.*’s study *User conceptualizations of derivative relationships in the bibliographic universe* (2018), which examined how users conceptualize and map relationships between related resources. Based on title pages of play scripts, DVD covers of film adaptations, and CD covers of soundtracks of *Peer Gynt* and *Romeo and Juliet*, users were asked to map the relationships between these resources and apply attributes to each. The majority of the produced concept maps (62/99, or 63%) featured a user-created “central node” that connected some or all of the three materials and featured attributes such as persons of responsibility while attaching attributes such as date of publication to the document nodes. As the central node featured persons of responsibility (e.g. Shakespeare) that were not necessarily responsible for all of the documents attached to that node (i.e. the music CD), the authors state that that “the central node cannot

entirely be interpreted as an FRBR work” and that instead it is “plausible to interpret the central node as an even more abstract collocating device” and “the central node more resembles a superwork entity” (p. 911). The findings that “users seem to prefer multi-entity conceptualizations including a superwork entity” (p. 912) emphasize the need to explore the development of entities representing multiple types of conceptual abstractions, namely the superwork and fictional world, in particular based on how users themselves conceptualize these.

Related niche-media focused Superwork research has been performed by McDonough *et al.* (2010), who found that in attempting to model computer video games, the number of related instances that can exist and the ability for this network to continue to grow over time provided a “compelling justification for the notion of a Superwork” (p. 31). This notion of the relationship network growing over time – one that is inherent to dynamic resource types – provides support for the creation of a franchise-level entity, as Lee and Jacob found that the explicit relationships prescribed by FRBR “might be too rigid to support the flexibility necessary for describing dynamic resources” (2011, p. 18). Additionally, McDonough *et al.* (2010) found that the ability to collocate many of these resources was important for users and that the “application of a Superwork entity may be the simplest means of enabling users’ preferred mode of searching” (p. 31). In Kiryakos *et al.*, (2017) the authors performed preliminary work on exploring the relationships between related entities in a multimedia franchise and the inability for existing entities to properly describe them, with this work being continued in Lee *et al.* (in press), where the Superwork concept was contrasted against existing entities in attempting to define its properties and characteristics.

In summary of this past research, it can be said that existing research into a Superwork-like entity has been conducted, though often within the limited context of the library and its materials, thus reducing its applicability to multimedia franchises and MAG objects which this dissertation is centered on. Unique aspects of these dynamic objects when compared to more traditional library materials justifies the exploration of new entities and model structures better able to represent them for the benefit of users, and studies have shown that users themselves can provide insight into how such entities and models may be structured. Research into existing models and entities have shown that certain entities within these models such as the FRBR Work or FRBRoo Complex Work may be able to represent a franchise-level entity, though a thorough comparison of these entities to existing portrayals of such entities as they are currently portrayed on the web has yet to be undertaken.

The inadequacies of existing bibliographic models in suitably describing certain resources has been discussed in some previous pop culture research. Studies focusing on videogames and anime (Jett *et al.*, 2016; Lee *et al.*, 2013) show that models such as FRBR are unable to adequately portray the complex relationships between these resources. This may be in part due to the intended use of certain models within the library context, thus limiting the anticipated diversity of resources and relationship types that are needed; the prior mentioned study by Bennet *et al.* (2003), which, based on a random sampling of 996 WorldCat records found only 25% of works existed as multiple manifestations, is

indicative of this. Aside from problematic predefined relationship issues with the FRBR WEMI entities, the structure of the model itself has been viewed by some as being too rigid to support dynamic resources (Lee and Jacob, 2011, p. 18). As multimedia franchises and specific mediums such as manga and anime within them tend to grow over time (i.e. with the addition of new chapters, volumes, episodes, home releases, etc.), the need to support such dynamic resources is vital for the overall representation of a franchise. That said, Teixeira (2010) argues that hierarchical representations like FRBR have increased benefits for certain types of works such as serial collections or works that exist across different publishers, editions, and, mediums. In summarizing Noerr *et al.*, Bennett *et al.* (2003) echoes this claim and suggests the primary benefits of FRBR are due to its hierarchical structure, allowing for entities at lower levels to inherit bibliographic descriptive data placed among higher ones. Additionally, Coyle (2004) states “it is reasonable to assume that a future cataloguing structure will embody some degree of hierarchy, especially in the need to express the relationships between multiple versions of the same work” (p. 168). This suggests that though a hierarchical representation of work collections may be beneficial, existing models i.e. FRBR may be inadequate in handling certain resource types. As an incorporation of web data into the model is desired, the entity types described by the web are also important to understand. Again, these fail to align well with existing models like FRBR, with rough comparisons to the Work or Expression levels often possible, but a clear WEMI structure rarely, if ever, represented by hobbyist data providers. Web data more often exists as descriptions of either the franchise (Superwork), a distinct series or universe within that franchise, or single media types. These facts and requirements guided the move away from attempting to conform diverse data into restrictive models, and instead create new entities that more readily align with how MAG object data is frequently portrayed, though with the maintaining of a hierarchical structure to enable the representation of complete franchises and relationships between various entities.

Aggregations and the OAI-ORE Model

Though much of the existing research into the OAI-ORE model has been performed in the context of archival descriptions, there are important takeaways for this dissertation in how the model has been implemented by various groups. Ferro and Silvello (2013), whose use of the OAI-ORE Proxy entity to connect resources with relationship property such as “isMetadataOf” (p. 224) rather than a more straightforward (and used in examples by the OAI-ORE documentation) “hasNext” type property encourages the connection of entities using more descriptive properties. Though the Proxy property is meant to be used with Aggregated Resources specifically, thus not being mandatory for Aggregation level connections, its ability to add semantic context to certain entities and relationships make it useful in contributing to primary research goals of increased granularity and semantic information; establishing relationships between entities is important, but when given more semantic meaning, they provided even more utility for users.

The importance of separate but linked metadata records, and OAI-ORE's role in achieving this, is echoed by Niu (2015), who emphasized that relationships between resources can and should be made explicit through the use of OAI-ORE's allowance of relational terms from other vocabularies. The work also details how the model can "represent an unlimited number of levels in a hierarchical description through recursive nesting" (p. 127) despite defining only two main entity levels. Again, the parallels to this research are clear, as an open model supportive of dynamic, changing entities, along with the support of open vocabularies to describe and connect these entities, facilitate the pursuit of the overarching research goals in this study.

Though types of relationships that are common between entities are not often detailed by Niu, Wang *et al.* (2013) specifies several in their work on the development of a clustering algorithm for the Europeana Data Model in order to create hierarchies based on semantic groupings. Though the content of the Europeana portal – works of cultural heritage sourced from numerous European museums and other cultural heritage institutions – differs from the nature of the multimedia objects described in this dissertation, the types of relationships revealed during Wang *et al.*'s study are common amongst most resources with relations and aided in guiding the boundaries of aggregation entities when formulating the hierarchical model. Such relation types included different views of the same object, different parts of the same object, derivative works, and thematic groupings (p. 252-254).

The work of Jerome McDonough has been instrumental in attempts to use OAI-ORE both in the pop culture space and in the representation of linked bibliographic hierarchies. In his works detailing the Preserving Virtual Worlds (PVW) project (2011, 2013b), McDonough discusses the preservation of digital games and interactive fiction, creating a PVW ontology and applying it to an OAI-ORE based model in order to represent the FRBR WEMI hierarchy for those digital works. In the modelling, each of the WEMI entities is given its own aggregation, connected hierarchically using the *ore:aggregates* property, i.e. the Work aggregates the Expression which aggregates the Manifestation, which aggregates the Item. This use of connected aggregations to represent a hierarchy, specifically the FRBR WEMI hierarchy, was similar to early research conducted by the author that likewise focused on FRBR as a way to hierarchically portray media franchise objects. Like Ferro and Silvello, the creation of unique properties in order to give relationships more semantic context is emphasized, such as McDonough's use of the *pvw:hasContextInformation* property being used to link an entity to a relevant URI. McDonough's again used OAI-ORE (2013a) in an attempt to develop a hierarchical representation of computer games. A major challenge encountered was the metadata creation process required considerable manual labor to portray the entire WEMI hierarchy for a game instance and its related entities. The lack of a proper scalable solution can be a significant issue, as data is sometimes gathered outside of reliable API methods, such as HTML scraping, and so manual entry may become an obstacle. Despite this, McDonough seems to acknowledge the capability of the OAI-ORE model in multiple areas related to the project, stating that "aggregations could be created for digital content being used as

representation information and context information, and appropriate links created in the aggregation resource map files” from his ontology (p. 59).

Chapter 3: The Portrayal of Bibliographic Data for MAG Objects

Whether attempting to determine what the Superwork represents, deciding how to formulate a suitable bibliographic hierarchy, or identifying what resources to aggregate to provide the most utility, an important consideration common throughout the research presented in this dissertation is what the current state of bibliographic data for MAG and other pop culture resources is among various data providers. For example, prior to creating entities for a new bibliographic model, or in determining the utility that data from a particular provide can contribute to an aggregation, the amount, granularity, and presentation of bibliographic data that is present among various providers of bibliographic data should be understood. These various providers typically fall into one of two categories, referred to in this research as **institutional** or **hobbyist**, with institutional referring to bibliographic descriptive data that typically appears in and is created by library and related cultural heritage institutions, and hobbyist referring to data produced by fans, collectors, or enthusiasts on the web. While institutional tends to serve a specific audience and is of limited granularity, it serves an important purpose for supplying and maintaining authoritative data. Data produced by hobbyist providers tends to be less formal and authoritative, but considerably more granular, often multilingual, includes a number of connections to related entities belonging to the same and different franchises, and in most cases is generally more suited to fulfilling a wider variety of search and retrieval needs that users may have. Because multiple aspects of the bibliographic data from both provider types has been central to informing several of the concepts and decisions in the research presented in this dissertation, detailing this data – which forms the content of this chapter – to provide an introductory context prior to detailing the studies themselves may prove beneficial for readers.

Bibliographic Data for MAG Objects as Portrayed by Libraries and Similar Institutions

Throughout this research, the type of bibliographic data provided by libraries and related institutions is referred to as **institutional** data. Institutions such as libraries, museums, archives, other cultural organizations, and corporations, have historically been creators and maintainers of bibliographic data for a variety of resource types. This is especially true for manga – the central resource to much of the research presented in this paper – as while a unique bibliographic resource type, it does share many similarities with more traditionally described objects such as monographs or serials. Though each individual institutional data provider will differ in a number of ways, they tend to share a number of core features that enable them to be grouped together and contrasted against the other data provider type, the hobbyist provider. The most significant attribute shared among institutional providers is the way bibliographic data is described; the descriptive properties and values institutional providers choose to use tend to focus on the resource as an object or item rather than focusing on the content. As a library or other institution may create a record to act as a finding aid or a description of their holdings, it is unsurprising for resources to be described in this way and not to describe the content to any significant

degree. Stemming from this notion of focusing on a MAG as an item, the following characteristics can be said to be shared amongst most institutional providers:

1. **Data created based on formal models and vocabularies:** Though the level of formality and constraints can vary, institutional providers such as libraries or corporations typically have strict rules governing the data they record (e.g. granularity, language), how it is recorded (e.g. the ontologies or authorities used), and where it is sourced from (e.g. requirements that certain data come from certain pages). These can be coded guidelines such as RDA, or simply a set of instructions governing how data is recorded into a corporate inventory list in a spreadsheet document. Although, depending on the institution, this has a limiting effect on the amount and granularity that will be recorded for a given object, it also leads to a level of consistency in the data, which enables the creation of bibliographic authority files, such as those maintained by the United States' Library of Congress⁶, Japan's National Diet Library⁷, and others. As a lot of MAG objects are Japanese in origin, yet described in English or other languages, consistency in fields such as the translations and transliteration of names and titles is of particular importance for many of the pop culture works discussed here.

2. **Limited target audience and use cases:** Both institutional and hobbyist providers will have some form of an intended audience in mind, but audiences for institutional providers tend to be more refined, and the intended uses of their data fewer in number. Manufacturers of physical copies of MAG objects (manga volumes, anime home releases, etc.) may record data about the items they produce, and such records typically are not public and are meant to be used as business records resulting in data such as the date a certain item was manufactured or printed, the place of its manufacture (rather than publication), or initial sale price. A library or other public institution will have a more diverse audience than a corporation, though the intended use case is often limited as well and is focused on enabling users to find a desired item in a collection, or browse items based on some description. Again, this is reflected in the data recorded, with titles, authors, subjects, and classifications being central to most records.

3. **Limited connections between related objects:** Institutional providers tend to have a limited number of relationships established between various related entities. While this is slowly changing through acts such as including links to relevant external sources within records (see Bailey, 2018), direct or indirect connections to related items is generally limited. Connections will occasionally be made through shared properties such as subject terms, authors, or titles, such connections are generally limited to what an institution has available in its collection or can make accessible to users. One effect of this is that individual catalogued objects are more likely to have connections to other objects rather than broader conceptual entities, meaning

⁶ <https://authorities.loc.gov/>

⁷ <https://id.ndl.go.jp/auth/ndla>

relationships such as sequential volumes in a manga series may be present, but connecting those volumes to other entities within their franchise hierarchy is unlikely; this again stems from institutions focusing on objects rather than content. This is not to imply that library or other institutional works are devoid of higher conceptual data from FRBR Work or related levels, but because such entities are not explicitly catalogued, any relationships to entities such as a franchise must be inferred from the descriptive data itself rather than being explicitly made. Additionally, as a result of the formal vocabulary usage mentioned in (1), connections that are made may lack any specific semantic meaning (e.g. spin-off, specific format adaptation), and the user may simply be informed that the relationship between two resources is that they are “related.”

The Media Arts Database

The primary institutional provider used throughout the research presented here was the Media Arts Database⁸ (MADB) which is a public database maintained by the Japanese government’s Agency for Cultural Affairs. The MADB contains four separate databases for manga, anime, videogames and media arts; two sample pages from the manga and anime databases can be seen in Figures 3.1 and 3.2. The choice to use the MADB as the main institutional provider was done for several reasons.

1. As a practical matter, access to the data for research purposes was granted by the database maintainers, which resulted in both a more convenient and deeper level of access than previous research using libraries where data had to be taken from MARC records or through HTML scraping.
2. Because the bibliographic data from MADB is sourced from numerous other institutional providers, e.g. National Diet Library, Kyoto International Manga Museum, Osaka Prefectural Central Library⁹, it can be viewed as representative of a variety of institutional providers. Evidence of this can be seen in Figure 1, where the bibliographic properties such as Author List (which is the MADB property resembling the common library / MARC property Statement of Responsibility), First Publication [date], and Category (i.e. DDC number) are similar to properties one finds in a library catalogue. While finding a library that has detailed catalogue entries of television or film animation is difficult, the properties shown in Figure 3.2 are of a similar level of granularity to the library properties given for manga in Figure 3.1, e.g. Number of Episodes, Date of Release and End.
3. An important consideration was that MADB catalogues all three of the MAG resource types this research focuses on. While the obvious benefit of this is that if desired, descriptive data for each resource is available from a single provider, another benefit is that the presence of multiple

⁸ <https://mediaarts-db.bunka.go.jp/>

⁹ See <https://mediaarts-db.bunka.go.jp/help/mg/library.html> for all providers

resource types on a single site makes the future possibility of establishing connections between them easier than attempting to do the same across multiple sites. While only limited connections exist across the database now, even between materials belonging to the same multimedia franchise, the structure and existing data values of the MADB mean establishing such connections may be feasible, provided the models and methods of connecting related materials exists.

4. Lastly, unlike traditional institutional providers, the MADB describes resources in varying ways, each represented by a unique web page type. These varying ways refer to different conceptual levels of thinking about a resource, similar to the aforementioned FRBR WEMI entities. The page in Figure 3.1 is an example of the *Comic Works* style page for manga (shown by the URL in the figure’s caption), which describes the resource in a way that more closely resembles the conceptual Work entity from FRBR rather than the Manifestation or Item. Additionally, there are *Book Titles* pages representing a published series of manga volumes, and *Books* pages representing individual volumes themselves. Looking at the properties such as Work Title, List of Authors, and First Publication, these properties and their values can be applied to the *One Piece* manga as a conceptual work, as these properties are agnostic of specific publications, editions, volumes, or arguably media formats.

The screenshot shows the MADB entry for the *One Piece* manga series. The page is titled "Manga Work: ONE PIECE([著]尾田栄一郎)".

Manga Work ID	MMT000042177
Manga Work Title	ONE PIECE
Manga Work Title-kan	One piece / ワンピース
Alternative / original Titles	-
In Roman Character	-
List of Authors	[著]尾田栄一郎
Author Reference ID	A100676651
Period	-
First Publication	-
Description	-
Category	-
Tag	海賊
Rating	-

On the right side, there are two sections:

- All Volumes** (31 Items): A list of volumes with columns for Book Title, Author, and Volume in total.

Book Title	Author	Volume in total
One piece(ジャンプ・コミックス)	[著]尾田栄一郎 / [編]ホーム社	87 Items
One piece / Eiichiro Oda ; [「E」overs [E] attning, Simon Lundstr [E] om], Swedish ed.(Manga)	Oda, Eiitir [R] o, 1975-[尾田栄一郎]	24 Items
ONE PIECE総集編(週刊少年ジャンプ特別編集増刊)	[著]尾田栄一郎	20 Items
ONE PIECE総集編"water seven"(集英社マンガ総集編シリーズ)	[著]尾田栄一郎	15 Items
ONE PIECE COLOR WALK尾田栄一郎画集(jump comics deluxe)	[著]尾田栄一郎	6 Items
- Title on Magazine** (4 Items): A list of magazine titles with columns for Title on Magazine, Author, Magazine, and Date.

Title on Magazine	Author	Magazine	Date
ONE PIECE	尾田栄一郎	ジャンプSQ.	2008/02/01

Figure 3.1 Screenshot of the MADB entry for the *One Piece* manga series. Full page at https://mediaarts-db.bunka.go.jp/mg/comic_works/81200.

メディア芸術データベース 日本語 English 繁体中文 한국어

MANGA ANIMATION GAME MEDIA ART

Anime Search Top Information on Works (Series)

Information on Works (Series)

[Return to Search Results](#)

Anime Series ID	ANS000302400	Distribution Format	劇場
Anime Work ID	ANT000302400		
Title	GHOST IN THE SHELL/攻殻機動隊		
Title-kana	ゴースト イン ザ シェル コウカクキドウタイ		
Date of Release	1995/11/18	Date of End	-
Time Slot / Duration	80分		
Number of Broadcasts	1		
Number of Episodes	-		
Distributor	松竹		
Broadcast Period	-		
Production Company	制作:Production I.G / 製作:講談社・バンダイビジュアル・MANGA ENTERTAINMENT Reference ID		

Original Source	士郎正宗
Director	押井守

Episodes **0 Items**

Episode No.	Episode Title	Date of Release	Notes
-------------	---------------	-----------------	-------

Materials **0 Items**

Material Name	Category	List of Authors	Access to this Material
---------------	----------	-----------------	-------------------------

Packages **0 Items**

Title	Volumes	Other Editions	Publication Format	Date of Publication	Access to this Material
-------	---------	----------------	--------------------	---------------------	-------------------------

Related Series **25 Items**

Figure 3.2. Screenshot of the MADB entry for the *Ghost in the Shell* anime series. Full page at https://mediaarts-db.bunka.go.jp/an/anime_series/5356

メディア芸術データベース 日本語 English 繁体中文 한국어

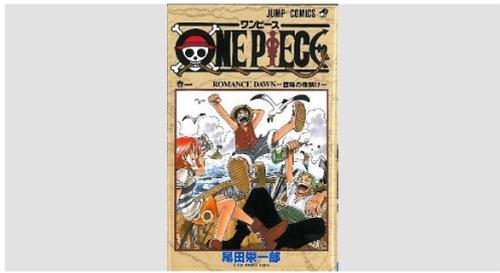
MANGA ANIMATION GAME MEDIA ART

Manga Home Individual volumes: One piece(ジャンプ・コミックス)巻1

Book : One piece(ジャンプ・コミックス)巻1

[Previous](#) [Next](#)

Images *All Rights Reserved.



Basic Information

Book ID	MMM000129116	All Volumes ID	MMG000129116
Book Title	One piece		
Book Title-kana	One piece / ONE PIECE / ワンピース		
Additional information on Cover	Romance dawn : 冒険の夜明け		
Additional information on Cover-kana	ロマンスドーン / ポウケンノヨアケ		
Volume	巻1	by Number	1.0
Additional Edition	-		
Description	-		

Author information

list of Authors	[著]尾田栄一郎 / [編]ホーム社		
Author Reference ID	A100283065 / A100676651 / A100676651		
Heading	尾田, 栄一郎(1975-) / オダ, エイチロウ(1975-)		
Author	尾田栄一郎	Author-kana	オダエイチロウ
Writer/Original Creator	-	Writer/Original Creator-kana	-
Collaborator	-	Collaborator-kana	-

Information

国立国会図書館(1/1)
※Please check with NDL-OPAC for the latest information on the collection.

Registration ID	000002634563 (Moving to NDL-OPAC)						
Book Information ID	C00101MMM000129116						
Date of Publication	1997/12						
Print Version	-	Number of Books Printed	-	Price	390円+税	Book Size	-

Figure 3.3. Screenshot of the MADB entry for first volume of the *One Piece* manga. Full page at <https://mediaarts-db.bunka.go.jp/mg/books/1127049>

Table 3.1. Manga page styles and their unique properties within the MADB

MADB Page Type (manga)	Sample Page Properties
Comic Works <i>Broad, franchise level data and links to multiple relevant Work entities within the MADB.</i>	Work title List of authors First publication <i>Link to MADB's Anime entry</i>
Book Titles <i>Data applicable to all volumes of a manga series, as well as a list of published volumes held in the MADB.</i>	Book title (all volumes) Total number of volumes Heading Imprint / Label List of volumes
Books <i>Data applicable to a single volume of manga, and holdings information for the data sources.</i>	Volume number Additional cover information (<i>subtitle</i>) Number of pages Cover image Physical Dimensions

Further properties used in the MADB description schema for manga can be seen in Table 3.1, which shows the three main manga page types, *Comic Works*, *Book Titles*, and *Books*, along with a number of properties unique to that page type. As the creation of a hierarchical model is central to much of this research, MADB presenting descriptive data as multiple page types which can be aligned to various hierarchical entities that form said model was a vital reason for its consideration as a primary data provider. While the final bibliographic model proposed does not fully align with the page types as divided by MADB – particularly as the model is meant to work with a variety of resource types whereas the MADB contains different page types for each resource – the clear and intentional isolation of specific data to specific conceptual entities allows for the data to be more easily aggregated into particular entities within the model. It allows not only for data for these various entities to be gathered from one provider, but the data sources for the MADB being libraries and other similar institutions means that much of this data can be authoritative in nature, thus fulfilling one of the main functions of institutional providers in the data aggregations presented in this dissertation.

Bibliographic Data for MAG Objects as Portrayed by Wikis and Other Fan-sites

Hobbyist data provider has been used to contrast the institutional provider type. A hobbyist provider is one that exists on the web, is generally created, curated, and open to contribution from other users, and typically features a larger amount of data of a type not available among institutional providers. Like institutional providers, hobbyist ones differ in the properties they choose to describe and how this data is presented, but a series of core attributes are shared and enable them to be viewed in a similar way. Again, like institutional providers, these providers view MAG objects in a similar way, though unlike the institutional, the focus of description is on the content rather than the item. While institutional

providers may have collections to maintain and provide access to, hobbyist providers traditionally lack these central repositories, and thus there is little need to focus on a MAG object as a carrier of content rather than the content itself. This content focus results in several defining characteristics shared among hobbyist providers:

1. **Data creation less constrained by formal models:** All of the hobbyist provider pages used in this research allow for “wiki” style contributions and edits to the descriptive data featured on their site, enabling data to be sourced from a global audience rather than a single (or consortium) collection. Although the forms that enable these public edits will have rules and guidelines which govern the way data is to be entered, there are typically no formal data models or widespread ontologies which these sites adhere to, unlike the institutional providers which tend to have either domain or single-institution rulesets. This does come with obvious drawbacks, such as a possible lack of consistency in formatting, incomplete or incorrect data, uncertain origins of data values, etc. Important to this research, however, is that this also enables the addition of a wealth of data that would otherwise not be able to be recorded, or that is not enabled by the data creation guidelines of institutional providers, with some examples shown in Table 3.2. As emphasized earlier, much of this unique data is related to the content rather than individual items. This results in the creation of the granular data that this research is focused on accessing, and so the importance of the inclusion of this data in fulfilling the data needs of users should not be undervalued.

2. **Diversity of audiences and use cases:** Though it varies from site to site, both the audiences of hobbyist providers, particularly larger ones, and the anticipated uses of their data tends to be more varied when compared to institutional providers. While a corporation’s private database may only need to serve specific business needs of its employees, and libraries aim to facilitate the finding and accessing of materials in their collection, the uses of hobbyist provider data may range from a collector seeking to validate information, someone unfamiliar with a medium seeking out recommendations or plot summaries, enthusiasts seeking to provide more complete data for a less popular series that is currently not well described, and many other possibilities. This lack of a single or limited number of expected audience types in conjunction with the more open data creation guidelines results in a greater amount and granularity of data available, as users are generally free to describe what they want in their own language. What defined vocabularies these sites do provide are generally more granular as well, as terms meant to populate fields such as tags or genres need only apply to specific pop culture resource domains resulting in considerably more semantically meaningful descriptive terms for these fields when compared to institutional authority terms, e.g. those found in the Library of Congress Subject Headings¹⁰, for resources such as manga and anime. A related point here is that an international

¹⁰ <http://id.loc.gov/authorities/subjects.html>

audience can lead to hobbyist sources providing multilingual data. For English language hobbyist providers, this multilingual data tends to be English and Japanese, as the objects of description in question, particular for manga and anime, are often in Japanese (even if the descriptive data is in English). The number of languages provided varies on a per-site and per-article basis, though there tends to be a correlation between the popularity of a site, the popularity of a franchise, and the number of languages provided. A popular series such as *One Piece* for example, is given titles in over 17 languages on the anime directory page AniDB, as shown in Figure 3.5; the newer and less popular 2017 series, *Berserk*, features just 3¹¹. Some sites like Wikipedia may also offer different language pages entirely, though multilingual data may still be accessible on a single page via DBpedia. Still, even these examples typically feature the main descriptive language, such as English, with some basic descriptive data (i.e. titles, authors) in the language of origin, which in many cases is Japanese. While this may seem an overall minor issue, the author has explained in previous studies (Kiryakos and Sugimoto, 2015; Kiryakos *et al.*, 2016) how this single provider multilingual data is vital for aggregating data from various providers, specifically in identifying related or matching resources that may be described in different languages but that still are candidates for aggregation, as the same resources are able to be identified even when different languages are being used to describe them.

3. Networks of relationships between resources: Both due to the lack of formal data creation rules and the desire to describe a greater amount and granularity of data, hobbyist providers tend to establish various types of relationships between resources. These relationships can be straightforward, such as to connect all relevant entities in a multimedia franchise, or relationships that reach across franchises to portray relationships such as inspirations or similar genres or archetypes. Importantly, the lack of restrictive existing ontologies mean that these connections can be described using semantically informative terms and not only vague relational ones. The presence of these relationship networks found among hobbyist providers is as vital as the granular descriptive data for reasons of inclusion in this research, and the importance of these connections will be explored in later chapters.

¹¹ See <https://anidb.net/a12414>

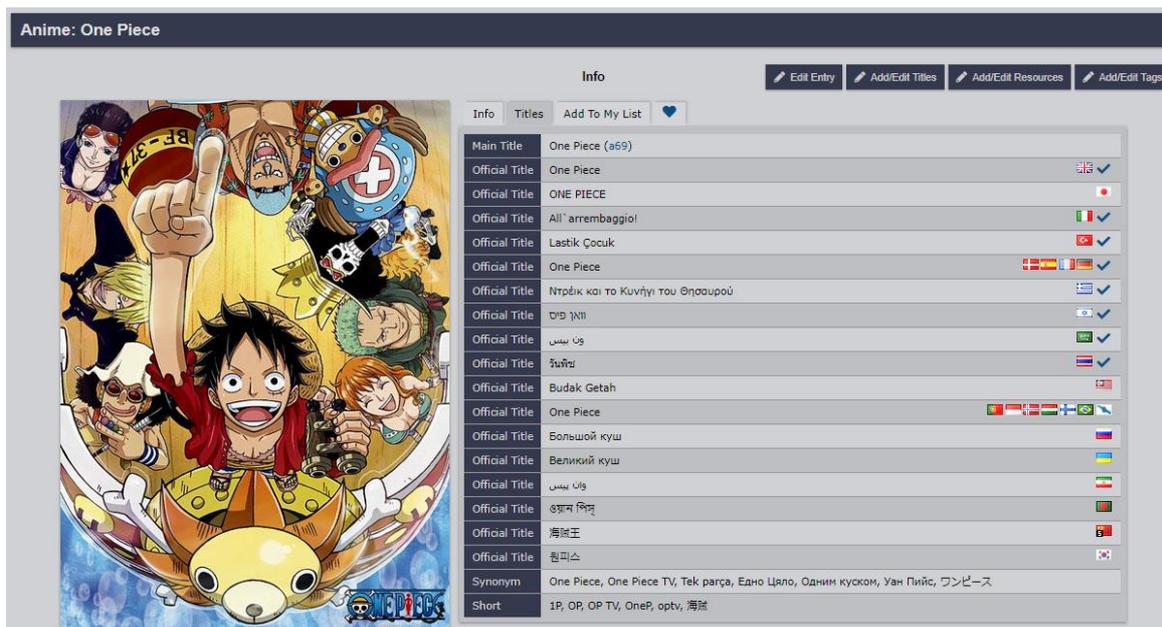


Figure 3.5. AniDB page for the *One Piece* anime illustrating the multilingual data available for the title property. Full page at <https://anidb.net/perl-bin/animedb.pl?show=anime&aid=69>

Table 3.2. Granular properties available from various data providers.

Hobbyist Provider	Sample Page Properties
MyAnimeList	<ul style="list-style-type: none"> Related Anime First Serialization (<i>Magazine name</i>) Themes Objectional content rating
AnimeNewsNetwork	<ul style="list-style-type: none"> Publication status Genres Related manga Character lists Character page links
AniDB	<ul style="list-style-type: none"> Resources (<i>External links for additional data</i>) Tags Main staff / cast Recommendations
Wikipedia	<ul style="list-style-type: none"> Related anime Demographic Chapter lists per volume (<i>en / ja</i>) Volume summaries

Mentioned in chapter 1, much of the data collection and analysis throughout the research that makes up this dissertation was focused on manga, the specific hobbyist data providers used reflected this. Main sources of manga data were sites extensive manga databases, namely AnimeNewsNetwork¹² and MyAnimeList¹³. Wikipedia / DBpedia were also used, along with some providers focused on other

¹² <https://www.animenewsnetwork.com/encyclopedia>

¹³ <https://myanimelist.net/manga.php> & <https://myanimelist.net/anime.php>

resource types, such as AniDB¹⁴. Even when the research was specifically focused on manga, the range of hobbyist providers used was fairly diverse, with data being sourced from various sub-groups within the category of hobbyist providers, some of which can be seen in Figure 3.6. With regards to the entity levels described among hobbyist providers, these also vary from provider-to-provider, with some like Wikipedia describing Superwork or Work level data, while medium-specific sites describe materials in ways more closely resembling Manifestation or Expression levels. At this point, readers need only understand the nature of institutional and hobbyist provider data and how they differ from one another, a summary of which is shown in Table 3.4; later chapters will outline the specific data available among these providers and their utility in hierarchy building and aggregation.

Ecosystem of Hobbyist Data Providers



Figure 3.6. A sample of websites that form the ecosystem of various types of hobby data providers found on the web.

¹⁴ <https://anidb.net/>

Table 3.3. Hobbyist data providers displayed in Figure 3.5.

Provider Type	Hobbyist Provider	URL
General Wikis	Wikipedia (English)	https://www.wikipedia.org/
	Wikipedia (Japanese)	https://ja.wikipedia.org/
Franchise	Bulbapedia	https://bulbapedia.bulbagarden.net
Wikis	Gundam Wiki	http://gundam.wikia.com/wiki/The_Gundam_Wiki
	Wookieepedia	http://starwars.wikia.com/wiki/Main_Page
	Ghost in the Shell Wiki	http://ghostintheshell.wikia.com/wiki/
	Dragon Ball Wiki	http://dragonball.wikia.com/wiki/Main_Page
Medium-specific	Moby Games	https://www.mobygames.com/
	AniDB	https://anidb.net/
Databases	Anime News Network	www.animenewsnetwork.com/encyclopedia/anime.php ; www.animenewsnetwork.com/encyclopedia/manga.php
	My Anime List	https://myanimelist.net/anime.php ; https://myanimelist.net/manga.php

Table 3.4. General summary of data provider types and their attributes

Provider Type	Primary Description Focus	Formality of Data Creation	Relationships	Primary Audiences	Authoritative Data
Institutional	Item, Object, Carrier	Domain or industry standards, corporate guidelines, formal ontologies	Few; Often query or keyword based; Limited semantics	Mainly patrons, users, or staff of the specific institution	Yes (agent / creator names, titles, subjects)
Hobbyist	Content	Site rules, FAQs, HTML form suggestions	Many; Resources connected both within and outside of single franchise; Semantically informative	Casual or hardcore fans, collectors, researchers, other enthusiasts	Generally no, though consensus data across many sites can exist; scope notes & other details often taken from authoritative sources

Chapter 4: The Representation of Multimedia Franchises using the Superwork Entity

As outlined in the introduction, in moving forward after the author's master's thesis research, primary goals were to not only include data from hobbyist providers as a way to better describe resources, but to also enable the inclusion of resource types beyond manga. The resulting examination of data on the web resulted in a new understanding of how MAG objects are being described – much of the descriptive data was for a conceptual entity that was seemingly outside of the scope of entities that were used in past research, namely the FRBR WEMI entities. Rather than describe individual items or works, such as single volumes of manga or a single release in a series of video games, users collect descriptions into single resources and describe series in their entirety or describe the larger multimedia franchise to which they belong. This collocating is sometimes done in the form of an entire database, such as the creation of wikis for the *Star Wars* or *Gundam* franchises, but it often takes the form of the creation of more singular entities, such as Wikipedia pages for a multimedia franchise. In addition to this being a common practice among all hobbyist communities across all media types, this method of description is unique enough from traditional bibliographic description to warrant further investigation. Importantly, if the goal is to include this data in aggregations, a model would must be capable of describing whatever entity these franchise-level descriptions could be attributed to. All of these factors led to the development of an entity – referred to as the Superwork – which seeks to be distinct from entities available in existing models, and that has been explicitly designed to represent and contain data for the multimedia franchises that these hobbyist web resources often describe. This chapter will outline the ideas behind the Superwork, what it is meant to represent, and examples of how it is currently portrayed on the web among hobbyist providers.

Encountering the Need for a Franchise-level Bibliographic Entity

The author's initial understanding of the need for an entity that could better contain some of the data that was present on the web came during the final stages of a paper written for the DCMi conference titled *Aggregating Metadata from Heterogeneous Pop Culture Resources on the Web* (Kiryakos *et al.*, 2015). In this paper, it was explained:

Classifying the more granular, manga-specific data, however, is a current work in progress; properties that describe things such as character relationships or story arcs are typically absent from traditional bibliographic description models, and thus work needs to be done to create classes and subclasses that are able to logically describe this data. (p. 70)

While the idea of a new class explicitly for franchise-level representation was not identified, it was clear that web resources, particular hobbyist providers, were describing objects in a way that did not fit well with existing models, or at least describing properties that were difficult to attribute to existing entities that were being used at the time. Descriptions on the web were for often for franchises that began as a single medium but expanded to become larger franchises. This establishment of a franchise may

maintain a dominant format, but the association between that franchise and a particular format can weaken. Henry Jenkins discusses this in his exploration of transmedia storytelling, stating that a franchise such as *Pokémon* “unfolds across games, television programs, films, and books, with no media privileged over any other” (2003, p. 2). The result is that over time, dynamic resources such as MAG objects contribute to the creation of multimedia franchises, and thus the creation of a unique entity that while not a single creative work, does refer to a network of specific and related creative works and the set of concepts that these works embody. The existence of this entity as described by users on the web led to the desire to create an entity capable of representing this level of description and containing its relevant descriptive properties and relationships. Throughout the development of the Superwork entity, guiding principles were how to differentiate it from existing conceptual bibliographic entities, how to adequately define its properties and boundaries to serve the needs of users, and how it compared it to the existing collocating entities that existed on the web via various hobbyist providers.

Defining the Superwork as Distinct from Related Entities in Existing Models

The idea of the Superwork as presented in this dissertation is an entity that can be used to represent a multimedia franchise in its entirety, free from the constraints of a single series, sub-series, medium instantiation, etc. In its most basic form, the Superwork can be understood as an umbrella entity that is the center of some number whole/part relationships. This means that the Superwork is a conceptual entity and does not refer to a single MAG object, a single series, or a single creative work, but instead a network of related works and the concepts and ideas that are associated with that network. Although there is, by design, some fluidity in determining how this network is formed and what works belong to it, some common relationships found within these networks are shared authorship, shared universe or unifying fictional world, shared characters, or a corporate branding. Flexibility in this definition and other boundaries of the Superwork are necessary to support the anticipated diverse set of users, user needs, and resource types that a franchise-level entity is expected to be used by. Because existing bibliographic models, namely FRBR and FRBR_{OO}, feature entities that are meant to represent abstract and conceptual levels of bibliographic resources, and because past research (Vukadin, 2014; Tallerås, *et al.*, 2018) has suggested such entities may be capable of portraying a collective work entity, if not a complete franchise itself, it is necessary to investigate these entities and, provided they are not able to represent the Superwork as it is discussed here, discuss their differences. This chapter also seeks to expand on the preliminary comparisons made between entities in Lee *et al.* (in press), which also discussed the Superwork in relation to entities like the Work, Complex Work, and Subject.

Superwork and the FRBR Work entity

The FRBR Work entity is defined as “a distinct intellectual or artistic creation” (IFLA, 2006, p. 17) – a definition vague enough to be open to interpretation and to allow some fluidity when determining whether a certain instance is a Work (see Zhang and Salaba, 2012). While FRBR documentation

examples do not use the Work to refer to an entity as broad as an entire multimedia franchise, some researchers have understood the definition as supporting the capability to do so. Svenonius (2000, p. 35) differentiates between the Work as “The set of all documents sharing essentially the same information” and the Superwork as “The set of all documents descended from a common origin.” Smiraglia (2007, p. 79) argues that this idea of the superwork and the FRBR Work are the same (although, he also discusses a similar concept that he refers to as bibliographic families which consist of derivations, sequels, adaptations, etc.). Vukadin (2014, p. 295) states the Work can encompass “both the totality of transmedia intellectual content and its parts: *The Matrix* as well as, for example, the film *The Matrix: Reloaded*,” though this is meant only in the context of the conceptual entity, as Vukadin goes on to state that the FRBR Work’s defined relationships would fail to “properly describe the essential whole/part nature of transmedia works” (p. 295-296).

The whole/part relationship that Vukadin emphasizes is important for distinguishing the FRBR work and the Superwork. Though the Superwork can be connected to its parts using various descriptive properties, the core of the relationship between a franchise-level Superwork and its various connected creative works is a “has part” / “is part of” relationship, a “has member” / “is member of” relationship, or more broadly, an “includes” / “is included in” relationship. The FRBR work-to-work relationship types are given as successor, supplement, complement, summarization, adaptation, transformation, and imitation (IFLA, 2006, p. 63) with no explicit whole-part or inclusion relationships given for non-aggregates or multi-part works, (though neither aggregates or multi-part works adequately represent relationships in a multimedia franchise); similar shortcomings were noted by Vukadin (2014) and Tallerås *et al.* (2018). Thus, while the FRBR Work definition offers a level of ambiguity that allows one to argue it may represent a franchise, the intended use and relationships prescribed within FRBR indicate that it does not imply an equal level of conceptuality and is inadequate at enabling the relationships that one can expect from a multimedia franchise, at least with any level of useful semantics. The types of relationships common between entities within a single franchise, such as significantly different intellectual content yet a shared same fictional world and brand, do not appear to align with what the FRBR Work is intended to represent.

Superwork and the FRBR_{OO} Complex Work entity

FRBR_{OO} is the object-oriented version of the FRBR family of conceptual models and is intended to act as an alignment with CIDOC-CRM, and a formalization of FRBR concepts, enabling better adoption and implementation, both inside and outside of the library world, using an ontological structure (Bekiari *et al.*, 2015). For this research, the main point of interest is how FRBR_{OO} attempts to define and clarify the Work concept as defined in FRBR. Though the FRBR_{OO} editors acknowledge that the Work definition as it exists in FRBR is open to interpretation, they respond to this by maintaining the vague idea of the Work as a superclass, while defining several other distinct work entities meant to represent and clarify the different interpretations of the original. These entities are the Work, Individual Work,

Complex Work, Container Work, Aggregation Work, Serial Work, Publication Work, Performance Work, and Recording Work. As previous research (Lee *et al.*, in press; Vukadin, 2014; Tallerås *et al.*, 2018, Kiryakos and Sugimoto, in press 2018a) has suggested the FRBR_{OO} Complex Work may be suitable for representing a superwork-like entity, a comparison between the Complex Work and the Superwork as we have defined it is worth investigating.

The Complex Work is defined as a class comprised of “works that have other works as members. The members of a Complex Work may constitute alternatives to, derivatives of, or self-contained components of other members of the same Complex Work” (Bekiari *et al.*, 2015, p.63). The Complex Work is associated with instances of the Work class through the “has member” property and allows for the “recognition of distinct sub-units, i.e. a complex work contained in a larger complex work” (p. 64). This definition is more closely related to the concept of a franchise-level superwork, though, like the FRBR Work documentation, the examples fall short of explicitly demonstrating this; examples given of the Complex Work (e.g. William Shakespeare’s two part play *Henry IV*, Walt Whitman’s poetry collection *Leaves of Grass*), while composed of separate parts, may imply an intention to be consumed as a single unit that a franchise-level entity does not. Members of a Superwork may not share a direct part relationship, as is the case with *Henry IV*, and may not have any intentional collection effort, as with *Leaves of Grass*. Based on the documentation, the Complex Work appears to have utility in representing distinct series or other sub-units within a franchise rather than the entire franchise itself. These findings were echoed in Lee *et al.* (in press), stating that the Japanese TV animation series *Mobile Suit Gundam* may be conceptualized as a Complex Work, but it is less suitable to make this claim for the whole *Gundam* franchise to which the show belongs.

Though Vukadin is more accepting of the use of the Complex Work, she stops short of saying it could be used as a franchise-level entity, instead stating that it provides support for the bibliographic modeling of transmedia relationships. Some support of the Complex Work is also given by Tallerås *et al.*, (2018) who view it as capable of collocating relevant works, but for fictional world levels of abstraction, defined as stories that “unfold based on (or at least referencing) a set of given characters, places, and events” (p. 901), FRBR_{OO} “does not provide any sophisticated semantics for expressing such relationships” (p. 912). This level of abstraction – and indeed, even more conceptual ones – as methods of relating creative works to form a Superwork may be not be ruled out, and so it is difficult to claim the Complex Work, as it is defined, is capable of representing every possible form of Superwork. Additionally, while it may represent a similar concept as a distinct series within a franchise, limitations on relationship types may restrict its actual usability in the modeling of related franchise entities.

Based on relevant documentation and previous research into the FRBR Work and FRBR_{OO} Complex Work, a total multimedia franchise in a way that the Superwork is defined in this research does not appear to align with the intended use of these entities. That said, FRBR_{OO} provides useful clarification and a better classification of concepts defined by the FRBR Work, and the FRBR Work is

still suitable in the portrayal of distinct intellectual creations. In other words, the concepts that they represent have roles in the creation and occasional portrayal of entities and relationship networks that the Superwork is comprised of, though the entities as currently defined may not. A generalized version of the relationship between these entities is shown in Figure 4.1, which portrays the franchise entity as the Superwork; contained within it are Complex Work entities representing distinct series within a franchise, and several Work entities representing individual mediums.

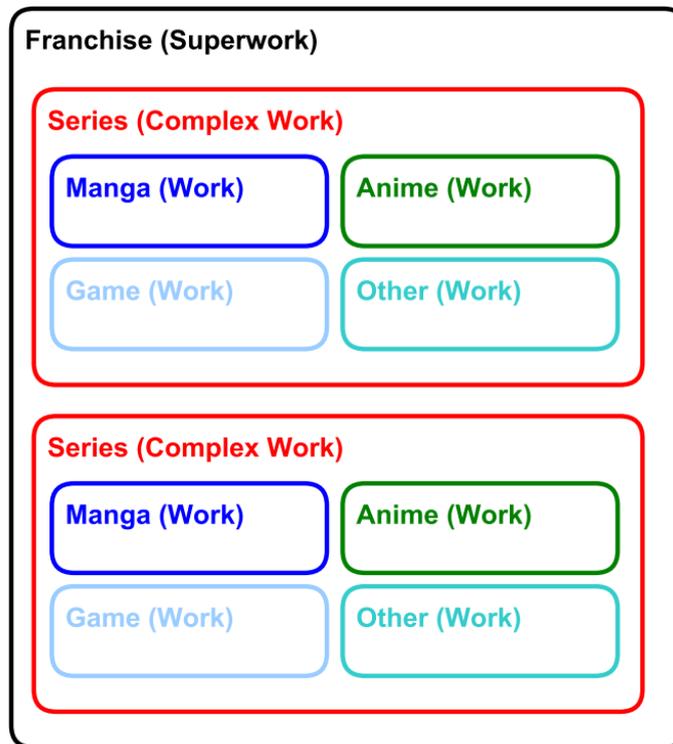


Figure 4.1. Euler diagram showing a simplified relationship between the conceptual Superwork, Complex Work, and Work entities.

A visualization of the Superwork an existing franchise can be found in Lee *et al.*, (in press), where the example of the *Gundam* franchise is used. The *Gundam* franchise refers to the Japanese science fiction / mecha multimedia franchise that began in 1979 with the animated TV series *Mobile Suit Gundam* and has since spawned numerous other TV series, films, manga, novels, plastic figures, large real-life statues, etc. This franchise consists of numerous distinct series e.g. *Gundam Wing*, *Gundam SEED*, *Gundam Unicorn*, which differ based on timelines, universes, characters, etc. but have the common theme of the featuring giant robots referred to as Gundam and belong to the same *Gundam* brand. In the conceptualization (see Lee *et al.*, in press, Figure 2) the Superwork entity, using an ontologically vague “includes” relationship, is connected to various individual components such as a TV series, film, original video animation, character, music soundtrack, plastic model kit, etc. The Superwork entity encompassing a variety of resource types from different series within the *Gundam* franchise is a rather distinct conceptual entity when compared to the Complex Work uses discussed earlier in the FRBR₀₀ documentation. Instead, the Complex Work appears to be suited to represent the series within the franchise (e.g. *Gundam Wing*, *Gundam SEED*) and any recursive Complex Work entities that belong to

them, rather than an entity representing the franchise as a whole, shown in Figure 6 as the “Series (Complex Work)” entity. That said, even representing single series using the Complex Work may be problematic, as the *hasMember* property may limit the types of relationships that can be portrayed between a Complex Work and other Works as a vertical or hierarchical one, excluding horizontal relationships (e.g. relationships between two *Gundam* series’ which will have a relation but not one that can be defined as a *membership*). Thus, while Figure 4.1 shows the conceptual relationship between the Superwork, Complex Work, and Work, an implementation meant to adequately model a franchise need not rely on these exact entities.

Superwork and Subjects

As the Superwork is meant to represent a fairly high level of abstraction, a concern is whether the Superwork is in fact an entity, or if it is instead a type of subject. In attempting to define the Superwork as distinct from a subject or theme, Lee *et al.* (in press) faced a similar challenge and questioned whether what was referred to as a Superwork “was not better conceptualized as a kind of super-topic” (p. 6) similar to a Group 3 entity from FRBR. The issue with treating the superwork as a subject was mainly a practical one; as subjects are typically the domain of various authorities (e.g. LCSH, NDLSH) “implementation would mean the insertion of countless new subject terms based on each franchise” (p. 6).

In addition to this implementation concern, a subject and a franchise-level entity may be thought of as implying two distinct concepts. It has been previously claimed in this dissertation that the Superwork is representative of a network of related works, and that it belongs to some number of whole/part or inclusion relationships. A subject, however, does not necessarily imply this type of connection, but rather describes the “aboutness” of a specific work. Thus, it may be possible for a work to be “about” a franchise without belonging to the network of resources that make up the Superwork. A news article about a particular franchise, for example, can be said to be “about” that franchise but is not necessarily a member of it. The two concepts are distinct, yet related; the Superwork is representative of some number of whole/part relationships that share a particular theme or subject, while a subject may be embodied by a group of related works that together form a franchise or Superwork; in other words, the “aboutness” of a work does not necessarily imply inclusion into the network of creative works that form the Superwork, or, at least, the network of creative works that form the unifying subject or theme that the Superwork embodies. The idea that the Superwork is related to a subject in that all works it is connected to may share and embody particular topics also helps to differentiate it from the FRBR Work, as this network of individual works and the concept that they embody are not a “distinct intellectual creation” even though they may stem from a single progenitor work.

Existing Web Entities Resembling Superwork-like Concepts

While Tallerås *et al.*'s study (2018) showed that users will create central nodes when instructed to conceptualize and map related resources, there remained the question of whether a similar practice currently occurs by the volition of users themselves, unprompted by researchers. In analyzing the FilePlanet website and examining user-made modifications for the PC game Doom, McDonough *et al.* wrote that the Superwork entity may help enabling users' preferred mode of searching (2010, p. 31), as the users were already collocating works by themselves, though it was not clear these collocations based on game engines would represent the same type of Superwork entity discussed here. Wikipedia, however, provides a larger case study, and does seem to support the idea that users will create Superwork-like entities for multimedia franchises.

Mentioned earlier in chapter 1, *The List of Multimedia Franchises* Wikipedia article¹⁵ is a collection of media franchises which, to qualify for inclusion in the article, must have “works in at least three forms of media, and must have two or more separate works in at least two of those forms of media” (para. 2). Some examples in this article include *Blade Runner*, *Jurassic Park*, *Garfield*, *Pokémon*, *Star Wars*, *Ghost in the Shell*, *Middle-earth*, *Marvel Cinematic Universe*, and *Gundam*. One trait common amongst some of these franchises is the existence of a single Wikipedia article that represents the franchise as a whole rather than a list of separate articles for each media format, though these format-specific articles do exist and are typically linked within the franchise article. For example, the articles for *Blade Runner*, *Godzilla*, and *The Matrix* have “(franchise)” directly written in their page titles, with others like *Jurassic Park* forgoing this title addition but stating in the opening sentence the article represents the total media franchise. Some examples like *Harry Potter* do not have a single franchise-level article but rather large articles for each medium (e.g. film series, novel series). Importantly, however, these medium-specific articles do discuss the entire franchise to some degree; the *Harry Potter* article is explicitly about the novels, but also features a substantial adaptations section that, in addition to linking to separate articles about the film series, video games, theme parks, etc., also describes each of them in some detail. Interestingly, the *Harry Potter (film series)* article lacks these descriptions about adaptations, and so only the article for the series of novels can be said to resemble the more explicit franchise article examples. The reason for this may be because the novels are the progenitor medium, though this is not always the case (e.g. the more detailed and franchise-representative article for *The Godfather* appears to be the article discussing the films and not the progenitor novels); further investigation on this issue is needed to determine how these decisions are made in Wikipedia.

In either of these cases, the franchise-level articles can be said to be a representation of a conceptual Superwork entity. Similar to the findings of selective attribution application by users in the Tallerås *et al.*, study (2018), these articles typically feature attributes relevant to the whole franchise

¹⁵ https://en.wikipedia.org/wiki/List_of_multimedia_franchises

rather than a specific medium (e.g. original story creator, first instantiation, unifying subjects or themes). They also connect different franchises based on abstract concepts such as crossovers or inspirations. A franchise-level entity is vital here, as works may be influenced by the themes, world, etc. present across an entire franchise and not just a single instance within that franchise, thus requiring franchise-level entities for an accurate portrayal of this relationship. Lastly, these Wikipedia entries both describe and link to a variety of articles for related series, media types, and other derivatives. If the franchise article represents the Superwork, then this network of connected articles can represent the inclusion network of instances that form the Superwork.

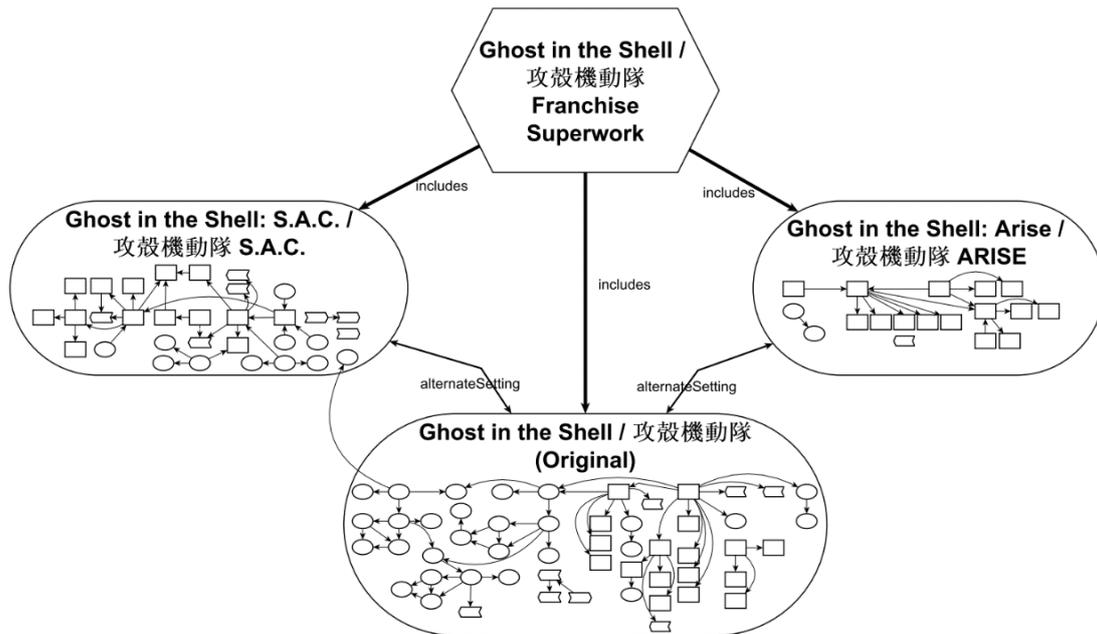


Figure 4.2. Abstract overview of entities and relationships between resources mentioned in the *Ghost in the Shell* franchise Wikipedia article (https://en.wikipedia.org/wiki/Ghost_in_the_Shell), with groupings representing distinct series within the franchise. Reprinted from Kiryakos and Sugimoto (in press, 2018b).

While the existence of franchise-level pages is arguably sufficient to claim that users on the web do create conceptual collocating entities for media franchises, an analysis of the franchise page and its relationships to other articles is helpful in revealing the relationship types that exist between different entities, between which entities certain relationships are located, and the boundaries of a Superwork through determining what types of media are included in certain pages. In attempting to visualize these relationships, a mapping of Wikipedia entities for the *Ghost in the Shell* franchise was carried out, the abstract overview of which is shown in Figure 4.2. The purpose of performing a mapping beyond the franchise page was to examine the relationships found among different Wikipedia pages related to the franchise and determine the diversity of resource types that were directly connected to either the franchise-level article or the series articles which it collocates. Figure 4.3 shows an excerpt from the Wikipedia visualization showing entities belonging to the *Ghost in the Shell* (Original) series and a series of sample relationship properties connecting them.

This series is the first instance of the *Ghost in the Shell* franchise and was originally conceived as a manga, though it now exists as multiple adaptations, of which the animated film version is shown in the figure. While the franchise-level Wikipedia article does have a link to both the *Ghost in the Shell* (1995 film) and *Ghost in the Shell 2: Innocence* entities, the various other entities linked to and mentioned in those articles e.g. novel sequels, unique releases, and soundtracks, are only featured in each respective film's article. This reflects the findings by Tellerås *et al.* (2018, p. 906-907) showing that users may not always connect every resource to the central node, but instead may connect a resource such as the music soundtrack for a movie to the movie resource itself rather than to the central node; a visualization of this indirect relationship has been reproduced in Figure 4.3. This also resembles the type of conceptual connection illustrated in Figure 4.1 supporting the use of entities resembling the Work and Complex Work as ways to intelligently organize and connect resources; in other words the Superwork need not be the entity that every single child entity is connected directly to. Though this finding is not surprising, as linking every possible media instance related to a franchise would result in Wikipedia franchise articles being overly lengthy, it does illustrate how users will collocate certain materials belonging to a single series, even if they also create a central node that is capable of having all possible resources connected to it. This action results in a content structure that resembles the simplified diagram shown in Figure 4.1, with users intelligently grouping content in logical ways while creating pages that represent conceptual entities to both contain descriptive data and encapsulate other pages based on membership or inclusion.

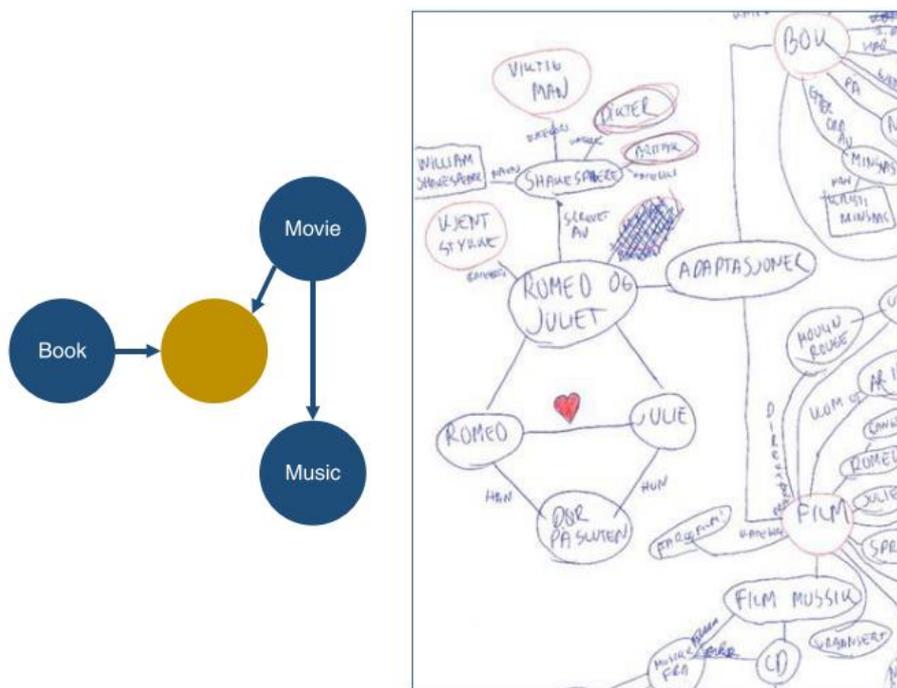


Figure 4.3. User-created concept map featuring a voluntarily generated central node showing indirect relations. Reproduced from Tellerås *et al.*, 2018 (Figure 8, p. 907).

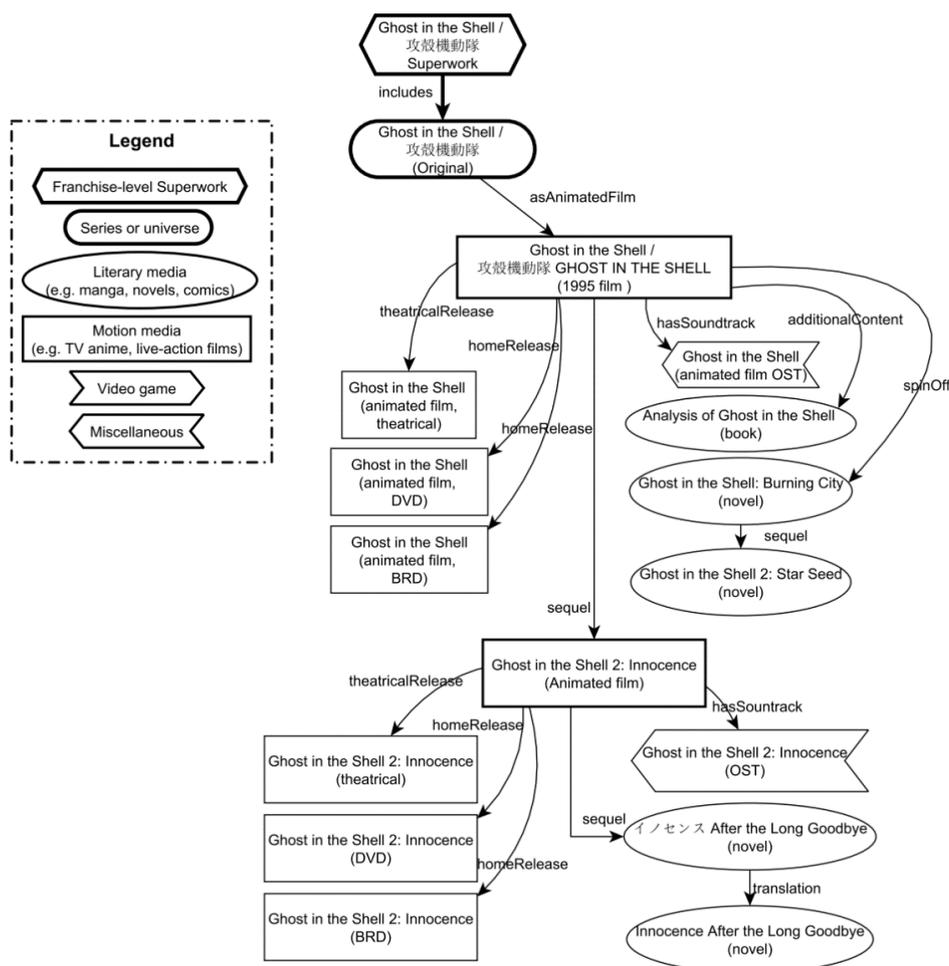


Figure 4.4. Excerpt of Figure 4.2 showing relevant entities and relationships mentioned in the *Ghost in the Shell (1995 film)* Wikipedia article ([https://en.wikipedia.org/wiki/Ghost_in_the_Shell_\(1995_film\)](https://en.wikipedia.org/wiki/Ghost_in_the_Shell_(1995_film))). Reprinted with permission from Kiryakos and Sugimoto (in press, 2018b).

The envisioning of the Superwork was based on – and made to – accommodate user-created hobbyist data, though it should be noted that there are some institutional examples of collocated series entities as well. Referring back the MADB screenshot in Figure 3.1, the *Comic Works* page can be representative of an entity resembling the Superwork or Complex Work. The page does leave out most manga-focused or medium-specific properties (e.g. publication date, physical dimensions) and contains a link to relevant related works in that other sections of the database, such as the anime work adaptation (if applicable), and so one can argue that this specific page type within MADB represents more than the “Comic Work” itself. To clarify this, however, connections between works in MADB would need to be more explicit and numerous to imply the same level of coverage as the Superwork, as currently the page type could be viewed as either the Superwork or a Complex Work-type entity representing a cross-media series, as shown in Figure 4.1. Regardless, the existence of this page type at an institutional provider makes the possibility of the future inclusion of hobbyist Superwork data more realistic, as a divided entity structure already exists and a possible location to contain such data is present.

Applicable Mediums and Boundaries of the Superwork Entity

Based on existing Wikipedia franchise articles, the boundaries on the types of media formats to warrant a direct connection to the Superwork appear to vary depending on the franchise and how important or prominent a given medium is to that franchise. Using the example of small statues or figures, any mention of them are absent from the central *Ghost in the Shell* franchise article, though they are briefly mentioned alongside other merchandise in the “Related Media – Other” section of the *Ghost in the Shell: Stand Alone Complex* page¹⁶. In contrast, “Gunpla” or “Gundam Plastic Models” are a significant aspect¹⁷ of the *Gundam* franchise and fandom, and thus hobbyists have created a separate Wikipedia page for these models¹⁸ as well as linked to this page from the central *Gundam* franchise article. The variation in importance for certain media types that exists per franchise imply that rather than attempting to prescribe any boundaries in this area, it is recommended that such definitions be left to the domains and communities themselves; this is consistent with conclusions from previous research the author has contributed to (Lee *et al.*, in press) and, due to the diverse nature of hobbyist communities, is the most flexible way to accommodate the diverse needs that different groups may have. While it may seem that not defining specific boundaries for the Superwork will result in an overly vague entity, the dynamic nature of MAG objects, multimedia franchises, and hobbyist communities on the web make attempting to define rigid boundaries an almost impossible task. Allowing users themselves to decide what new media types or derivative works are important enough to be defined within the boundaries of a Superwork concept is the most inclusive way of allowing users to incorporate data and relationships that they are interested in into a bibliographic model, as long as the entity based on that Superwork concept, as well as the model itself, are capable of adapting to this dynamic requirement.

The ability of a model to support this will be discussed in the following chapter, which details the creation of a bibliographic hierarchy incorporating the Superwork entity and its goal of better representing descriptive data for MAG objects.

¹⁶ https://en.wikipedia.org/wiki/Ghost_in_the_Shell:_Stand_Alone_Complex#Other

¹⁷ See Toys and Hobby Business sales data
<https://www.bandainamco.co.jp/cgi-bin/releases/index.cgi/file/view/8863>

¹⁸ https://en.wikipedia.org/wiki/Gundam_model

Chapter 5: The Creation of a Bibliographic Hierarchy for MAG Objects

The creation of the Superwork described in the previous chapter was a vital step in enabling the inclusion of important granular descriptive metadata for MAG objects into a bibliographic model. If the Superwork is concluded to be distinct from the Work and Complex Work entities, then its placement within a bibliographic model and its relationship to those entities – or more flexible entities representing similar concepts – is an important determination. Based on the understanding of the Superwork as a conceptual entity that can portray a multimedia franchise, this chapter details how the Superwork and other entities are used to create a bibliographic hierarchy that allows for the representation of any number of MAG or related resources. Discussion on issues surrounding existing models will be followed by a presentation of new entities – in addition to the Superwork – used to portray a bibliographic hierarchy, the relationships between these entities, and some additional discussion on relationships that a hierarchical model should enable outside of those required to build the core hierarchy itself, using relationships found among hobbyist data providers as guidance.

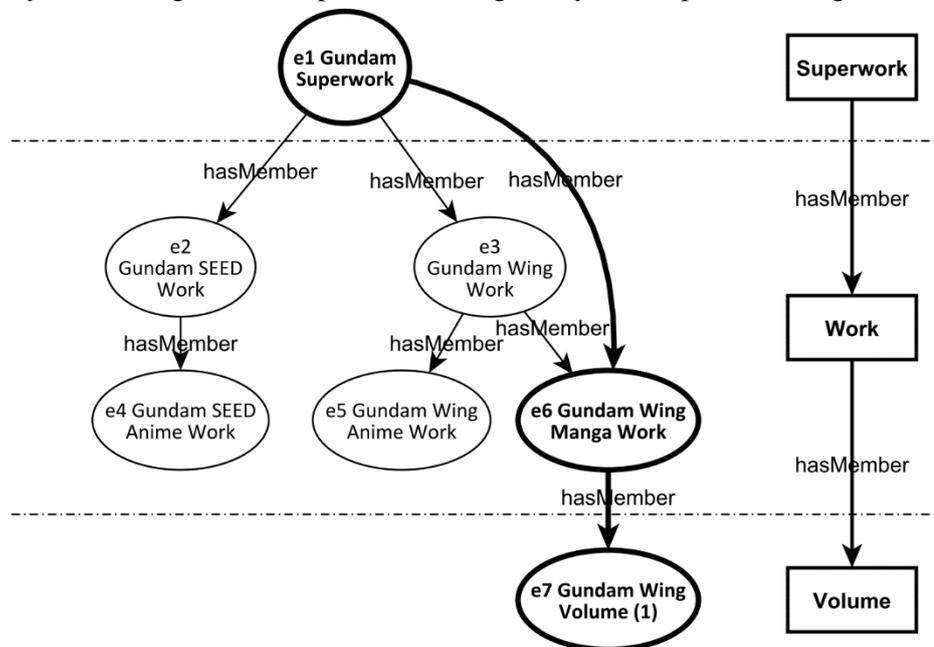


Figure 5.1. A previous iteration of the manga-centric bibliographic hierarchy (reprinted from Kiryakos and Sugimoto, in press, 2018a).

A Bibliographic Hierarchy for MAG Objects

Before an introduction of the current model, some outlining of its evolution may be helpful. An earlier version of this model was used in Kiryakos and Sugimoto (in press, 2018a) with a reprinted illustration shown in Figure 5.1. While the concept of a bibliographic hierarchy has remained similar, a separation of a layer within the hierarchy and a name change has occurred. Figure 5.1 shows the Superwork – Work – Volume connection used previously; despite there being a desire to include multiple resource types in the future, the research at the time (and the paper itself) had a focus specifically on manga. Thus, this iteration of the model does not place a series as a distinct entity separate from the Work, and

it also names the Volume specifically as the lowermost layer of the model. As research began to include other resource types, two primary changes were required.

First, because of the ubiquity of distinct series within single franchises, an intermediary entity was needed between the franchise-level Superwork and the medium-focused Work layers. The *e2 Gundam SEED Work* and *e3 Gundam Wing Work* entities are examples of such series entities, and their initial placement was also in the Work layer, as they were largely unincluded in the manga-focused discussions that were the focus of the paper, hence their lack of bolded emphasis in the figure. The inclusion of other resource types, requiring improved organization and division of distinct series from media formats, resulted in the choice to separate these two Work layer entities. Second, as the resource types expanded beyond manga, it was recommended to define the concept that the Volume layer represented in more medium-agnostic terms, as a ‘volume’ is a logical choice for manga and other literary works, but less so for anime or games, for example. Although the anticipation of entities beyond the Volume, such as an ‘Episode’ entity for anime, were expected and discussed in the study, describing the layer in a less manga-centric way is more inclusive of the variety of entities that a multimedia franchise can include. Note that this was not a fundamental model change to the model but simply a renaming, as the concepts which entities within this layer represent remain unchanged.

Based on these needs, the model evolved into one that is more readily able to support the types of entities and descriptive data that currently exist among web providers – beyond manga – and to represent multiple levels of conceptuality and instantiation for a given franchise or creative work; these features are enabled using four main layers / core entities – the **Superwork**, **Series**, **Work**, and **Object** . The generalized model is shown in Figure 5.2, with the model illustrated with examples from the *Gundam* franchise, is shown in Figure 5.4; Appendix II features additional entity descriptions and URL examples for Figure 5.4.

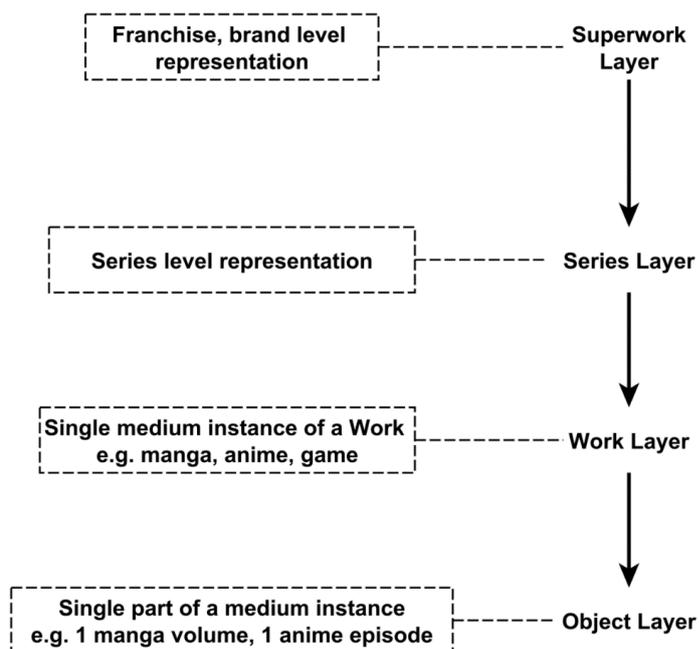


Figure 5.2. A generalized hierarchy for multimedia objects showing different entity levels required for adequate portrayal, based primarily on descriptive data from the web.

The Superwork layer is for the representation of the umbrella Superwork entity discussed throughout this dissertation. It is the highest level of abstraction and is the representative entity of an entire franchise. The Superwork is a central collocating node, allowing the application of certain bibliographic data that can also be applied to the entities below it in the hierarchy. Descriptive data such as original creators, original mediums and start dates, themes, tropes, or concepts common throughout the franchise can all be attributed to the Superwork.

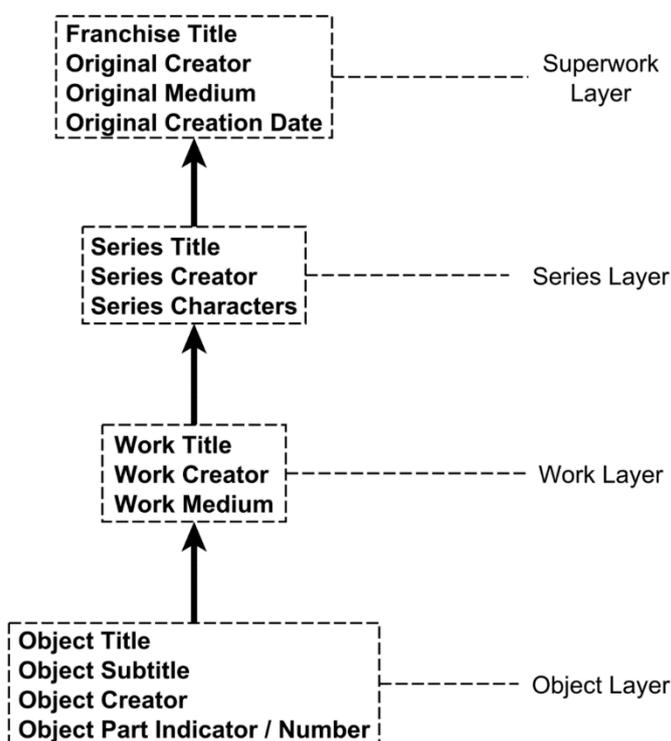


Figure 5.3. Illustration of the implied inheritance relationship between hierarchical entities using example properties.

In an ideal implementation of the hierarchy and a proper distribution of descriptive data, lower entities would have an *implied* inheritance relationship with properties from higher entities. Importantly, this means that properties attributed directly to the Superwork should be able to be logically applied to the Series, Work, and Object entities as well. This concept is also what enables the aforementioned *Comic Works* pages in the MADB (see chapter 3) to be viewed as similar to a Superwork level of description, as this page type does not include medium-specific properties that other manga pages within the MADB contain. Bennett *et*

al. (2003) summarizing Noerr *et al* also lists this type of property inheritance as a benefit to hierarchical representations such as FRBR. Referring to this inheritance as implied is done for two reasons. First, it is to indicate that this is not a type of strict property inheritance concept found in object-oriented programming. While the inheritance may eventually take this form, it would require a less fluid definition of each entity and their properties, which is not currently how they are defined. Second, the inheritance is implied because in a practical application of the model, the inherited properties may simply be implied rather than being explicitly stated. For example, the “original story creator” of the franchise Superwork would logically be the same throughout the entire hierarchy, though it would typically go unstated in the lower entities.

Figure 5.3 features an illustration of example property inheritance between entities within the hierarchy. Note that this again echoes the type of property separation found in the MADB. While aiming to decrease descriptive data workload by allowing the implied inheritance of desired properties, enabling this also encourages the logical separation of properties into their respective layers, maintaining the representative claims for each entity. For example, claims that the Superwork entity is representative of the total franchise may be contested if properties such as single mediums or part indicators such as episode or volume numbers are attributed to it. Relocating these properties to lower entities such as the Work and Object allow the Superwork to more accurately represent the franchise as it is meant to, while also providing places for data that is better attributed elsewhere. This also extends to more granular properties such as descriptions, with Superwork descriptions being broad enough to be applied to the total franchise, while Object descriptions instead being summaries or synopses of singular volumes, episodes, etc.

Additionally, various series or objects within a franchise can be directly linked to the Superwork, allowing for both the description of an entire franchise and the access of specific materials within it. The Superwork can serve a similar function to, and resemble, existing web entities such as the MADB *Comic Works* page or franchise-level Wikipedia articles. The Superwork can be used in cross-franchise relationships (e.g. inspirations, crossovers) where the franchise itself rather than a single instance is the most suitable target of the relationship. Note that while Figure 5.2 and later examples show the relationship as a linear one to the Series, entities under the Series can also be connected to the Superwork itself through an *includes* or other type of relationship, dependent on the user or community data organization preference. For the purposes of illustrating the full franchise hierarchy, linear examples are shown.

The Series layer is envisioned as similar in concept to the Complex Work entity (hence the use of the example *hasMember* properties in these figures), though without a constraint on entities it may be connected to and the relational terms that describe those connections. Mentioned previously, this entity was one of the later additions to the model that the author felt was required to fit as many franchise use cases as possible, particularly when the resources being examined expanded beyond manga. The Series was created mainly for the representation of single sub-brands, series, universes, etc., shown in Figure 5.4 as the *e2 Gundam SEED* and *e3 Gundam Wing* entities. Individual Series entities at this layer will belong to the same brand or franchise and may share certain traits (in this case, the use of giant mecha robots referred to as *Gundam* is common throughout each series), but may take place in a different universe, may have different characters, may be a different timeline, etc.

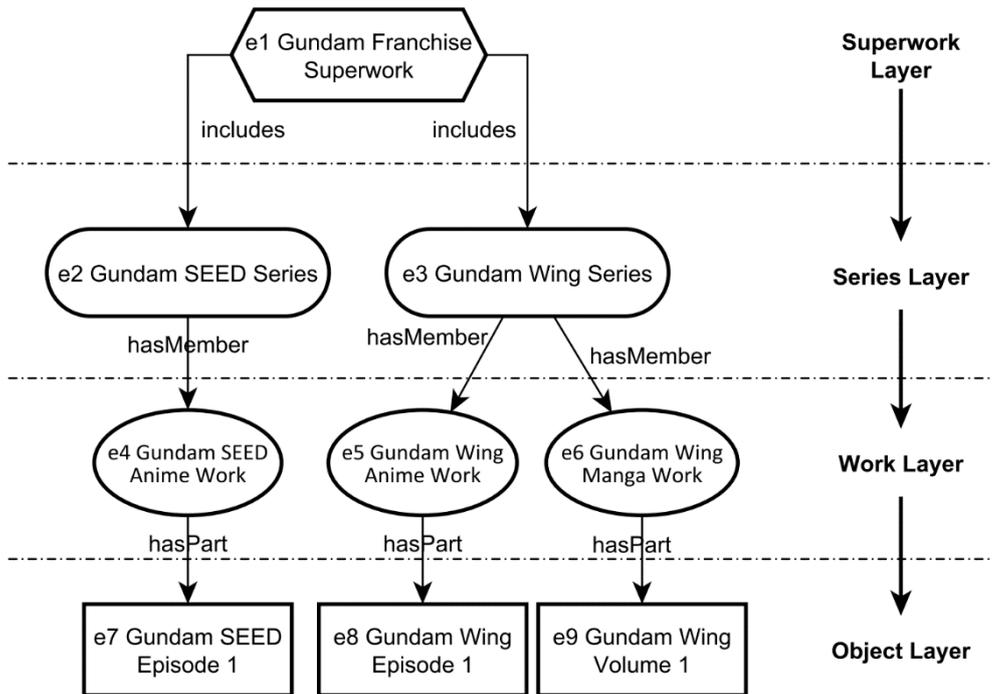


Figure 5.4. The hierarchical model with entities from the *Gundam* franchise.

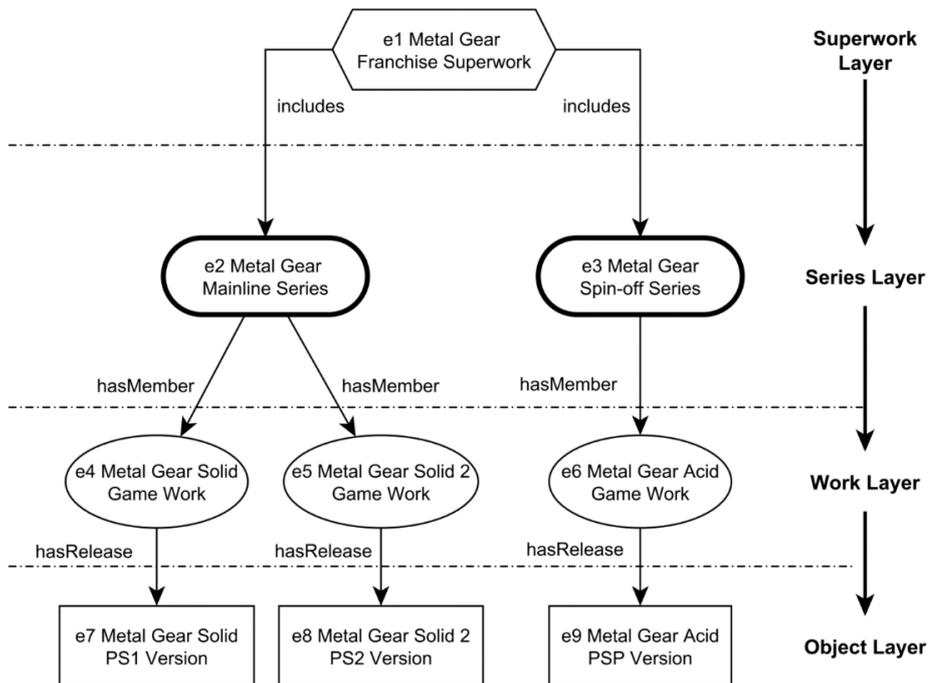


Figure 5.5. A hierarchy of a portion of the *Metal Gear* franchise illustrating the use of the Series entity to denote canonicity.

A similar though expectedly common way the Series can be used is to define some form of canonicity within a franchise, e.g. separating the mainline entries in a franchise from the spin-off works. Figure 5.5 illustrates an example of this using the *Metal Gear* series of games; like Figure 5.4, additional information regarding entities and example URLs can be found in Appendix II. This series consists mainly of two *Metal Gear* titles, with the story continued in several games in the *Metal Gear Solid* series¹⁹. Aside from specific ports of these titles, they form the mainline canonical version of the story. There are several spin-off titles, however, that share characters, locations, etc. but are not considered to be contributors to the mainline story; Series entities can be used to collect relevant works into each of these categories. A related use of the Series here can be to distinguish between “corporate works” and “fanfiction.” Past discussions that the author has been involved in were unable to resolve the issue of how to incorporate fan works, as these can be extremely significant sectors of some franchises. The use of the Series entity to indicate fan works while also including them within a franchise, should a community decide to, is a method to resolve this. Again, while the use of the Series in this way does not greatly differ from the use of multiple distinct series in Figure 5.4, the division of a franchise in this way is not uncommon, and the ability for the Series to support it and related uses should be noted.

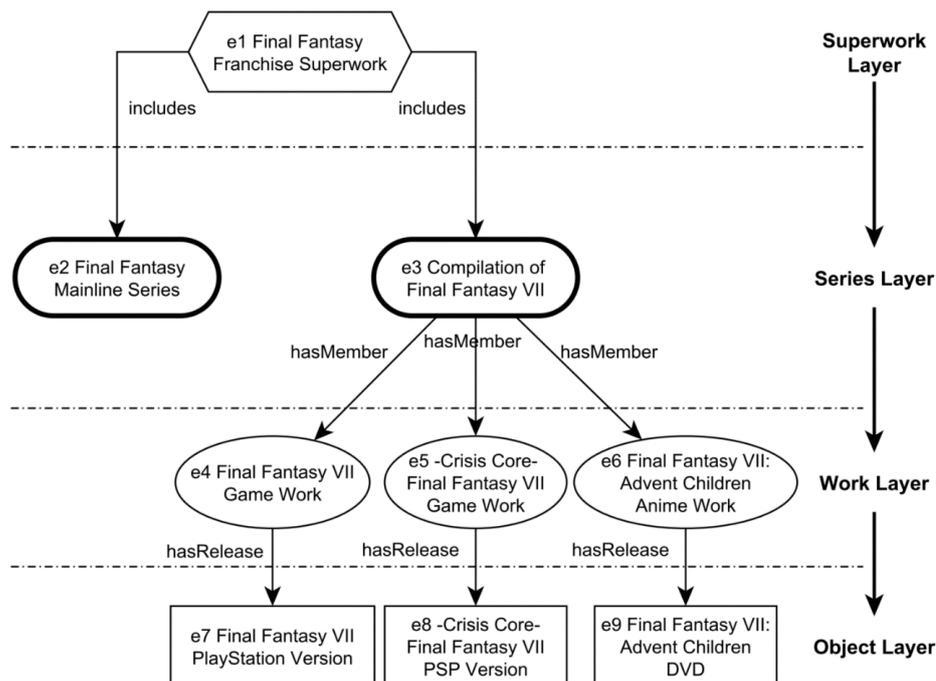


Figure 5.6. Hierarchy demonstrating a single entry in the Final Fantasy Mainline series, Final Fantasy VII, as its own Series entity.

Like the Superwork, a strict boundary on the Series entity is not defined, as different communities may hold different views on what this entity layer should represent and what type of Series is important enough to warrant distinction; an example of this can be seen in Figure 5.6. This figure shows two

¹⁹ See https://en.wikipedia.org/wiki/List_of_Metal_Gear_media

Series entities (bolded) representing the mainline *Final Fantasy* series alongside an entity for the *Compilation of Final Fantasy VII*. *Final Fantasy VII*, originally a Sony PlayStation videogame, became one of the most popular entries in the franchise, leading to a number of related follow-up works. For example, the unrelated (except by franchise) game *Final Fantasy VIII* has never received any significant follow-up media instances, whereas *VII* has had multiple spin-off games, films (one of which, *Advent Children* spawning the *Compilation* moniker²⁰), novels and related other media, and a modern remake is currently being developed. Fairly large sites dedicated to this single entry have also been established²¹ that are separate from more general *Final Fantasy* fan pages²². While some sites may view the *Final Fantasy VII* game as a single work of equal importance to the other entries in the series, others may emphasize the importance other *Compilation* and its additional works; the Series entity is meant to support either of these interpretations.

The final two entities are the Work and Object entities. The Work layer represents single media instances of a given Series, e.g. a manga, anime, or game adaptation, in their entirety, whether complete or ongoing, and agnostic of publication or edition variations. A completed 13 episode anime television show that may have a shared title and may be sold as a single release is represented as a single anime Work. Likewise, an ongoing manga publication that currently has 90 volumes and is still being produced would have a single manga Work entity, so long as all of these volumes are considered as contributing to the same manga work. Each unique episode and volume are represented as unique Object entities, which is meant for single instances or parts of a given work, e.g. single episode of an anime series, a single volume of manga, a single platform release of a video game. Note that the Object layer is meant to represent the content of a given object agnostic of publication or release variations. It represents the content of a single manga volume or video game release, for example, so long as their content is not altered significantly upon a different publication or porting to another platform. The specifics of a significant alteration are not defined here, and though this can lead to similar issues that other models face (e.g. debates regarding what FRBR entity a given resource is an instance of), such changes are typically highlighted by hobbyist providers and different entities will be created accordingly. Various examples of both the Work and Object can be seen in Figures 5.4 to 5.6.

It should be noted here that while these four entities form the core of this model, it does support additional entities should a community or institution require them. As the core entities have been created based on the nature of existing web data, some niche or institutional-centric entities may be needed by some communities. The lower-most Object layer represents a resource such a single volume of manga as it exists free from singleton publications. This means that an obvious exclusion from the core entities is one resembling the FRBR Item entity that is able to contain data such as physical dimensions or

²⁰ See https://en.wikipedia.org/wiki/Compilation_of_Final_Fantasy_VII & http://finalfantasy.wikia.com/wiki/Compilation_of_Final_Fantasy_VII

²¹ See e.g. <https://thelifestream.net/>

²² See e.g. http://finalfantasy.wikia.com/wiki/Final_Fantasy_Wiki

release dates. Though this data is rare on the web, institutions may connect an Item-like entity to relevant Object entities in order to meet their data description needs.

Entity boundary intentions and hierarchy summary

As discussed in relation to the Superwork in chapter 4, exact boundaries resulting in formal definitions are difficult to establish, as the diversity of resource types, communities, user needs, and other factors mean that a certain level of fluidity must be enabled by the entities and the hierarchy. That said, recommendations or expectations by the author as to what each entity can represent may help guide readers into better understanding the hierarchical structure and the variety of resources it is intended to represent. These have been formulated primarily based on the existing description divisions for various web resources, such as those encountered at the MADB, and while analyzing Wikipedia page networks for franchises (see Figures 4.2 and 4.4). Short descriptions and example values for each entity's *expected attribute* can be found in Appendix III and IV.

1. **Superwork.** As a franchise-level entity, candidates for the Superwork should be rather straightforward. Though identification of the Superwork is easier with larger franchises, smaller ones may still demand a Superwork entity, provided they have enough derivatives or unique works, particularly at the Series level; in other words, the existence of multiple related Series entities may be the trigger to establish the creation of a Superwork. *Expected attribute examples: Title, Original Creator, Original Creation Date, Original Medium, Genres / Tropes / Themes.*
2. **Series.** Previously mentioned, this level is intended to distinguish between distinct series, brands, universes, levels of canonicity, sub-series of prominence, or other important level of separation within a clearly defined franchise. For some franchises, this may be easier than others; *Gundam*, for example, has multiple clearly defined and distinct series which make for a straightforward separation. Other franchises may contain more subjective separations at the Series level that vary dependent on the needs of individual communities. Figures 5.5 and 5.6, for example, show a prominent sub-series and a mainline vs. spin-off distinction as warranting new Series entities, though other communities may disagree. A community dedicated to fan-fiction, for example, may see little need for a distinct canonicity Series entity, as all of their resources will inherently be non-canon. Other examples here may be when an adaptation becomes popular and diverges from the original content in its sequels; for example, sequels to the film *The Godfather* are not straightforward novel adaptations like the original film, thus the distinction between the films and novels may be done by establishing distinct series entities for both. In these and other cases, it is difficult for the author to make a recommendation that will satisfy all communities; rather, the Series entity is provided as a way to distinguish between major portions of a franchise that various communities may wish to distinguish at this level. This has the added benefit of the concept of the franchise Superwork being uniform across communities, with distinctions important to each

community being made at the Series level instead. *Expected attribute examples: Title, Creator, Original Release Date, Original Medium, Existing Mediums, Universe / Timeline, Canonicity.*

3. **Work.** The Work entity exists to represent a single medium instance belonging to a given Series. This serves an important function to not only indicate to users what mediums a particular Series has been adapted into, but also as a way to logically collect the Object entity parts into a single unit. It is the intention of the author that the several Work entities representing the same medium instances not be created unless there are significant content changes, though it is acknowledged this is a rather fluid boundary. For example, language or publication differences should not warrant separate Work entities, as the essence of the content of the Work remain unchanged between these differences. There may be cases, however, where a single medium instance changes enough across its parts to warrant new Work entities, dependent on community or organizational needs. For example, the *FLCL*²³ OVA series originally aired as 6 episodes between 2000 – 2001, with a follow-up series titles *FLCL Progressive* airing in 2018. While technically a sequel, the time lapse, and staff and character changes, may be a significant enough deviation to some to warrant treating it as a separate Work entity; this is enabled by the hierarchy. *Expected attribute examples: Title, Creator, Medium, medium-dependent properties e.g. for TV animation - Director, Original Network, Animation Studio, Licensed Networks, Original Broadcast, Total Episodes.*
4. **Object.** The Object should be used for representing parts of a Work entity. Like the Work, the intention is that content be the determining factor of Object entity creation; if the Work entity represents a manga publication agnostic of language or edition differences, then likewise the Object entity for ‘volume 1’ should represent different editions or language versions of that volume, save for significant content changes across editions. Though this decision is ultimately left to implementers, it is preferred to simply add a property (or properties) to the Object entity stating available languages rather than create a new Object for each. If such distinctions are required, additional entities outside of the four core ones (e.g. FRBR Item-like representation) can be added. A small deviation from the intention that significant content changes be made in order to spawn a new Object is present with video games, however, as the model treats different platform releases as different objects (see the Object layers in Figures 5.4 and 5.5). While the content does not always differ significantly between platform releases, the differences in required hardware required to run a video game is more significant than differences in other mediums and is typically treated as an important distinction among Hobbyist providers. This can occasionally lead to a rather large number of Objects for a single Work; for example, the original 1993 version of *Doom* exists across dozens of platforms as official and unofficial ports²⁴. Even in such cases (and perhaps even more so), platform and release differences matter enough to users to warrant this distinction at the Object

²³ see <https://en.wikipedia.org/wiki/FLCL>

²⁴ see https://en.wikipedia.org/wiki/Official_versions_of_Doom & https://en.wikipedia.org/wiki/List_of_Doom_source_ports

layer. *Expected attribute examples: Title, Subtitle / Part Title, Creator, Release Date, Medium, Format / Platform, Part Indicator.*

With the description of these entities, users should find that most web data they encounter has a logical place within the hierarchy. While fluidity in the boundary definitions is required, the fluidity should only serve to meet more niche needs and should not provide confusion as to what entity a given set of descriptive data should belong to; additional guidance will be shown in chapter 6 (see Table 6.1). Though the model is straightforward, its entities and intended flexibility enable several functions that are central fulfilling the research aims set out in this dissertation.

1. **Supporting relevant entities for MAG and related objects.** Though a not-insignificant amount of thought can (and has been) spent attempting to determine which FRBR WEMI entity a given niche resource type fits into, the four concept layers given here more readily align with those resources and the way they are typically described, from the abstract franchise concept to single objects. The model also supports the addition of other entities that various communities or institutions may desire in order to better describe a particular set of data.
2. **A hierarchical portrayal of a complete franchise.** The support of entities given in point 1 is not simply a modeling of entities based on descriptive data from web pages, but rather a method to create a bibliographic hierarchy meant to represent the way series, mediums, objects, or complete franchises exist in relation to one another. This portrayal can impact features such as search and retrieval, as users are able to, based off a single entry, view other entries in the same series, alternative media adaptations, other storylines belonging to the same franchise, and other relationships enabled by the entities and the hierarchy. For users, a hierarchical catalogue allows for faster browsing from broad searches, with the hierarchy enabling a direct path from the start of a search to the desired object of any entity level. (Noerr *et al.*, 1998).
3. **Entities meant to support granular data and semantic relationships.** The division of a franchise and its included works into various entity types as portrayed here allows for the incorporation of as much descriptive data as possible, as there exists multiple entities to which bibliographic data can be applied. Descriptive metadata belonging to a single medium may be unable to be included in a given description if that medium is not the object of the description, or more general franchise-level data may be excluded from the description of an object such as a manga volume, so as to not imply the record is representative of more than just that volume. The separation of these concepts into multiple entities allows such data to be included without confusion, and emulates some existing web providers, such as the MADB and various hobbyist sites. This separation of entities also allows for more relationships to be made amongst them, and provided vocabularies are available to do so, results in additional semantic context for users.

Relationships between core entities

Chapter 4 described the use of an *includes* property to connect the Superwork to either the Series entities or any other related franchise entities that one may wish to connect. To structure the hierarchy, Figures 5.4 to 5.6 show the use of additional example terms such as *hasMember* and *hasPart* to describe various other relationships. In looking to existing vocabularies for possible terms to describe these, some can be found among the DCMI Metadata Terms²⁵ vocabulary, such as *dcterms:hasPart*, *dcterms:isPartOf*, *dcterms:hasFormat*, and *dcterms:isVersionOf*, and *schema:exampleOfWork*, *schema:hasPart*, *schema:isBasedOn*, and *schema:isPartOf* from extensions of the Schema.org²⁶ vocabulary. While some of these may be useful for the relationships between non-Superwork level entities, the ontologically vague *includes* term may still be the preferred choice for direct Superwork relationships when one considers the breadth of resource types the entity is meant to include. Between certain entities, the *memberOf* property, which describes an intransitive meronymic part-whole relationship (Keet, 2006, p.1120), can be seen as similar in its relational implications to a *hasPart* / *partOf* property, e.g. a single manga volume can be understood as both being both a “part of” or a “member of” the manga work. As mentioned earlier, the FRBR_{OO} Complex Work entity also uses this property. Though it was discussed that a reliance on this as the only relationship type has limitations for total franchise portrayal, a *memberOf* property can adequately describe many relationships. As for the *hasPart* relationship types, though some researchers (see Prom and Habing 2002) have used the *dcterms:hasPart* and *dcterms:isPartOf* properties in a hierarchical portrayal, it may be problematic to use a *partOf* property to imply direct relation to the Superwork, as there will inevitably be entities that semantically require an explicit part relationship rather than a broader inclusion one (most obviously multi-part works). Like properties implying membership, properties indicating a “part” relationship have a role in connecting certain entities within a franchise network, though ideally these should not be used at the Superwork level. While the use of the aforementioned properties for the purpose of a generalized hierarchical portrayal are suitable, more semantically informative terms should be enabled by any implementation of the model; the use of these terms is discussed in chapter 6.

Intra-layer relationships between core entities

While the relationships between hierarchies displayed thus far have been “vertical” relationships designed to actually structure the hierarchy, the model also enables the use of “horizontal” relationships to connect entities within the same level. Figures 5.7 and 5.8 illustrate examples of such relationships. Figure 5.7 shows the same hierarchy from Figure 5.6, though with a simple connection made between with *e2 Final Fantasy Mainline Series* and the *e3 Compilation of Final Fantasy VII* entities. As was

²⁵ <http://dublincore.org/documents/dcmi-terms/>

²⁶ <https://bib.schema.org/>

stated, some may wish to portray Final Fantasy VII and its directly related works as a separate entity due to its importance, popularity, role as a progenitor of several works, or other reasons.

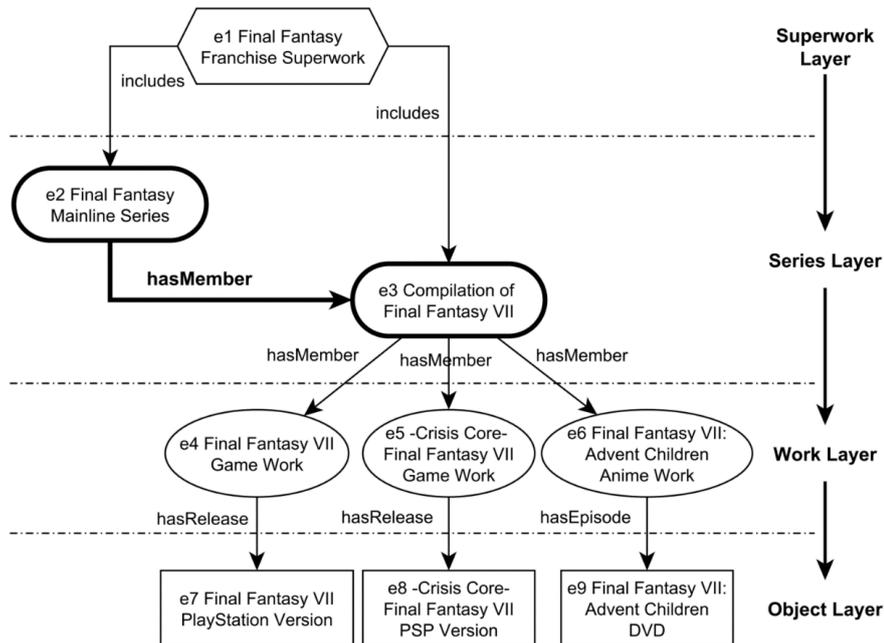


Figure 5.7. The *Final Fantasy* hierarchy from Figure 5.6, with bolded entities illustrating a Series-to-Series relationship. See explanatory paragraph for a note on accuracy of this portrayal.

Allowing for relationships, shown in Figure 5.7 as the *hasMember* relationship between two bolded Series entities, means this separation can occur while still maintaining the semantic connection stating that the *Compilation* is contained within the *Mainline* series. It should be noted that Figure 5.7 is meant only as an example of a Series-to-Series connection and not an accurate portrayal of the *Final Fantasy* franchise; because the *Compilation* includes entities besides *Final Fantasy VII* itself, e.g. *Crisis Core*, *Advent Children*, the entire *Compilation* is perhaps not a *memberOf* the *Mainline* series, but rather only *Final Fantasy VII* – the game itself – would be. The connection here is valid, though a *hasSubSeries* property, or something similar, would be more appropriate in an implementation.

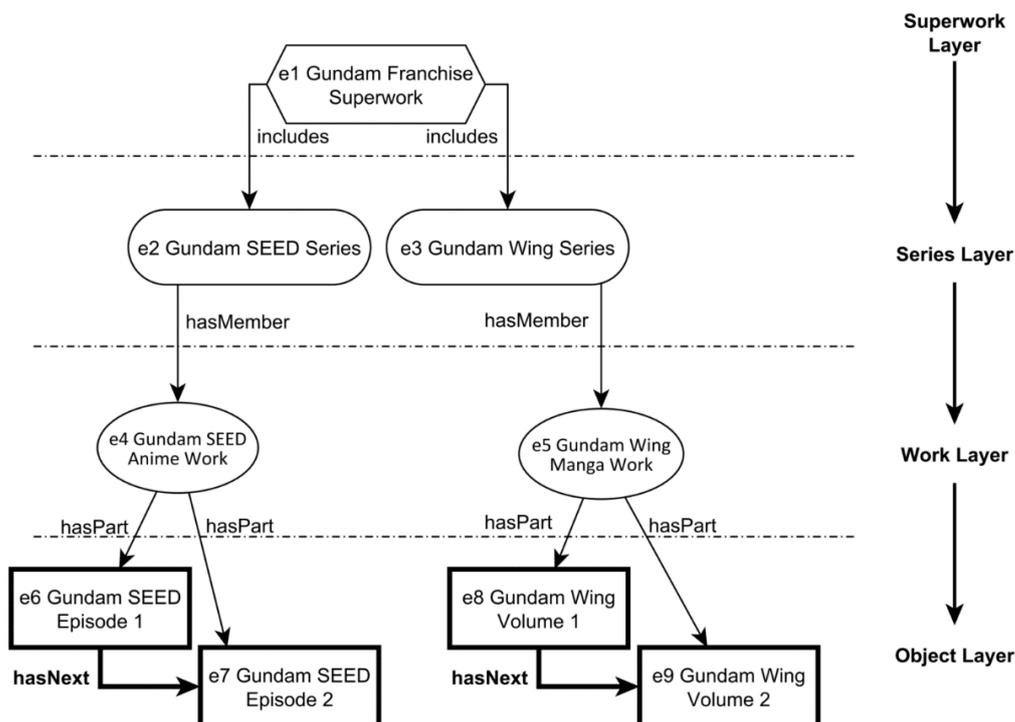


Figure 5.8. Sequential Object entities illustrated (bold), based on the *Gundam* franchise hierarchy shown in Figure 5.4.

Figure 5.8 illustrates another example of these horizontal connections, with a sequential relationship shown between bolded Object layer entities. Dynamic objects such as manga and anime will often get new volumes and episodes, which the model is required not only to contain as entities, but to place in a semantically informative order; the example *hasNext* property is a straightforward way to indicate this type of relationship. In addition to the examples in Figures 5.7 and 5.8, other horizontal relationships are enabled, with expected common types being additional Series-to-Series relationships to indicate retellings or alternative universes, or Work-to-Work relationships to communicate adaptations from one medium to another. In alignment with the goals of the model itself, these relationships are vital in enabling as much semantic context to be communicated to users as possible. These horizontal relationships will be examined again in chapter 6 on a discussion surrounding inter-franchise relationships.

Existing Semantic Relationships Found among Hobbyist Data Providers

Discussed previously in chapter 4 as a reason for pursuing the idea of a distinct Superwork, existing models typically have a predetermined set of entities and relationship types available. For FRBR, the Work-to-Work relationship types, given as successor, supplement, complement, summarization, adaptation, transformation, and imitation (IFLA, 2006, p. 63) may not be sufficient to describe the total range of possible relationship types that can be found in multimedia franchises, with Vukadin (2014)

and Tallerås *et al.* (2018) arriving at similar conclusions. The FRBR_{oo} Complex Work uses the *hasMember* property to connect to different Work entity types (e.g. Complex, Individual, Container) and does appear to be able to generally model the relationships that appear in a franchise, though based on examples in official documentation, the intention of the scope of this entity may be more limited than the Superwork as defined here. Similar thoughts appear in Lee *et al.* (in press), who state “it does not seem as though works that are not intended by the creators to be part of one another...are to be modelled using the complex work” (p. 2). Clarke (2015, as cited by Tallerås *et al.*, 2018) suggests that relationships between records and data enabled by the semantic web and linked data technologies is leading to new ways of conceptualizing data, with Tallerås *et al.* adding that this will require bibliographic models to “contribute to and enable flexible, complex semantics on different levels of abstraction” (Tallerås *et al.*, 2018, p. 912). Relating this to the Complex Work, the authors further state that the entity may be used to connect related works, but that it lacks the ability to provide sophisticated semantics for expressing relationships at the fictional world level of abstraction, which is described as a level of abstraction relating to shared contents (characters, places, events) of different documents. While a fictional world may be a component of a Work (or Works) and thus not a Work (or Superwork) itself, it does seem to be the basis for many relationships found amongst related MAG objects within a franchise, which aside from the brand, may share facets like characters, places, and events, which makes the ability to use descriptive relation types to semantically communicate this type of fictional world connection between works a useful tool in the portrayal of a franchise. In this dissertation, the author takes a similar point of view as some other researchers, which is that the Complex Work entity may be capable of connecting certain related Works and distinguishing them as different levels of abstraction, but that the entities and relationship types predefined in the model, if not disallowing the portrayal of some entity types and relationships entirely, restrict the semantics that could be communicated to users with the use of a more diverse set of entities and relationships.

As a primary reason for the creation of new entities and models is the need to better describe intra and inter-franchise relationships, it is worth exploring what these relationships may be. Appealing again to hobbyist sources on the web, a variety of relationships between entities and ways of describing those relationships are currently performed by users. While invoking the phrase MAG objects throughout this dissertation, there is an implication that a manga, for example, will have related media (adaptations or otherwise) in anime and video game mediums. Referring to the *List of multimedia franchises* article (Wikipedia, 2018b) mentioned previously, adaptations across media types for major franchises do appear to be the norm. In the list of *Franchises originating in comic books, manga, and printed cartoons*²⁷, 24 out of the 29 (82%) are listed as also having an animated TV series adaptation,

²⁷

https://en.wikipedia.org/wiki/List_of_multimedia_franchises#Franchises_originating_in_comic_books,_manga,_and_printed_cartoons

and 28 out of the 29 (97%) have video game adaptations. For *Franchises originating in video games*²⁸, 23 out of the 24 (96%) series in the list have a comic or manga adaptation, and 21 out of the 24 (91%) have either an animated TV or animated film instance. Though these numbers will be high due to the requirements of inclusion into the article (see para. 2: “a franchise must have works in at least three forms of media”), a “multimedia” franchise is expected to have works in multiple media forms, be they manga, anime, games, or other forms which the article also includes, such as live action films, literature, board games, and radio dramas, stage plays, and so on. Apart from this list-style article, more traditional articles for franchises will also detail adaptations. The *Ghost in the Shell* franchise article (Wikipedia, 2018a) lists several adaptations in both the article text itself and the Infobox, the latter of which is shown in Figure 5.9. While Wikipedia itself does not provide non-text relationship terms, some can be revealed in the DBpedia equivalent of this page²⁹, with properties such as *dbp:films*, *dbp:tv*, and *dvp:vgs* used for films, tv, and video game instances respectively. These may not provide much added semantic information but including the media type in the relation property itself is more descriptive than a more generic “adaptation” property, as it can help users or search tools that are seeking a specific adaptation medium.

Moving away from Wikipedia, it can be shown how other hobbyist providers describe such relationships. Though these other providers tend to focus on fewer media types, resulting in a lower overall number of connected entities for a given franchise, the relationships themselves are described in more semantically informative ways. One example of this can be seen in the *Related Manga* and *Related Anime* sections of series entries in the My Anime List encyclopedias. Figure 5.10 features a portion of screen captures from both pages³⁰ and more descriptive relationship types such as Sequel, Alternative Setting, and Side Story can be seen.

Ghost in the Shell	
	
Created by	Masamune Shirow
Original work	<i>The Ghost in the Shell</i>
Print publications	
Book(s)	<ul style="list-style-type: none"> <i>The Ghost in the Shell</i> <i>Ghost in the Shell 2: Man-Machine Interface</i> <i>Ghost in the Shell 1.5: Human-Error Processor</i>
Films and television	
Film(s)	<p>Animated:</p> <ul style="list-style-type: none"> <i>Ghost in the Shell</i> <i>Ghost in the Shell 2: Innocence</i> <i>Ghost in the Shell: Stand Alone Complex - Solid State Society</i> <i>Ghost in the Shell 2.0</i> <i>Ghost in the Shell: The New Movie</i> <p>Live-action:</p> <ul style="list-style-type: none"> <i>Ghost in the Shell</i>
Animated series	<ul style="list-style-type: none"> <i>Ghost in the Shell: Stand Alone Complex</i> <i>Ghost in the Shell: S.A.C. 2nd GIG</i> <i>Ghost in the Shell: Arise - Alternative Architecture</i>
Games	
Video game(s)	<ul style="list-style-type: none"> <i>Ghost in the Shell</i> <i>Ghost in the Shell: Stand Alone Complex (2004 video game)</i> <i>Ghost in the Shell: Stand Alone Complex (2005 video game)</i> <i>Ghost in the Shell: Stand Alone Complex - First Assault Online</i>
Miscellaneous	
OVA	<ul style="list-style-type: none"> <i>Ghost in the Shell: Stand Alone Complex - The Laughing Man</i> <i>Ghost in the Shell: S.A.C. 2nd GIG - Individual Eleven</i> <i>Ghost in the Shell: Arise</i>

Figure 5.9. Infobox screen capture for the *Ghost in the Shell* Wikipedia franchise article, listing various media instances. Full page at https://en.wikipedia.org/wiki/Ghost_in_the_Shell

²⁸ https://en.wikipedia.org/wiki/List_of_multimedia_franchises#Franchises_originating_in_video_games

²⁹ http://dbpedia.org/page/Ghost_in_the_Shell

³⁰ <https://myanimelist.net/manga/1023> & <https://myanimelist.net/anime/43>

Related Manga

Sequel: [Koukaku Kidoutai 1.5: Human-Error Processer](#)

Character: [Dominion](#)

Alternative setting: [Koukaku Kidoutai: Stand Alone Complex](#), [Koukaku Kidoutai: Arise - Sleepless Eye](#)

Side story: [Innocence: After the Long Goodbye](#)

Adaptation: [Ghost in the Shell](#)

Related Anime

Adaptation: [Koukaku Kidoutai: The Ghost in the Shell](#)

Alternative setting: [Ghost in the Shell: Stand Alone Complex](#), [Ghost in the Shell \(2015\)](#)

Sequel: [Ghost in the Shell 2: Innocence](#)

Alternative version: [Ghost in the Shell 2.0](#)

Figure 5.10. Screen captures of the related work sections for the *Ghost in the Shell* manga and anime pages from My Anime List. Full pages at <https://myanimelist.net/manga/1023/> and <https://myanimelist.net/anime/43/>

Fans organizing and describing these relationships in their self-created and curated resources echoes results from previously mentioned studies, such as those by Lee *et al.* (2013, 2015), which show that hobbyists have a complex understanding of these resources and that more semantic descriptive data is an important yet frequently unmet user need. These same studies discuss the benefits this can have on search and retrieval, and the explicit naming of relationship types may be a straightforward way of addressing some user queries such as “I watched x and am curious if it has a sequel or prequel”. Further examples showing how users describe related entities can be seen on AniDB³¹ which features not only named relationships for directly related works, shown in Figure 5.11, similar to the My Anime List examples, but also relationship graphs that visualize connected anime works with explicit relational terms, with an example of this for the *Ghost in the Shell* 1995 film shown in Figure 5.12.

Though predefining a set of relationships between a Superwork and other related entities, both those that make up the hierarchy shown earlier or those that may be added by other communities, may be difficult due to the number of possible resource types that a multimedia franchise can include, it may also be ill-advised as it may lead to earlier discussed problems that exist with the Work and Complex Work entities. It is clear, however, that the hierarchical model defined earlier should support the inclusion of the types of semantic relationships that are currently found among hobbyist providers, as well as any addition unforeseen relationship types that users may want to describe in the future. In determining how to establish relationships between entities to provide a sufficient level of contextual information about a given video game, McDonough (2013b) discusses the variation in user preference, and in highlighting the difficulty in relating a game to all content information objects in its environment, he states that “a given user will consider some documents as valuable context information and others

³¹ <https://anidb.net>

as irrelevant, but those judgements will vary from user to user” (p. 8). In discussing the boundaries of a superwork, Lee *et al.* (in press) highlight the need for flexibility in order to accommodate the needs of possible communities that may use such an entity. A model that allows for some variation to accommodate user needs, particularly when discussing the user needs of web data providers who may have a diverse audience, is a requirement if one is seeking to adequately meet those needs; the enabling of this flexibility as an accommodation to meet user needs is a central feature which this model has sought to support.

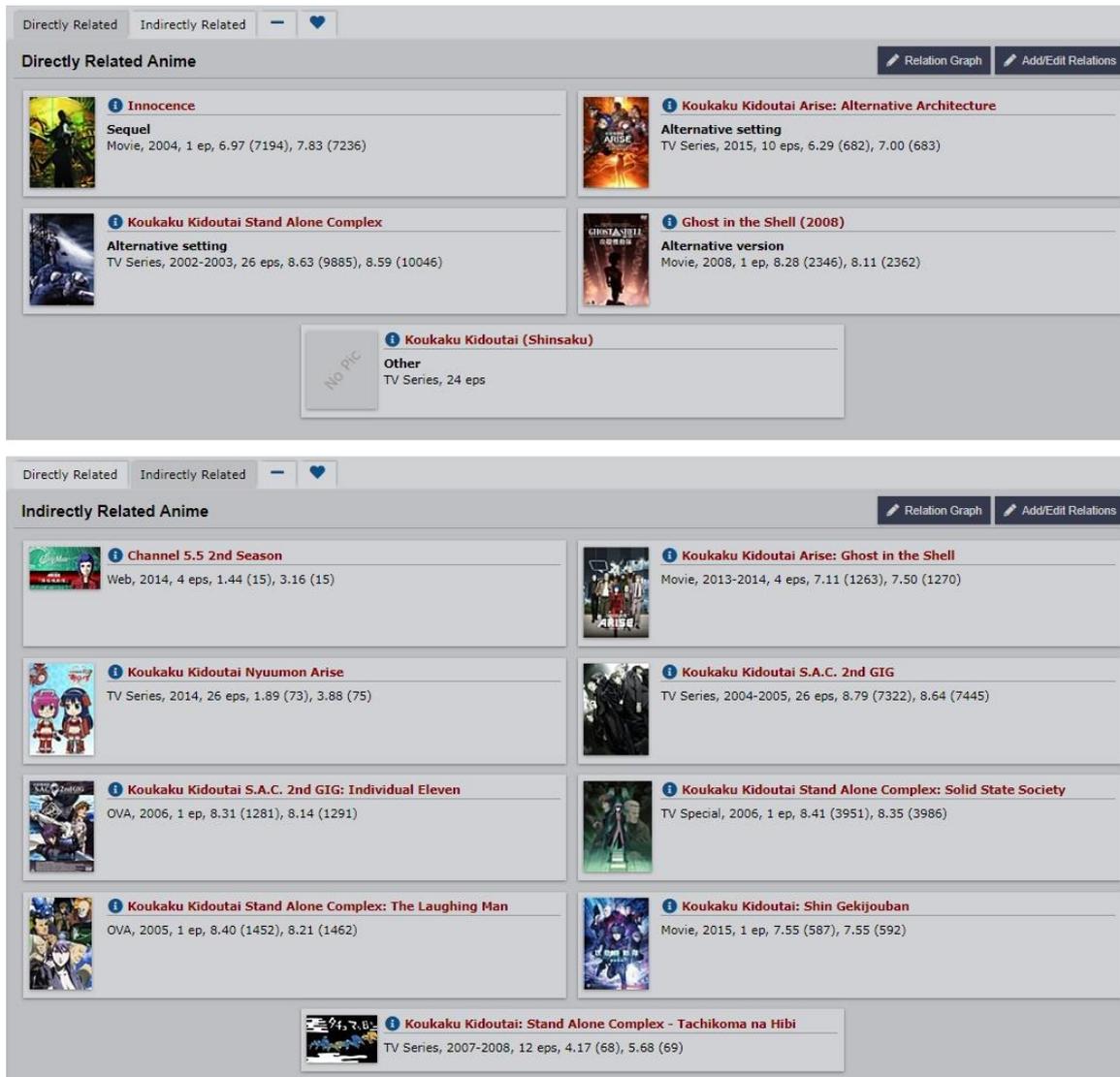


Figure 5.11. Screen capture of the AniDB entry for the 1995 *Ghost in the Shell* film showing Directly and Indirectly Related Anime, with relationship types given for the Directly Related entries. Full page at <https://anidb.net/a61>.

To summarize, this chapter has presented a conceptual hierarchical model for MAG and other multimedia objects that is meant to, particularly when compared to existing bibliographic models, provide users with a level of description that they have not only expressed as desiring, but have taken upon themselves to do on the web. This has been done through the development of distinct entities

representing levels of conceptuality matching those that descriptive metadata exists for, forming these entities into a connected hierarchy to represent the network of creative works that a franchise is made up of, and, again based on how users themselves work with such data, supporting flexible entities and relationship types to enable a high level of semantic information that the model can portray. To form the entities that make up this proposed hierarchy and gather data with which to describe them, the following chapter details the use of the OAI-ORE metadata aggregation model.

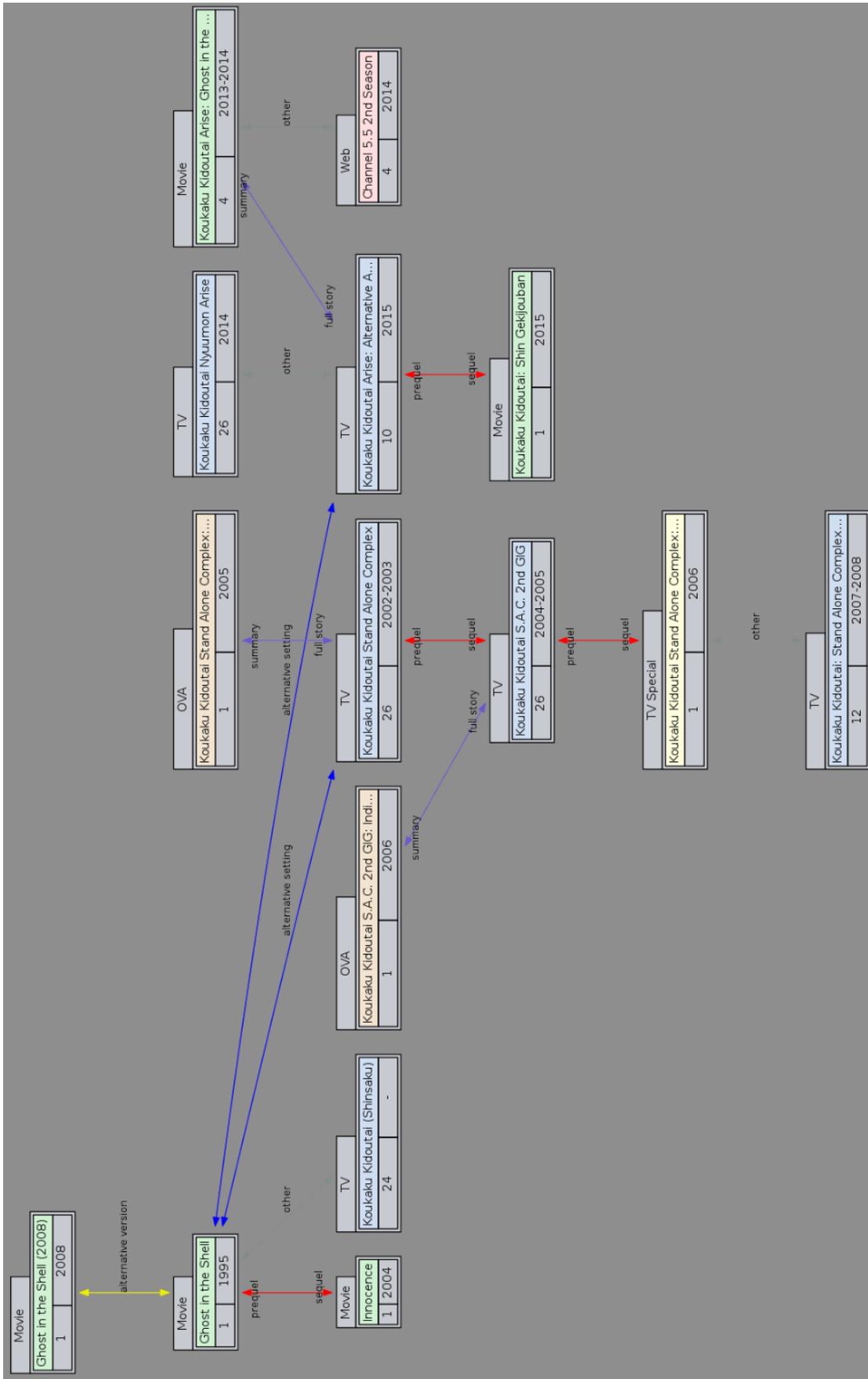


Figure 5.12. Screen capture of the AniDB Relation Graph for *Ghost in the Shell* listing various entities and their relationship type. Full page at <https://anidb.net/perl-bin/animedb.pl?show=rel&aid=61>

Chapter 6: The Aggregation of Descriptive Metadata from Institutional and Hobbyist Data Providers

First identified by the author as a possible solution to the broad issues with which this dissertation addresses in related past studies (Kiryakos, 2015), the aggregation of descriptive metadata has been the guiding principle to many of the ideas presented here. Although early research focused mainly on library data, which resulted in a simple increase in an amount of available data rather than a significant increase in granularity, it was clear that metadata aggregation was a useful way of better describing resources using already-created data. In addition to the reliance on web data, these early experiments relied mainly on the use of the Europeana Data Model (EDM), which was eventually found to be less than ideal for the purpose of working with MAG objects due to factors such as the focus on items (i.e. museum or gallery objects) rather than content or concepts. The result was a suggestion by the author (Kiryakos *et al.*, 2016) to either create an alternative aggregation model, or adopt an existing one that was more capable of working with various multimedia resource types. Thus, an investigation, and eventual use of the OAI-ORE aggregation model took place, and continued research was performed with this model as the basis. This chapter outlines this research and expands upon the benefits of metadata aggregation in fulfilling research goals, an outline of the OAI-ORE model itself, and its utilization in the creation of the bibliographic hierarchy described earlier.

Overview of the OAI-ORE Aggregation Model

Core entities within OAI-ORE

The OAI-ORE model was created in response to the lack of the ability to refer to a group of logically connected Web resources as an aggregation. For example, when wanting to refer to an entire multi-page Web document, one may use the URI of the first page. This URI does not refer to the complete set of documents, however, as it simply identifies the first page. In seeking to remedy this, OAI-ORE enables one to identify an aggregation, describe the constituents or boundaries of that aggregation, and finally assign a URI and produce a machine-readable document for that aggregation (Lagoze *et al.*, 2008b).

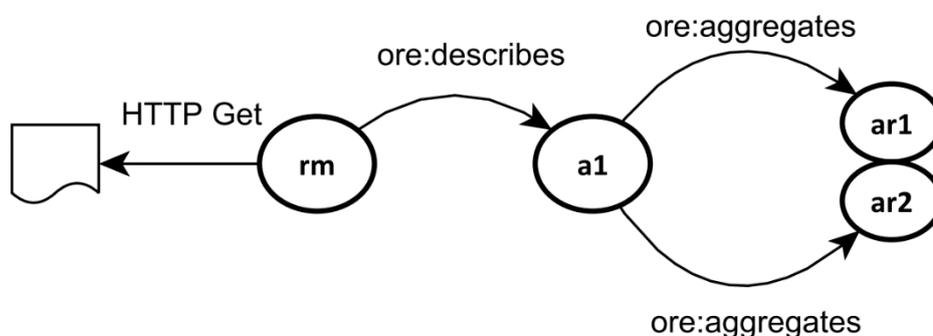


Figure 6.1. The three core entities of the OAI-ORE aggregation model (Resource map, Aggregation, and Aggregated Resource) and their relational properties. HTTP Get property added to illustrate point of access by the user.

The data model performs this using three main entities: **Aggregation**, **Aggregated Resource** and **Resource Map**, illustrated along with their relationships in Figure 6.1.

1. **The Aggregation** (*a1* in Figure 6.1) is a conceptual resource that is a set of some other group of resources that together form the Aggregation. In the context of this research, these Aggregation entities form the Superwork, Series, Work, and Object entities that form the hierarchical model outlined in the previous chapter. Each of these hierarchical entities is an Aggregation which aggregates resources that represent the entity in question.

2. **Aggregated Resources** (*ar1*, *ar2* in Figure 6.1) are the group of resources that the Aggregation is a conceptual collocation of. Importantly, OAI-ORE allows for these to be existing web resources, and for the Aggregation entity to identify them as Aggregated Resources. In other words, web pages and entries provided by existing data providers on the web act as the Aggregated Resource for this research.

3. Lastly, the **Resource Map** (*rm1* in Figure 6.1) is a resource which describes the Aggregation entity and is the location of all descriptive metadata for a given Aggregation. The Resource Map entity can be expressed in various formats, such as RDF/XML or JSON-LD, and is where users access metadata about the Aggregation and Aggregated Resources. Metadata also exists here for the Resource Map itself as a number of mandatory metadata properties, namely the authoring body for the aggregation labeled with the *dcterms:creator* property, and the last modified timestamp using the *dcterms:modified* property. The Resource Map is also used to assert metadata for the Aggregation using various preferred vocabularies.

These core entities enable the creation of the bibliographic hierarchy as defined earlier, with a visualization of this shown in Figure 6.2. In the most basic utilization of OAI-ORE to form the hierarchy, existing web resources, sourced from multiple data providers, are the Aggregated Resources (*ar1* – *ar4*) within the hierarchy. A set of Aggregated Resources deemed to be describing the same entity (e.g. Superwork, Work) form a conceptual Aggregation. In Figure 6.2, *ar1* and *ar2* can be viewed as web pages describing the same Work level entity, *a3*, while *ar3* and *ar4* are describing the same source Object, *a4*. Lastly, descriptive metadata for the Aggregation and Aggregated Resources are able to be described, distributed, and accessed via the Resource Map (not pictured in this figure). Like the hierarchical model itself, this aggregation model is relatively straightforward, yet allows for a high level of entity and relationship complexity which directly address research goals central to this dissertation. Descriptive data from multiple providers is attributed to a single aggregated entity, while those entities and the conceptual hierarchical levels that they represent can be based directly on existing portrayals, be they by hobbyists or institutions.

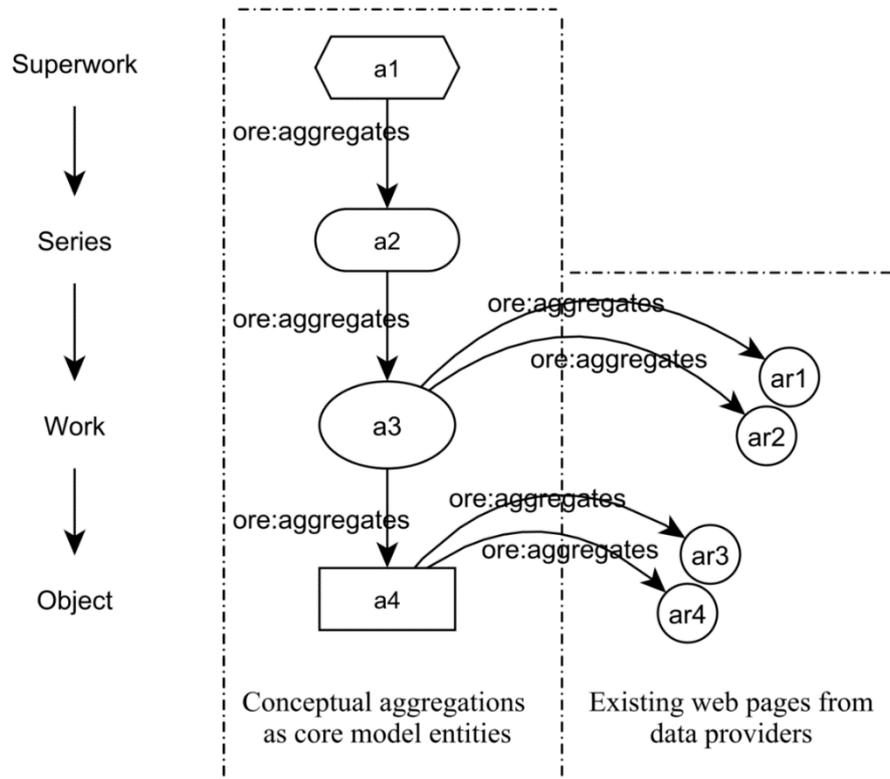


Figure 6.2. The role of Aggregations ($a\#$ entities) and Aggregated Resources ($ar\#$ entities) in the portrayal of the bibliographic hierarchy.

Nesting aggregations within OAI-ORE

One feature of OAI-ORE is that an Aggregation itself can be considered an Aggregated Resource for another Aggregation entity. This allows one to nest aggregations and create recursive relationships consisting of Aggregations of Aggregations. Figure 6.3 illustrates an example of such nesting, using the example of describing multiple language editions of a volume for the *One Piece* manga.

This usage of nested aggregations within the model allows for the easier inclusion of additional resources over time, supporting the nature of dynamic resources such as anime and manga. A use of this within the hierarchy would be to connect added Object entities (e.g. volumes, episodes) to a Work entity over time. The ability to create additional aggregations and nest them in this way also allows for different communities to describe entities in different ways, using varying levels of descriptive granularity. For example, if the Object aggregation entity for a single volume is defined as an aggregation entity representing descriptive data that is agnostic of a single publication, then publication-specific data may be attached to a separate aggregation and nested accordingly. Discussed in a previous chapter, this is an expected act by institutions that may wish to describe publication specific data resembling the FRBR Item entity. This type of nesting and hierarchical representation supported by the model allows for the benefits of hierarchical representation for multimedia works (Teixeira, 2010), while more readily supporting the flexibility for dynamic resources when compared to other existing models.

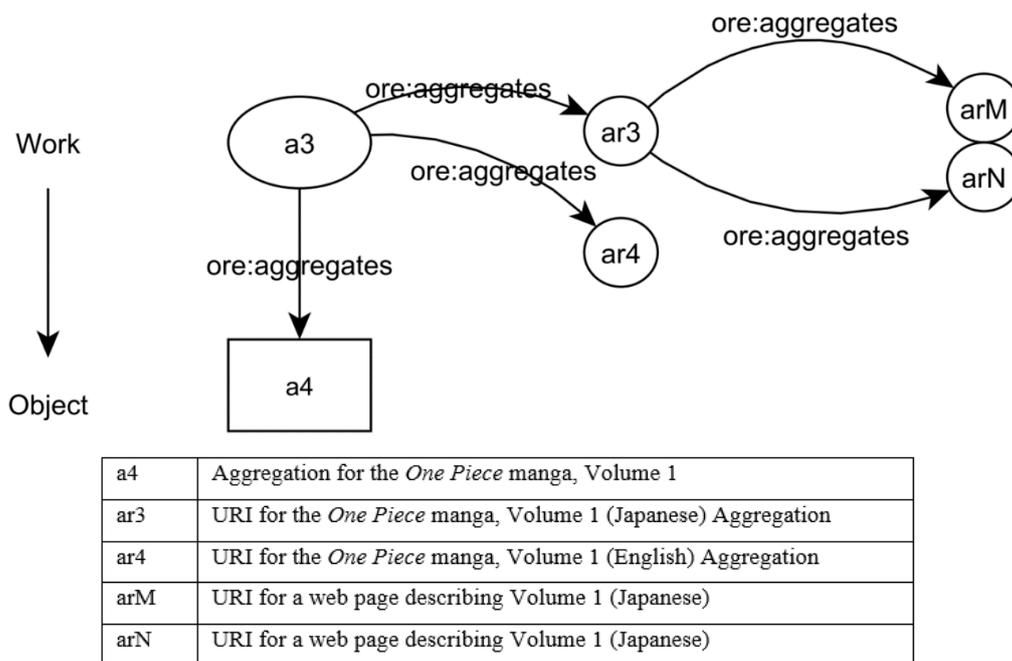


Figure 6.3. A depiction of nested aggregations as part of the *One Piece* hierarchy.

The role of OAI-ORE's Proxy resource

Another entity found within the OAI-ORE model is the Proxy, which is an “Aggregated Resource in the context of a specific Aggregation” (Lagoze *et al.*, 2008a). As typical Aggregated Resources are existing Web entities, their existing URIs have no inherent meaning that is unique to a given aggregation. Assigning a separate, unique Proxy URI to a given Aggregated Resource allows one to make an assertion about that resource that is contextually specific to a single Aggregation. Figure 6.4 shows a usage of proxies with manga – in this case using the invented *manga:nextVolume* property – to connect two proxy entities (p1, p2) representing sequential manga volumes. Though usage of the Proxy in this way is straightforward, sequential relationships such as these are both common and important in the context of resources such as manga and anime. It should be noted that the Proxy is meant for use with existing web entities that need Aggregation-specific context. As some of the Aggregated Resources within the hierarchy are nested aggregations, and thus already carry some Aggregation-specific context, the Proxy may not be needed in many cases (and it is indeed an optional entity). Additionally, with enough descriptive data relationships such as sequential order may be inferred without the use of Proxies, but the ability to use them to add semantic information within the context of a specific Aggregation makes them important to include, nonetheless. An earlier mentioned study by Ferro and Silvello (2013) showed useful non-sequential examples of Proxy usage via an “isMetadataOf” property to connect two related resources through connected proxies.

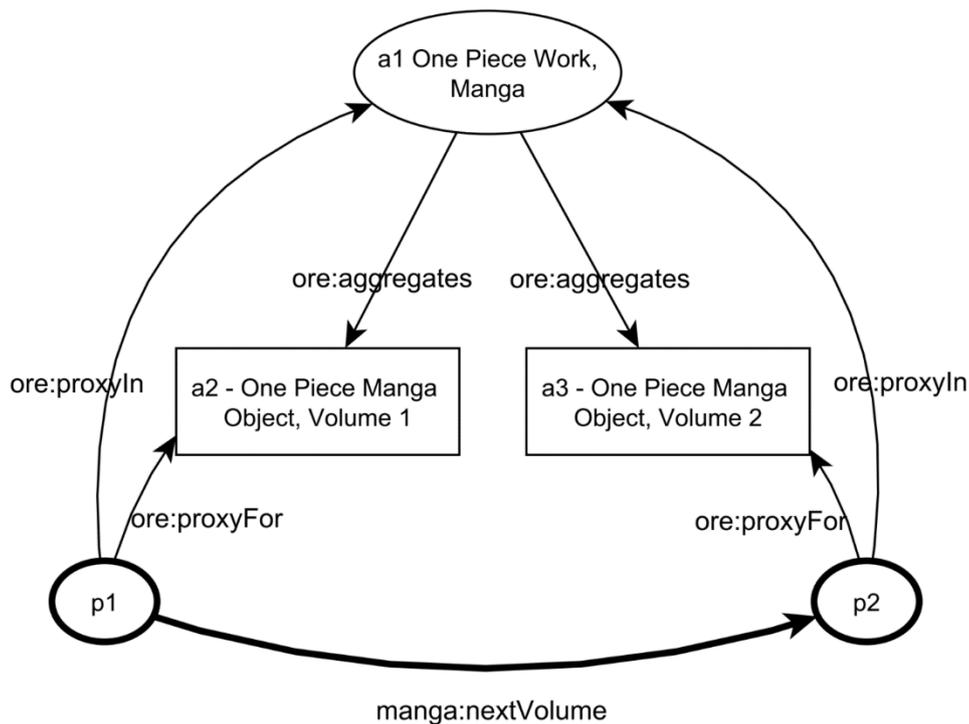


Figure 6.4. Use of the Proxy resource and a sample *manga:nextVolume* property to assert a sequential relationship between two Aggregation entities.

Determining Bibliographic Entities Based on Web Data

As the use of the OAI-ORE model relied upon the aggregation of data from numerous web providers, the entity that a given web provider's data was describing had to be identified. Particularly for the creation of the bibliographic hierarchy, it was not enough to simply identify to which series or franchise a given set of web data was describing. In addition, the core hierarchical entity that a data provider's descriptions most closely aligned to also had to be identified. The analysis process in order to make this determination differed between providers. It was typically, however, a manual process, involving the examination of descriptive properties from a number of page types available at each provider to determine which of the core entities the page in question was describing. As MADB has different named pages for different bibliographic entities³², the entity can be determined by examining the page type or URL identifier, provided these page types align with the core entities themselves. While some, such as the *Comic Works* and *Books Titles* page types for manga may align to the Superwork and Work entities respectively, this is not the case with all entities and with other mediums (i.e. anime and games) in MADB. Examining the properties themselves, which is also the main method of determining the entity being described among hobbyist providers, is a more widely applicable method. Exemplar property types that are useful identify the hierarchical entity being described by a data provider is shown in Table 6.1, column 2, with column 3 illustrating specific properties taken from various data providers.

³² See Chapter 3, *Bibliographic Data for MAG Objects as Portrayed by Libraries and Similar Institutions*

Although this determination is to be done by a user, thus leaving open the opportunity for differing interpretations as to which entity a given page is describing, this not only is in line with the goal of supporting flexibility for various communities, but occurs with applications of FRBR and related bibliographic models, as entities themselves are not always strictly defined and claims about to which entity a given resource belongs is often fluid and open to interpretations (see Zhang and Salaba, 2012). Thus, examining the properties listed on each page and using them to determine what entity the data represents was determined to be a suitable method. While future work may provide more rigid property suggestions for each entity based on a more thorough analysis of more hobbyist providers, the property types available in 6.1, along with the recommendations and descriptions made in chapter 5 and Appendix II provide reasonable boundaries for allowing users to determine the entity type a given page is describing.

Table 6.1. Property types requiring analysis in order to determine entity level and the identification of like-resources (column 2), along with example relevant website properties from multiple providers in column 3.

Model Entity	Entity Identification & Matching Resource Property Types	Website Example Properties
Superwork	Franchise Title Original Creator <i>Descriptions of multiple mediums & series</i>	Title; Name Authors; Written by; Original Creator
Series	Franchise Title Series Title Creator	Title; Name Title; Name Authors; Written by; Original Creator
Work	Franchise / Series Title Creator Creator Role Format / Medium	Title; Name Authors; Written by Directed; Story & Art; Written by Type; Release; Format
Object	Title Subtitle / Alternate Title Creator Creator Role Format / Medium Part Identifier	Title; Name Subtitle; Alternate; Part Title Creator; Authors; Written by; Directed; Story & Art; Written by Type; Release; Format Volume; Issue; Episode; Disc

Identification of Related Resources Across Data Providers

Once the entity level of a given resource has been determined, data must be examined to determine exactly what the resource in question is, i.e. to what franchise it belongs, what part of a multipart work it is, etc. The more important function this serves with regards to metadata aggregation within the OAI-ORE model is to identify resources that are of a level of similarity, dependent on the entity type, which enables their data to be logically aggregated. This level of data examination requires a deeper level of analysis than the entity identification, as more values need to be examined and then matched across providers, with the numbers of property values that require matching generally increasing the lower down the hierarchy one descends. For example, in addition to determining a property:value pair of

Format:Manga as an indicator that a page is describing a manga, a *Volume Number:[number]* property-value pair must be the same (e.g. Volume: 1) in order to aggregate data for a given manga volume. Table 6.1, column 2 shows sample property types from data providers that are useful in the identification of matching resources across the hierarchy, with column 3 featuring relevant properties from web providers. Like the identification of entity levels, this process was mainly performed manually.

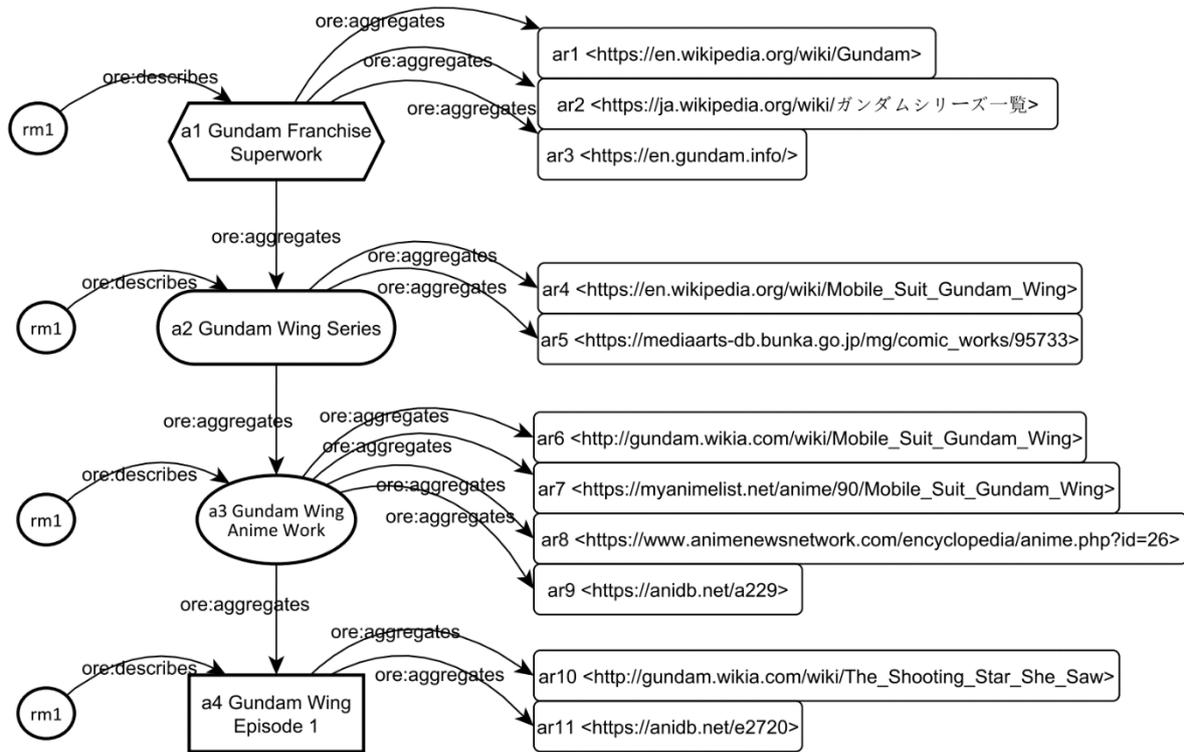


Figure 6.5. The formation of a bibliographic hierarchy for a portion of the *Gundam* franchise, with entities formed based on the descriptions of various data providers.

Portraying Existing Web Resources as a Bibliographic Hierarchy

Based on the analysis of data from various providers to first determine their respective entity level, and second to identify the actual resource to locate candidates for aggregation, the bibliographic hierarchy as it was earlier defined can be created. Using the OAI-ORE model as the basis for hierarchy formation, existing web pages act as Aggregated Resources, and matching Aggregated Resources for each hierarchical entity are collocated into a conceptual Aggregation, forming the core entities (Superwork, Series, Work, and Object) defined by the model. Figure 6.5 depicts a hierarchy for a portion of the *Gundam* franchise utilizing existing web data providers as Aggregated Resources. Aggregated Resources for the *a1 Gundam Franchise Superwork* Aggregation are sourced from Wikipedia and another general *Gundam* franchise page. Connected to this is the *a2 Gundam Wing Series* level aggregation, which also uses Wikipedia in conjunction with the MADB page as representations of the *Gundam Wing Series* without a restriction on the medium. The *a3 Gundam Wing Anime Work* entity relies on several hobbyist pages to act as its Aggregated Resources, similar to the final *a4 Gundam*

Wing Episode 1 entity which also aggregates hobbyist pages. While this is not the totality of pages that can be used for each entity, the level of granularity increase that this group of sources provide compared to single-sourced data is nonetheless significant.

The result of the organization of existing resources into this model is rather interesting and has unique implications for how pop culture objects that are a part of multimedia franchises can be portrayed. The structure of the hierarchy and the range of conceptual entities that this hierarchy covers means that descriptions for instances of various conceptual levels within a franchise, be they for representation of the franchise itself or for single instantiations of a creative work, can be logically connected to other entities to provide added context within a franchise. Additionally, with the use of existing web data acting as Aggregated Resources, the created entities do not appear to be based on arbitrary conceptualizations of layers within a multimedia franchise, but rather distinct levels of description that have been created based on consensus from how data is currently described by a number of different providers, both institutional and hobbyist. In other words, the connection of existing web resources to the conceptual aggregation entities supports the claim that the establishment of these entities are based on the way that data for multimedia objects are commonly portrayed on the web.

Providing Increased Granularity for Bibliographic Descriptions using Metadata Aggregation

In addition to using OAI-ORE to form the core entities that form the bibliographic hierarchy, the metadata aggregation functions can be used in order to provide the aggregated entities with descriptive metadata. In the context of the research goals of this dissertation, this aggregation of existing data, mainly from hobbyist providers, is used to provide a greater level of granular, semantic data across all entity types that data providers record descriptive data for. The extent to which the granularity is increased is based on several factors, such as the entity level in question, the number of data providers available, the enthusiasm for bibliographic description of the series or franchise in question (which is normally correlated with popularity) and the type of data being described. A visualizing of the increase in descriptive metadata is shown in Figure 6.6, with a list of the RDF properties and values available in Appendix I. While the Resource Map entity will not be a focus in this dissertation, this RDF data within the Resource Map entity is how the data is made accessible on the web, though as mentioned, other syntaxes such as JSON-LD are available for this purpose.

This figure shows the *al One Piece Manga Work* Aggregation entity and the amount of bibliographic descriptive data that is available from a single source, shown on the top half and provided by MADB, compared to an aggregated addition of a second source, the My Anime List page for the same manga work. Unlike prior research that dealt mainly with the aggregation of institutional providers (Kiryakos and Sugimoto, 2015), the aggregated data here is not simply greater in total amount but also granularity and utility for users, as it adds descriptive data such as synopsis, character information, and genres. Though the MADB data is useful as a provider of authoritative data, Japanese language data, and information on rare items that institutions may hold, the granular hobbyist data such as subjects and

genres, character data, and synopses directly addresses the user needs of more semantically informative data for MAG objects.

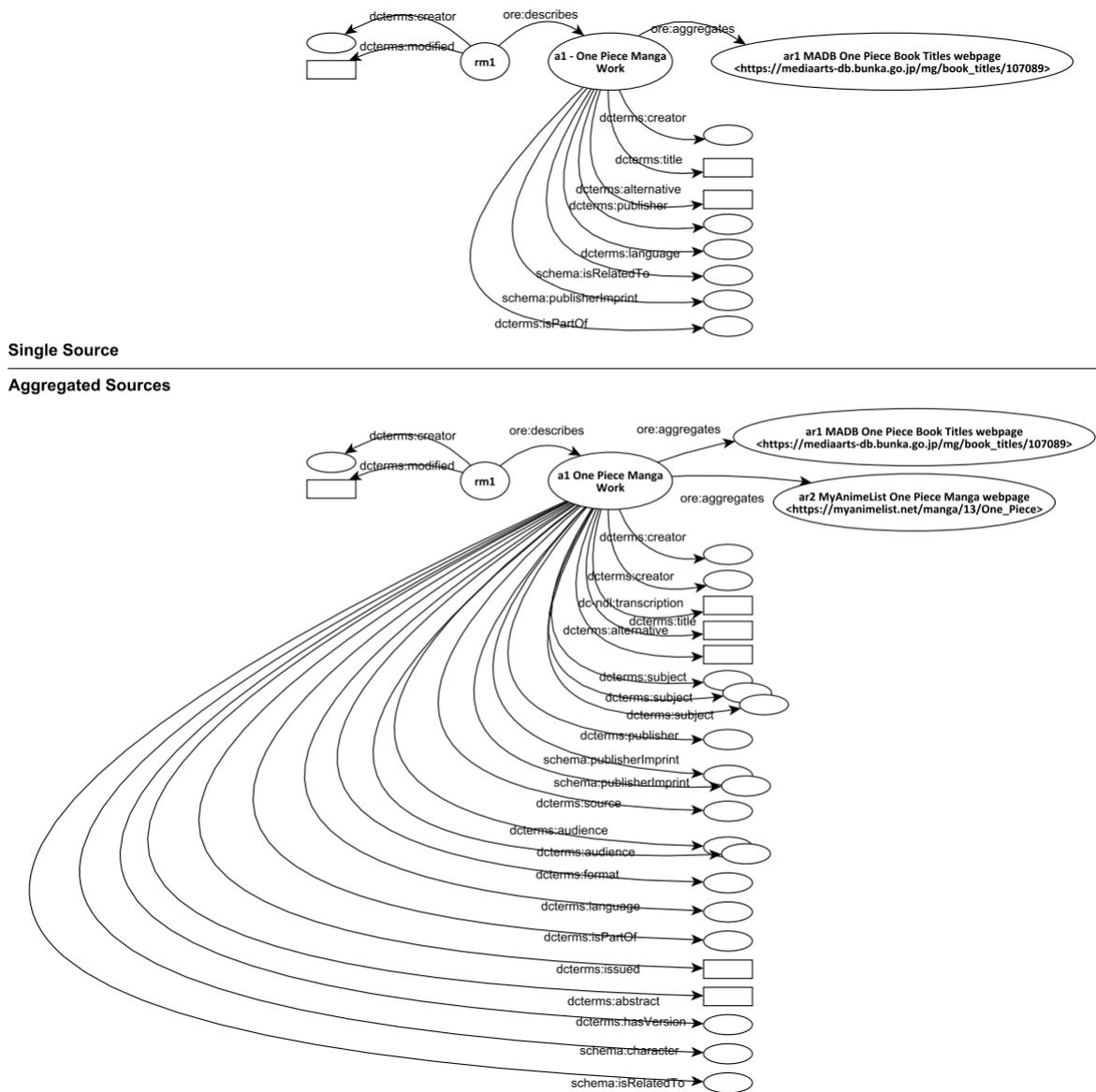


Figure 6.6. Bibliographic data available from a single source (top) along with an aggregated second source (bottom) for the *One Piece* manga Work entity. Full property list available in Appendix I.

Supporting Semantic Relationships Described by Hobbyist Providers

The OAI-ORE examples thus far have used the aggregation-to-aggregation property *ore:aggregates*, which aside from being mandatory, is the most standard method of connecting aggregation entities. For expressing different types of relationships and providing added semantic context, OAI-ORE allows for the use of relational terms from any vocabulary (Lagoze *et al.*, 2008a). Two additional properties defined for use within the model are *rdfs:seeAlso* and *ore:isSimilarTo*, though these are stated as having

specific use cases within the OAI-ORE model³³. For the purposes of describing a hierarchy and its relationships in more semantically meaningful ways, external vocabularies can be used. As stated previously, Niu (2015) also found the ability for OAI-ORE to describe relationships using open vocabularies to be beneficial for users and encouraged its use within implementations of the model.

As was discussed in chapter 5, basic terms such as *includes*, *hasMember* and *hasPart* can be used to structure core entities within the hierarchy, though as the ability to support semantically informative relationships was a key factor in decisions regarding formation, structure, and implementation of the model, relational terms beyond these may be useful. Studies by Lee *et al.* (2013, 2015) showed that despite users having a complex understanding of facets such as genres, art styles, and character types for resources like anime, existing library vocabularies were deemed inadequate in meeting user needs due to insufficient granularity, and that better described metadata supported by medium-specific controlled vocabularies were a way to address this. The metadata shown in Figure 6.6 are property mappings based on DCMI Metadata Terms, Schema.org, and DC-NDL. While DC terms is sufficient for most of the bibliographic properties from all data providers, some missing terms have been sourced from Schema.org; specifically, some granular, manga-specific properties can be found in the Bibliographic and ComicSeries vocabulary extensions. In addition, some Japanese specific properties were taken from DC-NDL, the Japanese National Diet Library vocabulary based on Dublin Core. Even though these vocabularies (aside from the ComicSeries extension) are not specifically targeted at pop culture media formats, the inclusion of granular data values with adequate properties does increase useful semantics.

Figure 6.7 illustrates a similar hierarchical model for the *Gundam* franchise shown in Figure 5.4, but with example relationship properties. Despite the limited diversity of entities in this figure, it shows that the relationship examples here are more semantically descriptive than they would be if portrayed in existing models, specifically those with restrictions on the types of relationships that can be made between entities. The *e2 Gundam SEED Series* and *e3 Gundam Wing Series* entities, if portrayed as FRBR_{OO} Complex Work entities, for example, would be limited to as *hasMember* relationship, thus being unable to communicate the (example) *alternateUniverse* connection. Differing communities may see certain relationships here as more important or relevant and may have more granular properties for these relationships that they wish to use, though the level of relationship granularity seen in the AniDB relationship chart, shown in Figure 5.12 appears to be the norm. Similar to being able to determine when and with what entities the granularity of aggregated metadata can be improved upon, the descriptive granularity of these relationship types is dependent on outside factors, such as the availability of relevant vocabularies and an adequate level of description in order to be able to infer such types of relationships (though thankfully, this level of description is typically provided). Nevertheless, the ability for the model to support this function is important, and again in support of

³³ See <https://www.openarchives.org/ore/1.0/datamodel.html#ore:similarTo>

various communities, their needs, and the improved description and portrayal of MAG and related objects.

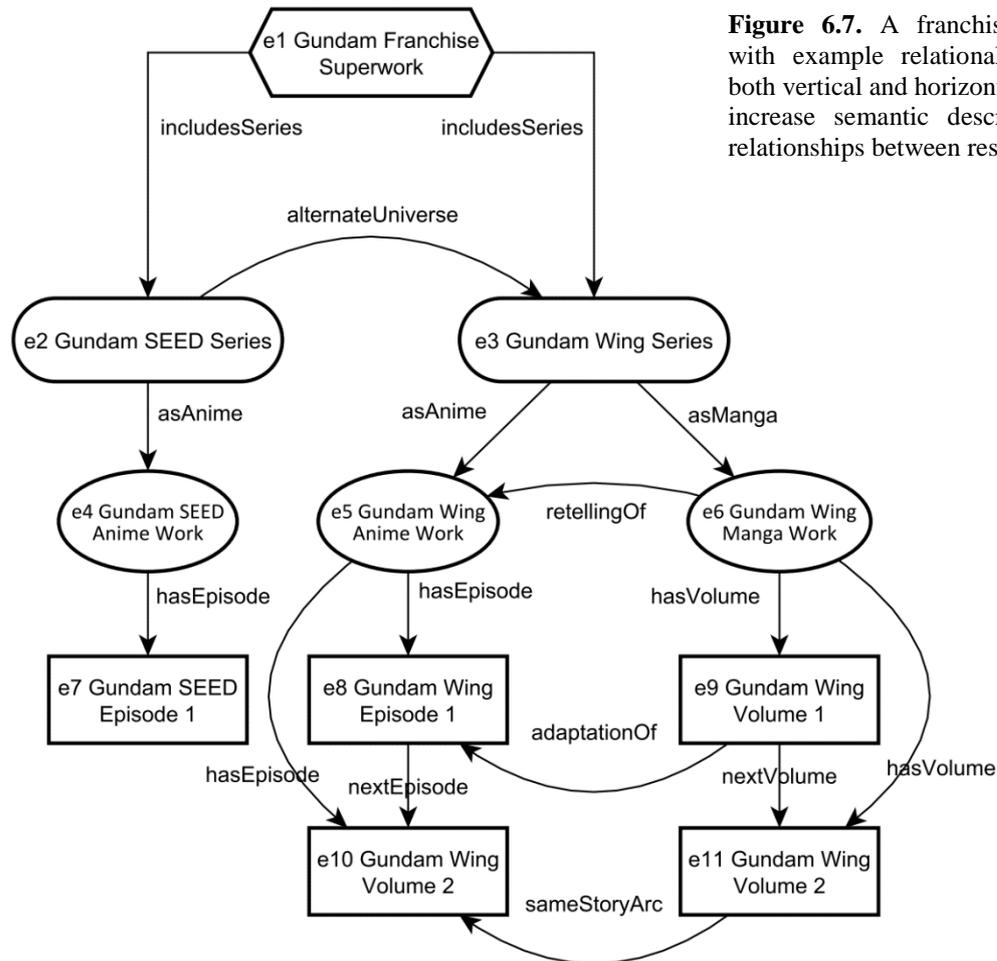


Figure 6.7. A franchise hierarchy with example relational properties, both vertical and horizontal, to enable increase semantic descriptions and relationships between resources.

While the entities given in Figure 6.7 would still include the OAI-ORE mandatory *ore:aggregates* property, the addition of these example terms enables a considerably higher level of semantic context than more basic relational terms, particular if Semantic Web technologies and the improvement to search and retrieval they can provide are considered. The use of open vocabularies in this way is one of the key functions enabled by the hierarchy and its use of OAI-ORE, with their being a near limitless possible addition of semantic data thanks to a lack of restrictions on vocabularies and resource type inclusions. Another feature enabled through the use of open relational terms is the expansion on the facets of a given resource that relationships can be based on. While often based on the content of a continued storyline, many more general relational terms such as those describing a prequel or other sequential relationship are done based on the carrier itself (i.e. *hasNext* to indicate Volume 2 follows Volume 1). The support for semantic relationship terms, such as the example of *alternateUniverse* in Figure 6.7, results in entities being connected based more on the content than the item, further appealing to user needs regarding semantic context. These again depend on vocabularies and available data, but the ability to describe relationships such as a guest character or a spiritual sequel means that users are

able to find related items of interest that would otherwise go undescribed due to less flexible models and vocabularies.

Inter-franchise relationships

Throughout the research presented in this dissertation, relationships between entities that belong to the same franchise has been the focus, mainly due to the importance of the formation of the bibliographic hierarchy that primary research goals rely upon. As detailed in chapter 5, these hierarchy-forming relationships are mainly “vertical” and are used to connect Superwork, Series, Work, and Object entities. Within a franchise hierarchy, relationships can also be “horizontal”, as shown in Figures 5.6, 5.7, and 6.7, be they to connect distinct yet related series, or for a more straightforward usage of connecting sequential entities. These relationships can also take place across different franchise hierarchies to describe various relationships such as crossovers, guest characters, or other inter-franchise exchanges, the portrayal of which was a main reason for the creation of the Superwork entity. Figure 6.8 illustrates an example of such a relationship that exists between the *Gundam* and *Dynasty Warriors* franchises.

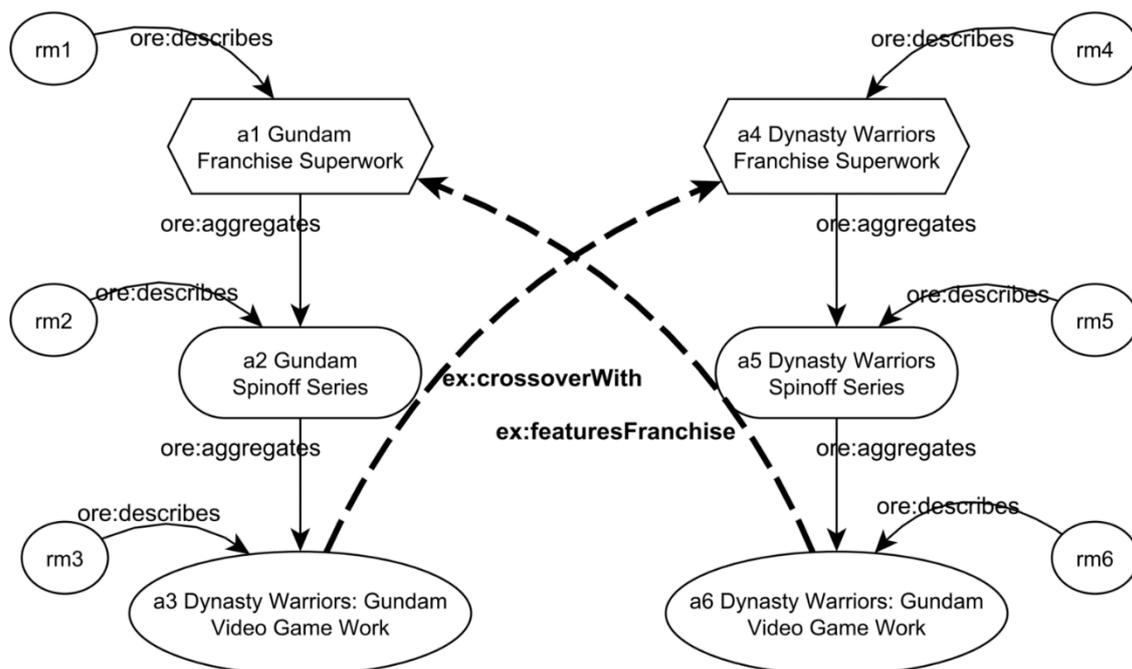


Figure 6.8. A depiction showing relevant relationships made between different franchise hierarchies, both of which feature the *Dynasty Warriors: Gundam* video game.

Entries in the *Gundam* franchise typically involves a science fiction / futuristic setting involving giant robots (“mecha”) doing battle. *Dynasty Warriors* is mainly an action video game franchise based on the historical Chinese *Records of the Three Kingdoms* tales. Apart from its mainline series, the *Dynasty Warriors* games feature many spin-off series that feature the gameplay of the mainline series, but with settings, characters, enemies, etc. from other franchises, such as *One Piece* (resulting in the game *One Piece: Pirate Warriors*), *The Legend of Zelda* (*Hyrule Warriors*), and the *Dynasty Warriors: Gundam*

game. While according to intellectual property rights, there may be a “true” franchise hierarchy to which this game belongs, this concern is not relevant for portrayal of conceptual entities, particularly with regards to how they are described by hobbyists; the *Dynasty Warriors: Gundam* game appears in fan wikis for both franchises³⁴. Thus, the placement of the same game entity in both hierarchies is supported, as are relationships across these franchises, centered on these two entities. The relational properties used in Figure 6.8 (*ex:crossoverWith*, *ex:featuresFranchise*) are again example properties, though as crossovers between franchises are not rare (see e.g. Menard, 2015, p. 12; Jenkins, 2006a, p. 96; Jett, *et al.*, 2016, p. 514), the ability of the model to support these is necessary for proper representation of relationships between franchises and provides additional semantic context and search and discovery features for users.

The Role of Authoritative Data from Institutions

Because much of the presented research has focused on granular and semantic descriptive data, hierarchical portrayals on the web, and conceptual entities like the Superwork, a majority of the discussion has surrounded hobbyist data rather than institutional. As the research moved towards higher level franchise portrayal, much of the item level data held by institutions became of limited use when it came to creating aggregated entities that would be of use to a wide group of users. The aggregated data does have its uses for institutions, such as simply providing more granular and semantic data for their audiences to more adequately meet their expectations, or possibly to complete data that is desired, yet missing, from current institutional records (see blank data values within MADB in Figure 3.1). The role of institutional data itself, however, is arguably of lesser utility to the aggregation overall, and institutions appear to have more to gain than contribute to aggregated entities; however, as was discussed in prior chapters, there does appear to be an important role for institutional data providers as contributors of authoritative data. While the intended flexibility of the model allows for multiple views and interpretations of the same entity from different providers, properties such as names or titles may have a “true” version, though determining this can be difficult due to the variation among providers. Looking at the creator of the *One Piece* franchise 尾田 栄一郎, his name may appear transliterated as Eiichiro Oda; Eiichirō Oda³⁵; Oda Eiichirou³⁶; Eiichiro ODA³⁷; Oda, Eiichiro³⁸, and with different orderings of his Japanese given and family names. Though these variable display names should be supported, authoritative data from institutions can be used within the aggregations using properties such as *skos:prefLabel*, with *rdfs:label* or *skos:altLabel* being utilized for name variations from other providers. This allows for an interesting utilization of authority data with fan created data and can enable

³⁴ *Gundam* Wiki entry: http://gundam.wikia.com/wiki/Dynasty_Warriors_Gundam

Dynasty Warriors (Koei) Wiki entry: http://koei.wikia.com/wiki/Dynasty_Warriors:_Gundam

³⁵ https://en.wikipedia.org/wiki/Eiichiro_Oda & http://dbpedia.org/page/Eiichiro_Oda

³⁶ <https://anidb.net/cr4013>

³⁷ <https://www.animenewsnetwork.com/encyclopedia/people.php?id=4639>

³⁸ https://myanimelist.net/people/1881/Eiichiro_Oda

the opportunity to align aggregated entities with other web authority sources such as the Virtual Internet Authority File (VIAF).

Summarizing the Use of the OAI-ORE Model

The decision to use the OAI-ORE model as a basis for fulfilling the objectives of this dissertation has overall been very successful. As this research constantly looks towards the diverse range of hobbyist providers and their audiences in determining how to best structure and describe multimedia resource data, a suitable base model able to conform to this diversity of needs is necessary. The flexible nature of the OAI-ORE model in the creation and portrayal of entities, the establishment of relationships between them, and their description using existing descriptive metadata from web providers, has been central in the pursuit of the research goals presented in this dissertation. With a relatively simple set of core resource types, the OAI-ORE model is able to point to existing web resources as the targets of aggregation, use matching data from these web resources to form the core entities of the developed hierarchical model, increase the amount of descriptive bibliographic data for them via metadata aggregation, and allow for the connection of any relationships with users may want to establish, be they within or outside of single hierarchies. The intended flexibility with regards to central aspects of this research such as defining entity boundaries, utilizing various data providers, describing and relating entities using various open ontologies – all of which are meant to support as wide a range of resource types, semantic data, and user needs as possible – are all readily supported by OAI-ORE, and whether any of these aspects change in their constraints or concepts moving forward, the model is intended to be capable of readily supporting such changes.

Chapter 7: Discussion on Findings, Implications, Limitations, and Challenges

This chapter outlines additional thoughts and lessons learned throughout the research that has been conducted as a part of dissertation and the studies that preceded it, various implications related to the research and the obtained results, and various limitations that were encountered by the author.

Lessons Learned Throughout this Dissertation and its Related Research

Distinguishing between the carrier and the content

One of the earliest fundamental lessons learned in the course of this research was that the differences in how institutions and hobbyist providers described data was, at its core, based on how they viewed the same resources. Whether it is a library that views an item as a physical (or digital) object that it has within its collection and must provide location and access to if desired, or a corporation that views objects as the physical (or digital) output of their business, institutional data providers typically hold the item itself to be the most important aspect of a multimedia object, whether or not they view the object in the context of its greater franchise or in relation to other objects. Hobbyist providers, and by extension the users that create the data found within hobbyist providers, have shown that they are more interested with the content contained with the items (though, collectors concerned with the building and maintenance of a collection are an exception to this). The reasons for this are as diverse and numerous as the users themselves, with some enjoying a piece of media from a genre they may be unfamiliar with and are interested in seeking out similar resources based on features like shared themes, character archetypes, art styles, gameplay mechanics, while others may be creatives themselves and are seeking inspiration based on material they have already consumed or ideas they have an interest in seeing how others have tackled. This is not to suggest that the focus on the content is a superior way of thinking about a given resources; the reasons that institutional providers are concerned with the carrier / item itself is understandable and expected based on the traditional intended use cases and services they provide. What is important in understanding the distinction between these two ways of thinking about the same object is that users seeking information on various pop culture objects have specific and diverse needs that need to be addressed if an institution seeks to adequately service them.

The importance of user behavior

The creation of bibliographic models, vocabularies, and concepts generally, have traditionally been created by researchers or other expert agents, then implemented (or imposed) on users. This is not to say that users are not considered when developing these models, but rather that experts tend to design systems based on their own ideas and understandings of user tasks, and what they perceive the optimal solutions to be. Even today this appears to be the case - Tollerås *et al.* (2018, p. 895) states that even IFLA's new Library Reference Model (LRM) builds on user task concepts that can be traced to those developed by Charles Cutter in 1904. The problem with developing bibliographic models while

considering user needs as understood by experts rather than user needs as studied by users themselves has been identified by others e.g. Zhang and Salaba, 2009; Coyle, 2016; Tallerås *et al.* 2018.

An insightful aspect of the research conducted by the author was the opportunity to examine how users themselves developed the bibliographic models and methods of descriptions which they themselves intended to use. Though one may argue that even these have constraints such as the available article templates in Wikipedia or other tools used to create fan wikis, the number and diversity of fan resources on the web makes it difficult to levy this claim against all possible hobbyist providers. Though the specifics on resource portrayal, relationship types, descriptive property focus, and so on vary from provider to provider, from community to community, and from medium to medium, the author has learned several important lessons from examination of user-created data hobbyist data that can be applied to these resource types generally. These lessons echo prior research involving various pop culture media types, with the most notable being that the more granular descriptive data that is able to be used in the description of resources, the better. The diversity of possible user needs, the enthusiasm of various fans, and the benefits that this provides in the searching of related materials means that most hobbyist providers will choose to describe more data rather than less. Though a caveat here is that more data requires more organization, lest single resources be too big for proper digestion by users, existing web data created by users (and the models developed by this research) have been shown to be logically organized in a way to prevent this issue. Figures 7.1 through 7.3 show an example of this using Wikipedia pages for the *One Piece* franchise and manga.

Media

Manga

Main article: List of One Piece manga volumes

Further information: Lists of One Piece chapters

See also: List of One Piece chapters (1–186), List of One Piece chapters (187–388), List of One Piece chapters (389–594), List of One Piece chapters (595–806), and List of One Piece chapters (807–current)

Written and illustrated by Eiichiro Oda, *One Piece* has been serialized in the manga anthology *Weekly Shōnen Jump* since July 22, 1997.^[35] The chapters have been collected into *tankōbon* volumes by Shueisha since December 24, 1997.^[36] In total, there are 914 chapters and 89 *tankōbon* volumes.^[37] Oda teamed up with Akira Toriyama to create a single crossover of *One Piece* and Toriyama's *Dragon Ball*. Entitled *Cross Epoch*, the one-shot was published in the December 25, 2006, issue of *Weekly Shōnen Jump* and the April 2011 issue of the English *Shonen Jump*.^[38] Oda collaborated with Mitsutoshi Shimabukuro, author of *Toriko*, for a crossover one-shot of their series titled *Taste of the Devil Fruit* (美食! 悪魔の実!! *Jitsushoku! Akuma no Mi!!*, lit. "The True Food! Devil Fruit!!"),^[39] which ran in the April 4, 2011, issue of *Weekly Shōnen Jump*. The spinoff series *One Piece Party* (ワンピースパーティー *Wan Pīsu Pāī*), written by Ei Andō in a *super deformed* art style, began serialization in the January 2015 issue of *Saikyō Jump*.^[40]

The *One Piece* manga was licensed for an English language release by Viz Media, who published it via chapters in the manga anthology *Shonen Jump*, since the magazine's launch in November 2002, and in bound volumes since June 30, 2003.^{[41][42][43]} In 2009, Viz announced the release of five volumes per month during the first half of 2010 to catch up with the serialization in Japan.^[44] Following the discontinuation of the print *Shonen Jump*, Viz began releasing *One Piece* chapterwise in its digital successor *Weekly Shonen Jump* on January 30, 2012.^[45] In the United Kingdom, the volumes were published by Gollancz Manga, starting in March 2006,^[46] until Viz Media took it over after the fourteenth volume.^{[47][48]} In Australia and New Zealand, the English volumes have been distributed by Madman Entertainment since November 10, 2008.^[49]

In Poland, Japonica Polonica Fantastica is publishing the manga,^[50] Glénat in France,^[51] Panini Comics in Mexico,^[52] LARP Editores and later by Ivrea in Argentina,^{[53][54]} Planeta de Libros in Spain,^[55] and Edizioni Star Comics in Italy.^[56]

Figure 7.1. Paragraph describing the *One Piece* manga on the main franchise page, with links to more granular entries. Full page at https://en.wikipedia.org/wiki/One_Piece#Manga

Volume list [edit]

Volumes 1–20 [edit]

Main article: *List of One Piece chapters (1–186)*

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 ^[4] ISBN 4-08-872509-3	June 30, 2003 ^[8] ISBN 1-56931-901-4
2	<i>Buggy the Clown</i> <i>Versus!! Bagī Kaizoku-Dan</i> (Versus!! バギー—海賊団)	April 3, 1998 ^[16] ISBN 4-08-872544-1	November 19, 2003 ^[17] ISBN 1-59116-057-X
3	<i>Don't Get Fooled Again</i> <i>Itsuwarenu Mono</i> (偽れぬもの)	June 4, 1998 ^[18] ISBN 4-08-872569-7	March 17, 2004 ^[19] ISBN 1-59116-184-3
4	<i>The Black Cat Pirates</i> <i>Mikazuki</i> (三日月)	August 4, 1998 ^[20] ISBN 4-08-872594-8	July 14, 2004 ^[21] ISBN 1-59116-337-4
5	<i>For Whom the Bell Tolls</i> <i>Tagatameni Kane wa Naru</i> (誰が為に鐘は鳴る)	October 2, 1998 ^[22] ISBN 4-08-872619-7	November 17, 2004 ^[23] ISBN 1-59116-615-2

Figure 7.2. The more granular layer which lists single volumes in the *One Piece* manga. Full page at https://en.wikipedia.org/wiki/List_of_One_Piece_manga_volumes#Volumes_1%E2%80%9320

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
<p>1. "Romance Dawn" (Romance Dawn —冒険の夜明け— "<i>Romance Dawn: Bōken no Yoake</i>")</p> <p>2. "They Call Him 'Straw Hat Luffy'" (その男「麦わらのルフィ」 "<i>Sono Otoko 'Mugiwara no Rufi'</i>")</p> <p>3. "Enter Zoro: Pirate Hunter" (「海賊狩りロロノア・ソロ」登場 "<i>Kaizoku-Gari Roronoa Zoro' Tōjō</i>")</p> <p>4. "The Great Captain Morgan" (海軍大佐「斧手のモーガン」 "<i>Kaigun Taisa 'Onote no Mōgan'</i>")</p> <p>5. "The King of the Pirates and the Master Swordsman" (海賊王と大剣豪 "<i>Kaizoku Ō to Daikengō</i>")</p> <p>6. "Number One" (一人目 "<i>Hitorime</i>")</p> <p>7. "Friends" (友達 "<i>Tomodachi</i>")</p> <p>8. "Nami" (ナミ登場 "<i>Nami Tōjō</i>")</p>			
<p>Seven-year-old Monkey D. Luffy tries to join "Red-Haired" Shanks' pirate crew, but is rejected as too young. He accidentally eats a <i>devil fruit</i> which causes his body to gain the properties of rubber, but makes him unable to swim. After an ordeal with mountain bandits, Luffy abandons his plan to join Shanks' crew; instead, he vows to 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, frees the young <i>Coby</i> from a slave's life in <i>Alvida's</i> pirate crew and saves three-sword-wielding bounty hunter <i>Roronoa Zoro</i> from being executed by the <i>Navy</i>. With Zoro Luffy's first crewman, they set sail for the <i>Grand Line</i> (the sea where the <i>One Piece</i> – the treasure of the last king of the pirates – is supposedly hidden, and meet thief (and expert navigator) <i>Nami</i>.</p>			

Figure 7.3. The most granular layer describing single manga volumes, including chapter lists. Full page at [https://en.wikipedia.org/wiki/List_of_One_Piece_chapters_\(1%E2%80%93186\)](https://en.wikipedia.org/wiki/List_of_One_Piece_chapters_(1%E2%80%93186))

Figure 7.1 shows a short excerpt describing the *One Piece* manga, choosing to place more granular data in other page entities, which the excerpt also links. The list of volumes that form the manga are given their own article, shown in Figure 7.2, with some minimal descriptive data like multilingual titles, release dates, and ISBN numbers, but again, presumably due to length (the manga has over 807 chapters at the time of writing), divides further descriptive data into additional pages while linking to them. Figure 7.3 shows this bottom layer of granular division, which in Wikipedia's case, is an article featuring the same data shown in Figure 7.2's volume article, but with additional information such as a volume synopsis and a chapter list for each volume. While the separation of data in this manner may be due to several practical concerns, such as Wikipedia's own guidelines or the simple illegibility of having an article with all of the combined data in it, it does also indicate users can engage in logical – if not complex – modeling of related data through the division and relation of various web entities. This is the across most hobbyist providers, with the visualization feature of AniDB discussed earlier (see Figure 5.12) being one of the more impressive examples. Increased granularity in entity portrayal,

bibliographic description of each entity, and relational terms that alone can provide significant semantic context, are all vital aspects of fan-created hobbyist data provider information. While it is not clear how much of this granular data needs to be included among each institutional data provider, past studies have shown a general dissatisfaction with regards to the granularity of data for these resource types, even if the institution in question is a library and the catalog is meant to mainly act as a finding aid rather than a complete information repository. It is clear that data from the web can be used in some quantity, even by authoritative institutions, to provide data for their users, whether this is caused by a lack of related materials in a collection with which individual resources can be related to for additional context, or concerns like the Tate who state that they do not have the resources to create biographies for every individual in their collection, and thus have outsourced the provision of artist biographies to Wikipedia (Bailey, 2018).

Supporting flexibility within models

A key lesson that has come to be realized over the course of this research is that the importance in using flexible and fluid concepts, particularly when dealing with niche resources, and even more so when taking user needs into account, should not be understated. Early research by the author that utilized models such as EDM and FRBR was successful in achieving the outlined goals, though they required a constant reworking of data and framing of various concepts for multimedia objects in order to work with these models – models which did not take these object types into consideration during their development. While the switch from EDM to the OAI-ORE model was done mainly for practical purposes, as it was able to better aggregate and describe data without significant ingestion, the open nature of the model e.g. the ability to connect entities in various ways and describe them using various vocabularies, resulted in an opportunity to develop new bibliographic hierarchies, describe resources using any number of existing vocabularies, and establish near infinite number relationships between related objects to provide additional semantic context. The results of the adoption of these more flexible models and methods were a change in the author's understanding of how these resources should be viewed, portrayed, and described, the extent and importance of available relationships between resources, and the ways in which models can allow users to be included in areas of bibliographic modeling and description.

Implications and Limitations of the Research and Results

Repeated throughout this dissertation has been the importance of understanding, considering, and in some cases emulating user-generated models, relationships, and data, when seeking to improve the bibliographic landscape of multimedia objects. Based on this desire, novel concepts and models have been presented which have future implications in areas such as how multimedia franchises and their entities are portrayed and how bibliographic hierarchies on the web can be established using existing resources. In addition, however, the emphasis on user-based portrayals and other creating entities and

relationships based on hobbyist portrayals resulted in several limitations as well. This section discusses both the implications and limitations of core aspects of this research.

Implications and limitations for the Superwork entity

The creation of the Superwork entity was a continuation of significant past research and has important implications in both the portrayal of a complete bibliographic hierarchy for MAG and related objects that have been the focus of multiple studies, and in the enabling of the inclusion of a wealth of semantic data, both as descriptive metadata pointing directly to the Superwork entity, and as a way to enable inter-franchise relationships. Broadly speaking, the implication of the Superwork entity is the improved support of existing collocating activities for multimedia objects that users on the web currently engage in, and the enabling of improved levels of description, entity portrayal, and relationships that can aid those wishing to better meet the needs of users of these resources.

Although the concept of the Superwork remains at the core of much continued research, the decision to place relatively few restrictions on the entity can produce problems in implementation. As stated previously, the lack of predefined relationship types and resource boundaries was a purposeful one and is meant to support the flexibility of dynamic resources, unforeseen semantic information, and diverse groups of communities and their respective needs. Though the author does feel this is a better option than attempting to define a set of boundaries or properties, conflicting views on what a Superwork is, what resources it should be directly connected to, and other not-insignificant differences would no doubt arise with an implementation of the Superwork concept across multiple data providers. Even amongst a relatively small group of users, conflicting ideas as to what a collocating node represents and the resources it should have a relationship with are not uniform (see Tällerås *et al.*, 2018, p. 906 – 907). The diversity of user groups and their needs (including amongst different fan wikis created for the same franchise) means that even domain-specific Superwork implementations may result in conflict. Admittedly, this may be an unavoidable result of the level of conceptuality and range of users this singular entity is meant to work with, though it remains to be seen how any conflicting views of the Superwork would be handled amongst communities themselves. Despite this concern, however, the utility the Superwork provides in the role of increasing semantic data and relationships, and the implications on areas such as search and retrieval that this has, mean that a franchise-level entity, particularly for multimedia franchises, should be represented by any bibliographic model hoping to adequately represent and describe such resources for the benefit of users.

Implications and limitations for the bibliographic hierarchy

The bibliographic hierarchical model defined within this dissertation succeeded in meeting the objectives of being able to represent a wide variety of resource types, sourced from a wide variety of data providers, described in varying ways by a variety of communities, both hobbyist and institutional. Analyzing bibliographic data and determining what bibliographic entity it could belong to always

resulted in an applicable entry into one of the four core entities defined, along with the expected additional entities (such as a FRBR Item-level one) that the model was created to support the inclusion of. While this model was developed based on a variety of existing ideas and concepts, e.g. existing bibliographic models such as FRBR and FRBR_{OO}, the organization of institutional data (i.e. MADB), the understanding of how hobbyists model hierarchical data on the web was again instrumental in the hierarchy's formation. Even at a mainly conceptual level (rather than a real-world implementation), the implications that this method of development had on the model enable it to sufficiently portray data as it exists across a number of different entities from both institutional and hobbyist providers.

The basing of the hierarchy on various existing structures means that there may be a bias towards the MAG resources that were the focus of much past research, though this remains to be determined. Early research focused on library data, library-centric models, and manga as a resource, which would imply that a model developed from such research may be less capable in dealing with resources that were not manga, or at least non-textual. Whether because of the close relationship within a franchise between manga, anime, and games, or the inclusion of the MADB as another primary resource, this did not appear to be the case, and developed paradigms were able to work well with resources such as anime and manga. Though not well studied here, there is little reason to think the model would be unable to be successful in working with other related resource types such as live action films or television dramas. Still, despite the goal of supporting as many resource types as possible, there may be even more niche mediums that do not conform well to hierarchy as it has been established. A note here, however, is that despite the mentioning of resources such as plastic models or other physical merchandise, data providers that describe these to the same level of MAG and related objects are rare, and thus the impact that a limited level of support for them within a hierarchy (beyond a simple entity inclusion) remains to be determined.

Another limitation here is the lack of a developed automation process designed to identify what resource and entity level a given web resource is describing. While the identification of entity levels based on descriptive data was performed manually, as research has progressed, a more well-defined set of properties that can be used to determine which of the defined hierarchical entities a web resource is describing has been developed. Despite the fluid boundaries of some of the entities, manual property analysis did not result in conflicting entity application, and so the evolution of said property list has helped the research move forward regarding this issue. Nevertheless, any practical implementation of the hierarchy would necessitate an automated method of performing this task, whether or not the defined list of properties used to identify entities was changed.

Implications and limitations in the use of the OAI-ORE model

As with the Superwork entity and the bibliographic hierarchy, the demonstrated flexibility of the OAI-ORE model and its ability to fulfill specific tasks in pursuit of the research goals is an important takeaway from this dissertation. This research illustrated that the relatively modest core resource types

and features of the OAI-ORE model can be used in more complex implementations – in this case, the creation and description of a bibliographic hierarchy. While the use of the OAI-ORE model for aggregating existing web, resources is main and stated intended use of the model, the creation of bibliographic hierarchies in this way is not, and this research has demonstrated that the entities and relationship structure provided by the model are capable of doing so. This has direct implications in the research outlined in this dissertation, as the hierarchy was created using Aggregations, and Aggregated Resources formed the basis of those entities, but there are also wider implications for others seeking to create similar hierarchies or other complex structures using the simple set of resources the model provides. This research can help inform others that require a flexibility in entity portrayal or the relationships between them, whether this flexibility is needed for a description of niche object types, meeting the needs of diverse communities or user groups, or other studies where existing defined concepts present some limitations similar to those presented here.

Another limitation regarding data automation concerns was encountered in the context of identifying matching data across a set of data providers, and in the actual creation of bibliographic metadata in various syntaxes. Though the collection of bibliographic metadata from various providers and the mapping of their data to properties is relatively straightforward, throughout this research it has mostly been a manual task; the automation of the aforementioned data analysis functions would make this process considerably less time consuming. A similar conclusion was noted by McDonough (2013a) who found the process of creating metadata for videogames as multiple entities within OAI-ORE was straightforward but manually intensive. Though not presented in this dissertation, creation of bibliographic data was output as RDF/XML, though implementation, particular among hobbyist providers, may be more realistic if the JSON-LD format were used instead (which the OAI-ORE model does support) due to the a more widespread use of this syntax amongst these providers. Related to this issue is that the descriptive metadata applied to Aggregation entities manually does not inherently state the actual data provider source. Any automated data application to an aggregation entity should attempt to identify to which provider a particular metadata value has been sourced from.

Methodology limitations

While much of the data collection and analysis was performed manually in this research, it is yet to be determined how automated replacements for some processes would affect the type and amount of data able to be obtained. The more manual approach here allowed for data not easily available through APIs to be gathered, analyzed, and organized into entities and Resource Map files in whatever way the author desired. Though existing automated processes could replicate much of this, particular the data collection and organization phases, accessing certain desired content from certain web providers may prove challenging. For example, the chapter summaries for *One Piece* volumes shown in Figure 7.3 are accessible via DBpedia, though the summary for every volume is grouped into a single dbp:summary

value³⁹ rather than being separated by volume. As a limitation, this means that the manual processes used in this research may have resulted in a more thorough and ideal level of data access and aggregation output for each entity within the hierarchy than could be expected if a feasible automated approach were used instead. Though this is not an insurmountable limitation, it does raise the question of which processes used in this research are available for sensible automation and the role manual efforts must play beyond that. This discussion extends into the importance of manual work in the curation of the final aggregated data, as even if all data collection, analyses, and organization was done using automated methods, community or institution-specific needs would typically result in not wanting the totality of this aggregated data to be available. Though this curation is generally a positive, as the final data aggregation output can be tailored by each implementer to meet specific user needs, it may also be seen as a type of limitation, as some portion of the process will inevitably need manual intervention. Nevertheless, provided that labor is available for this final data process, it is an overall benefit of the OAI-ORE model, as the descriptive metadata made available to users can be thoughtfully curated by those responsible for dissemination.

Other Encountered Issues

Over the time this research has progressed, models such as BIBFRAME and LRM have become more complete and perhaps closer to some form of implementation. While the author sought to differentiate the Superwork and hierarchy from FRBR and FRBR_{OO}, significant work on BIBFRAME and LRM as alternative data models has yet to be undertaken, and it is unknown if these models can better represent a franchise-level entity as it has been envisioned in this dissertation. BIBFRAME defines its Work entity as “the highest level of abstraction” and that it is “the conceptual essence of the cataloged resource: authors, languages, and what it is about (subjects)” (Library of Congress, 2016); because the franchise-level entity is arguably a level of abstraction above the essence of a catalogued resource (as typically, single creative works and not franchises are the catalogued resource), it appears that this would not be suitable for representing a franchise, though no definitive answer can be given without a more thorough analysis. That said, Zapounidou *et al.* (2017) demonstrated that in attempting to map FRBR to BIBFRAME 2.0, some important information is lost. Specifically, they found that mapping a Work with multiple Expressions results in a loss of information regarding the common progenitor (see p.22 & p.24). As much of the hierarchies presented here are dependent on common progenitors at multiple levels, alternative results would need to be found if BIBFRAME use was desired. LRM also includes conceptual entities which may be able to represent the franchise as a subject but may fail to portray the includes or membership relationships that Superwork and its connected entities have, similar to issues regarding the Complex Work. Further clarification on these issues should be examined in the future.

³⁹ See [http://dbpedia.org/page/List_of_One_Piece_chapters_\(1%E2%80%93186\)](http://dbpedia.org/page/List_of_One_Piece_chapters_(1%E2%80%93186))

Throughout the creation of the Superwork entity, both during solo work by the author and research carried out with collaborators, dozens of franchises have been used as examples to determine how well the Superwork is able to handle the diversity that exists among multimedia franchises with regards to genres, resource types, medium focuses, types of progenitor works, etc. An issue not often discussed, however, is when exactly a Superwork came into existence, as larger, established franchises were often used as examples. As the model developed into the form presented in this dissertation, this question remains, though another has also arisen - what is the utility of the Series entity, if any, if only a single series exists? While a definitive answer has yet to be formulated, brief examination appears to suggest that once a creative work spawns a derivative that warrants a new Series layer entity (e.g. a related but new universe, a spin-off series that differs from the mainline one), then both that secondary Series entity and a Superwork entity is formed. Though this appears to be valid for the examples used throughout the author's Superwork research, it remains a hypothesis at the moment, and the question of how and when a Superwork and multiple Series entities are created have yet to be definitively answered.

Chapter 8: Conclusions and Future Research Opportunities

The conclusion of the dissertation is presented in this chapter, along with a summary of the presented ideas and thoughts on future research avenues.

Summarizing the Research Findings Presented in this Dissertation

This dissertation has sought to address core issues related to the portrayal and description of multimedia franchises and the network of creative works and objects that they contain. In addressing these issues, the research presented here examined how a multimedia franchise may be portrayed as a single entity, how that entity may form relationships with other core franchise entities in the formation of a bibliographic hierarchy, and how a flexible metadata aggregation model may be used to both portray that hierarchy and provide it with granular and semantically informative descriptive metadata, sourced from a variety of data providers, to better represent and describe these resources for diverse group of fans that these resources have developed.

A number of contributions have resulted from the research presented here. First, the research has developed a conceptual Superwork entity that has been designed to enable a more formal representation of a multimedia franchise at its highest conceptual level, and to support the descriptions and relationships that frequently occur here. Second, a bibliographic model for multimedia franchises was created to model entities as they are most commonly described by various data providers and establish relationships between them to form a hierarchy for a given franchise. The creation of unique entities and their placement in a hierarchy allows for users to understand how a creative work exists in relation to other resources that they may also have an interest in accessing. The flexibility of this model, both in its structure, entities, and relational and descriptive properties, ensures that a diverse range of resource types beyond manga, anime, and games can be incorporated, and that a diverse set of user needs from various user groups can be better met. Lastly, this dissertation has contributed to existing research that has sought to use the small number of entities within the OAI-ORE data model to create more complex structures for representation on the web, namely the bibliographic hierarchy that was formed using OAI-ORE aggregated entities using existing web resources.

In the creation of the Superwork entity, a multimedia franchise is able to be portrayed at its highest conceptual level. This conceptual entity represents the network of individual creative works from numerous media types that together form a single franchise. This entity resembles collocation efforts that users have taken it upon themselves to perform in various places on the web, such as Wikipedia. Additionally, it enables the description of semantic relationships that may otherwise go undescribed. Such relationships include inspirations, crossovers, trope origins, and other relations that are often best attributed to a franchise in its entirety and not a single series or work that resides within that franchise.

While the Superwork enables a variety of interesting relationships, there are of course others to be made between series, creative works, or single multimedia objects. The bibliographic hierarchy

consisting of the Superwork, Series, Work, and Object entities, supports such relationships, while simultaneously forming the set of creative works that the franchise is a representation of. Apart from the “vertical” relationships connecting the entities in the creation of the core hierarchy, “horizontal relationships” between entities at each level support dynamic resources such as manga, anime, and video games, which this research has focused on, and allow for the ongoing inclusion of new entities and resources into the hierarchy as a franchise continues to grow over time.

The OAI-ORE aggregation model serves as a basis for the creation of the Superwork and other entities, their connection using various relational properties to form a hierarchy, and the description of the entities using existing web data. The allowance of open vocabulary use allowed for the portrayal of the hierarchy as defined, the connection of entities, both within a single hierarchy and between hierarchies, using relational terms of any desired level of granularity, and the description of each entity based on the aggregation of metadata sourced from existing data providers on the web. This model enables the creation of hierarchical entities based not simply of abstract levels of conceptuality found within a franchise, but rather on how hobbyists and institutions themselves have chosen to formulate and describe multiple layers of resources. While the levels of descriptive granularity and semantic relationships can vary due to factors such as the popularity of a franchise and the availability of descriptive data created by volunteers on the web, as hobbyists have shown to be generally enthusiastic about creating interesting, informative, and useful data for multimedia objects, the reliance on web data was proven to be successful.

While the work presented in this dissertation has not been without challenges or limitations, the achievement of the overarching research goals has been successful. The creation of entities and a hierarchical structure that better conform to how users have shown they view multimedia objects was achieved. The definition of a bibliographic hierarchy that allows for flexibility in how it portrays entities and relationships allowed for the development of models and concepts that serve the needs of a diverse range of pop culture fans. Lastly, a suitable base data model, OAI-ORE, enabled the building of the hierarchy as it was envisioned, while also supporting the increase in semantic relationships and granular, bibliographic metadata that was required to better meet the needs of users was utilized. Although further research is needed for the refinement and implementation of several ideas, the broad goal of the betterment of bibliographic description for multimedia objects was successfully carried out.

Possibilities for Future Research

One issue that has not been touched upon, either in this dissertation or in any significant manner in the research conducted previously, is that of the preservation of bibliographic data and resource descriptions. A primary reason for this is the use of existing web resources as targets of OAI-ORE aggregation and not the heavy ingestion of descriptive metadata from these resources. Though metadata is indeed created and offered for access via the Resource Map entity, because such data has not yet been created in mass quantities, the ideas of storing it for the purposes of preservation have remained in their infancy.

Numerous works on the importance of fan-created works and the need to properly describe and preserve them exist (e.g. see Bullard, 2015, 2016; Gursoy, 2015; Hart *et al.*, 1999; Johnson, 2014; Price and Robinson, 2017; Versaphile, 2011), and while the preservation of creative works and bibliographic data are certainly different, the author takes the position that the wealth of semantic data that users have created is worth some amount of preservation effort. It is not an unrealistic scenario that a popular hobbyist web provider, with web hosting funds provided voluntarily and maintenance or backups not institutional mandated, could easily disappear, descriptive data along with it. As the currently inadequate state of institutional data has been highlighted throughout this research, the disappearance of hobbyist data providers should not be left to chance. One avenue for future research is the investigation of this issue and possible solutions, one of which may be an implementation of the metadata aggregation described here, with some hobbyist data being aggregated with and hosted by institutional providers.

Though data from both institutions and hobbyists were used in order to formulate all of the core concepts presented here, a practical implementation amongst data providers was not carried out in this research. As the number of various communities that can exist for the multiple mediums that form a multimedia franchise can be vast, initial implementation among existing hobbyist providers may prove difficult. A more realistic option at this stage, mentioned briefly earlier in this dissertation, is the implementation of a form of this model within the MADB. As this provider already features the portrayal of multiple hierarchical entities, separate databases for the three main MAG objects that this research focused on, and rudimentary connections between them are already enabled, the aggregation of multiple data providers and the storing and portrayal of this data at the MADB is a possibility that will be explored in the future.

This dissertation has emphasized the importance of users throughout, though the data models and concepts developed during the presented research was performed mainly through an attempt to observe and understand user data as it exists on the web. Most of this data is created and curated by users, the ideas discussed align with multiple user communities and the findings align with limited user studies that have been done, particularly with relation to the Superwork and user collocation activities. Based on these notions, the author is confident that the ideas in this research would be accepted by users. Nevertheless, a user study with a focus group consisting of fans representative of the diverse users of hobbyist providers may reveal hereto unmet needs that future developments of the entities, models, or aggregation methods may help address.

As much of this research has been focused on user-created data and understandings of how multimedia objects exist and are related to one another, the role of these crowdsourced user descriptions in areas such as participatory cultures or collective intelligence should be explored. Collective intelligence and participatory cultures on the web have been the focus of much past research (see e.g. Levy, 1999, Jenkins, 2006b and 2013, Martin and Steinkuehler, 2010), with most studies looking at areas such as social implications, group dynamics, changes in relationships between consumers and

creators, and others. Areas of future research within these areas may include topics such as how participatory contributions to bibliographic descriptions differ from traditional, i.e. institutional descriptions (specifically with regards to group or crowdsourced descriptions rather than solely a hobbyist vs. institutional difference), or how the importance of the Superwork or franchise-level descriptions on the web may be the result of a distributed network of individuals participating in the creation of descriptive data. Regardless, the wealth of hobbyist data on the web is evidence for the effectiveness of collective intelligence, and so the analysis of this data as an important product of participatory cultures should result in a number of interesting research streams.

Acknowledgements

All of my sincerest thanks to those who have given me various forms of support throughout my masters and doctoral research, or more generally throughout my stay in Japan. All of my academic advisors and other faculty and staff who have helped academically and in support for various other activities have been an inexplicable help, most notably my supervisors Sugimoto and Nagamori. Whether it has been in the help or guidance of research over the years or just general conversations about anything and everything, my utmost thanks goes out to you. You've been supportive academically, financially, emotionally, and in other, significant-to-quality-of-life ways. I'm also thankful for being given the opportunity to explore research opportunities and make academic (and friends) overseas at conference and other visits. While I unfortunately take this opportunity for granted more often than I should, I am extremely grateful for these amazing events. As an extension of this, a big thanks to all of the interesting researchers that we have collaborated with, both in Japan and abroad. Those from the US, Taiwan, Thailand, Germany, Korea, and other nations have been both welcoming and hospitable during my visits, but a pleasure to collaborate with. The members of the mdlab who have come, gone, and are still here, have of course also been great additions to my study and to my stay. The interactions have always been genuine, fun, and helpful. The input that all of these members have given to my research over the years has, again, been of a level of help that is difficult to explain. I hope I've been able to provide some help as well, but I think we all have been able to build off of one another in really constructive ways.

Outside of school and outside of Tsukuba, a lot of life support has come from my family and friends, both in Japan and in Canada. My parents and my sisters get the biggest thanks for the emotional and financial support. I'm sure they'd like to me communicate (and visit) more than I have, but I hope they are aware of my appreciation. Nana and my too-little-visited Tokyo friends have also been necessary in keeping my sane over the past several years, so a large thank you to all.

I believe the handful of people I am close with – that is, everyone that this section refers to – knows that my thanking skills are not the best and that an accurate acknowledgement of the support you have all provided for me would be difficult for me to communicate, but hopefully these words communicate some level close to it. At least, it's more than you have probably heard me say in the past five years.

References

- Asahi Shimbun (2017), "Anime News: Japan's anime market in 2016 worth more than 2 trillion yen", 10 December.
- Bailey, M. (2018), "Tate outsource artist biographies on its website to Wikipedia." Available at <https://www.theartnewspaper.com/news/tate-uses-wikipedia-entries-on-artists-for-its-website> (accessed 12 September 2018).
- Bekiari, C., Doerr, M., Le Boeuf, P., Riva, P. (eds.) (2015): "FRBR object-oriented definition and mapping from FRBRER, FRAD and FRSAD (version 2.4)" Paris: ICOM/CIDOC.
- Bennett, R., Lavoie, B. F. and O'Neill, E. T. (2003), "The concept of a work in WorldCat: An application of FRBR", *Library Collections, Acquisition and Technical Services*, Vol. 27 No. 1, pp. 45–59.
- Branch, F. , Arias, T. , Kennah, J. , Phillips, R. , Windleharth, T. and Lee, J. H. (2017), "Representing transmedia fictional worlds through ontology", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 12, pp. 2771–2782.
- Bullard, J. (2016), "Motivating invisible contributions: Framing volunteer classification design in a fanfiction repository", In *Proceedings of the 19th International Conference on Supporting Group Work - GROUP '16*, pp. 181–193.
- Bullard, J. (2017), "Warrant as a means to study classification system design", *Journal of Documentation*, Vol. 73. No. 1, pp. 75–90.
- Carlyle, A. (1996), "Ordering author and work records: an evaluation of collocation in online catalog displays", *Journal of the American Society for Information Science*, Vol. 47 No.7, pp. 538-554.
- Carlyle, A. (1999), "User categorisation of works: toward improved organisation of online catalogue displays", *Journal of Documentation*, Vol. 55 No. 2, pp. 184–208.
- Coyle, K. (2004), "Future considerations: The functional library systems record", *Library Hi Tech*, Vol. 22 No., pp. 166–174.
- Coyle, K. (2016), *FRBR, Before and After: A Look at Our Bibliographic Models*, American Library Association, Chicago, IL.
- Davenport, G. (1998), "Very Distributed Media Stories: Presence, Time, Imagination", In *Proceedings 4th International Euro-Par Conference on Parallel Processing*, 47–54. Springer-Verlag: London.
- Dena, C. (2009), "Transmedia practice: theorising the practice of expressing a fictional world across distinct media and environments. Media and communications", PhD thesis, University of Sydney, Sydney.
- Ercegovac, Z. (2006), "Multiple-version resources in digital libraries: Towards user-centered displays", *Journal of the American Society for Information Science and Technology*, Vol. 57 No.8, pp. 1023–1032.

- ESA (Entertainment Software Association) (2016), "Essential Facts about the Computer and Video Game Industry." Available at: <http://essentialfacts.theesa.com/Essential-Facts-2016.pdf> (accessed 7 June 2018)
- Fee, William. T. B. (2013), "Where Is the Justice... League?: Graphic Novel Cataloging and Classification", *Serials Review*, Vol. 39 No. 1, pp. 37–46.
- Ferro, N. and Silvello, G. (2013), "Modeling archives by means of OAI-ORE", In Agosti, M., Esposito, F., Ferilli, S. and Ferro, N. (Eds), *Digital Libraries and Archives. IRCDL 2012. Communications in Computer and Information Science*, Vol. 354, Springer, Berlin and Heidelberg, pp. 216-227.
- Giovagnoli, M. (2011), *Transmedia Storytelling: Imagery, Shapes and Techniques*. Pittsburgh, PA: ETC Press.
- Gursoy, A. (2015), "Evaluating fan fiction metadata for preservation use", In *Proceedings of the Association for Information Science and Technology*, Vol. 52, No. 1, pp. 1–4.
- Hart, C. et al. (1999), "The bibliographical structure of fan information", *Collection Building*, Vol. 18, No. 2, pp. 81–89.
- IFLA Study Group on the Functional Requirements for Bibliographic Records (2006), *Functional Requirements for Bibliographic Records*, (September 1997). Available at: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>
- Jenkins, H. (2003), *Transmedia Storytelling*. *MIT Technology Review*, 28. Available at: <https://www.technologyreview.com/s/401760/transmedia-storytelling>
- Jenkins, H. (2006a), *Convergence Culture: Where Old and New Media Collide*. NYU Press.
- Jenkins, H. (2006b), *Fans, Bloggers, and Gamers: exploring participatory culture*. NYU Press.
- Jenkins, H. (2013), *Textual Poachers: Television Fans and Participatory Culture*, Routledge.
- Jett, J. et al. (2016), "A conceptual model for video games and interactive media", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 3, pp. 505–517.
- Johnson, S. F. (2014), "Fan fiction metadata creation and utilization within fan fiction archives: Three primary models", *Transformative Works and Cultures*, Vol. 17, pp. 1–15.
- Keet, C.M. (2006), "Part-whole relations in object-role models", in Meersman, R., Tari, Z. and Herrero, P. (Eds), *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. OTM 2006. Lecture Notes in Computer Science*, Vol. 4278, Springer, Berlin and Heidelberg, pp. 1118-1127.
- Kiryakos, S. (2015), "Model for Aggregating Manga Metadata Across Institutions: Improving the Granularity of Manga Bibliographic Data on the Semantic Web", master's thesis, University of Tsukuba, Tsukuba, Japan.
- Kiryakos, S., and Sugimoto, S. (2015), "A Linked Data Model to Aggregate Serialized Manga from Multiple Data Providers", In Allen R., Hunter J., Zeng M. (eds) *Digital Libraries: Providing*

- Quality Information. ICADL 2015. Lecture Notes in Computer Science*, Vol. 9469. Springer, Cham, pp. 120–131.
- Kiryakos, S., and Sugimoto, S. (in press, 2018a). “Building a bibliographic hierarchy for manga through the aggregation of institutional and hobbyist descriptions”, *Journal of Documentation*.
- Kiryakos, S., and Sugimoto, S. (in press, 2018b). “The Representation of a Multimedia Franchise as a Single Entity: Contrasting Existing Bibliographic Entities with Web-Based Superwork Portrayals”, *LIBRES – Library and Information Science Research*.
- Kiryakos, S., Sugimoto, S., Nagamori, M., & Mihara, T. (2016), “Aggregating Metadata from Heterogeneous Pop Culture Resources on the Web”, *In Proceedings of the International Conference on Dublin Core and Metadata Applications*, 2016, pp. 65–74.
- Kiryakos, S., Sugimoto, S., Lee, J.H., Jett, J., Cheng, Y., & Downie, J.S. (2017), “Towards a Conceptual Framework for Superworks”, *In Proceedings of the 7th Conference of Japanese Association for Digital Humanities*, September 11-12, 2017 (pp.47-49). JADH.
- Langford, B. (2010), *Post-Classical Hollywood: Film Industry, Style and Ideology since 1945*. Edinburgh University Press.
- Lagoze, C., Van de Sompel, H., Johnston, P., Nelson, M., Sanderson, R. and Warner, S. (2008a), “ORE specification – abstract data model.” Available at: www.openarchives.org/ore/1.0/datamodel (accessed June 8, 2017).
- Lagoze, C., Van de Sompel, H., Johnston, P., Nelson, M., Sanderson, R. and Warner, S. (2008b), “ORE user guide – primer. Open Archives Initiative object reuse and exchange”, Open archives initiative. Available at: <https://www.openarchives.org/ore/1.0/primer> (accessed 11 May 2018).
- Leazer, G. H. and Furner, J. (1999), “Topological Indices of Textual Identity Networks”, *In 62nd Annual Meeting of the American Society for Information Science, Information Today*, Medford, NJ, pp. 345-358.
- Lee, J. H., Shim, Y., & Jett, J. (2015), “Analyzing User Requests for Anime Recommendations”, *In Proceedings of the 15th ACM/IEEE-CE on Joint Conference on Digital Libraries - JCDL '15*, ACM, New York, pp. 269-270.
- Lee, J. H., Cho, H., Fox, V., & Perti, A. (2013), “User-centered approach in creating a metadata schema for video games and interactive media”, *In Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries - JCDL '13*, ACM, New York, pp. 229-238.
- Lee, J. H., Jett, J., Cho, H., Windleharth, T., Kiryakos, S., Disher, T., & Sugimoto, S. (in press), “Reconceptualizing Superwork for Improved Access to Popular Cultural Objects.” Paper presented at 2018 Association for Information Science & Technology (ASIS&T), Vancouver, Canada.

- Lee, S. and Jacob, E. K. (2011), "An Integrated Approach to Metadata Interoperability", *Library Resources & Technical Services*, Vol. 55 No. 1, pp. 17–32.
- Levy, P. (1999), *Collective Intelligence: mankind's emerging world in cyberspace*. Perseus.
- Lemke, J. (2004) "Critical Analysis across Media: Games, Franchises, and the New Cultural Order", *Approaches to Critical Discourse Analysis*, Vol 7 No. 1, pp. 1–22.
- Library of Congress. (2016), "Overview of the BIBFRAME 2.0 Model." Available at: <https://www.loc.gov/bibframe/docs/bibframe2-model.html>
- Martin, C. and Steinkuehler, C. (2010), "Collective Information Literacy in Massively Multiplayer Online Games", *E-Learning and Digital Media*, Vol. 7 No.4, pp. 355–365.
- McDonough, J., Kirschenbaum, M., Reside, D., Fraistat, N., & Jerz, D. (2010), "Twisty little passages almost all alike: applying the FRBR model to a classic computer game", *DHQ: Digital Humanities Quarterly*. Available at: <http://digitalhumanities.org:8081/dhq/vol/4/2/000089/000089.html>
- McDonough, J. P. (2011), "Packaging videogames for long-term preservation: Integrating FRBR and the OAIS reference model", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 1, pp. 171–184.
- McDonough, J.P. (2013a), "A tangled web: metadata and problems in game preservation", *The Preservation of Complex Objects*, Vol. 3, pp. 49-62.
- McDonough, J.P. (2013b), "Some assembly required: reflections on XML semantics, digital preservation and the construction of knowledge", In *Proceedings of Balisage: The Markup Conference 2013, Balisage Series on Markup Technologies*, Montréal, Canada, 6-9 August 2013, Vol. 10.
- Menard, D. (2015), "Entertainment Assembled: The Marvel Cinematic Universe, a Case Study in Transmedia", master's thesis, Liberty University, Lynchburg, Va.
- Merčun, T., Žumer, M. and Aalberg, T. (2012), "Presenting and exploring the complexity of bibliographic relationships", In Chen HH., Chowdhury G. (eds) *The Outreach of Digital Libraries: A Globalized Resource Network. ICADL 2012. Lecture Notes in Computer Science*, Vol. 7634 Springer, Berlin, Heidelberg, pp. 63–66.
- Merčun, T., Žumer, M. and Aalberg, T. (2017), "Presenting bibliographic families using information visualization: Evaluation of FRBR-based prototype and hierarchical visualizations", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 2, pp. 392–411.
- Noerr, P., Goossens, P., Matei, D., Otten, P., Peruginelli, S., & Witt, M. (1998), "User benefits from a new bibliographic model: follow-up of the IFLA functional requirements study." *64th IFLA general conference*. Available at: <http://www.ifla.org/IV/ifla64/084-126e.htm>.
- Niu, J. (2013), "Hierarchical Relationships in the Bibliographic Universe", *Cataloging & Classification Quarterly*, Vol. 51 No. 5, pp. 473–490.

- Niu, J. (2015), "Transforming archival education", *Archives and Manuscripts*, Vol. 43 No. 2, pp. 120-130.
- Nozawa, S. (2013), "Characterization", *Semiotic Review*, Vol. 3. Available at: <https://www.semioticreview.com/ojs/index.php/sr/article/view/16> (Accessed: 14 May 2018).
- Picard, M. and Pelletier-Gagnon, J. (2015), "Introduction: Geemu, media mix, and the state of Japanese video game studies", *Kinephanos: Journal of Media Studies and Popular Culture*, Vol. 5 No. 1, pp. 1-19.
- Pineda, R. A. (2017), "Japan's Manga Market Grows 0.4% in 2016, Digital Sees 27.5% Increase", Available at: <https://www.animenewsnetwork.com/news/2017-03-18/japan-manga-market-grows-0.4-percent-in-2016-digital-sees-27.5-percent-increase/.113129> (accessed 12 June 2018).
- Price, L. and Robinson, L. (2017), "Fan fiction in the library", *Transformative Works and Cultures*, Vol. 25.
- Prom, C.J. and Habing, T.G. (2002), "Using the Open Archives Initiative protocols with EAD", In *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries*, ACM, New York, NY, January, pp. 171-180.
- Reid, C. (2016), "North American Comics Market Reaches \$1.03 Billion. Publishers Weekly." available at: <http://www.publishersweekly.com/pw/by-topic/industry-news/comics/article/70897-north-american-comics-graphic-novel-market-reaches-1-03-billion.html> (accessed 3 July 2018).
- Smiraglia, R. P. (2007), "Bibliographic families and superworks." in Taylor, A. G. (ed.), *Understanding FRBR: what it is and how it will affect our retrieval tools*, *Libraries Unlimited*, Westport, Conn., pp. 73-86.
- Smiraglia, R. P. and Leazer, G. H. (1999) "Derivative bibliographic relationships: the work relationship in a global bibliographic database", *Journal of the American Society for Information Science*, Vol. 50 No. 6, pp. 493–504.
- Smiraglia, R. P., Riva, P., Žumer, M. (2014), *The FRBR family of conceptual models: toward a linked bibliographic future*, Routledge, London.
- Svenonius, E. (2000), *The intellectual foundation of information organization*. Cambridge, Mass: MIT Press.
- Tallerås, K., Dahl, J. H. B., & Pharo, N. (2018), "User conceptualizations of derivative relationships in the bibliographic universe", *Journal of Documentation*, Vol. 74 No. 4, pp. 894–916.
- Teixeira, M.T. (2010), "Contributions to FRBRization. The library catalogue in a new digital environment: how FRBR can help produce better display in OPACs", master's thesis, Instituto Universitario de Lisboa, ISCTE, Lisbon.
- Versaphile (2011), "Silence in the library: Archives and the preservation of fannish history", *Transformative Works and Cultures*, Vol. 6, pp. 1–5.

- Vukadin, A. (2014), “Bits and Pieces of Information: Bibliographic Modeling of Transmedia”, *Cataloging & Classification Quarterly*, Vol. 52 No. 3, pp. 285–302.
- Wang, S., Isaac, A., Charles, V., Koopman, R., Agoropoulou, A. and van der Werf, T. (2013), “Hierarchical structuring of cultural heritage objects within large aggregations”, in Aalberg, T., Papatheodorou, C., Dobрева, M., Tsakonas, G. and Farrugia, C.J. (Eds), *Research and Advanced Technology for Digital Libraries*. TPD L 2013. Lecture Notes in Computer Science, Vol. 8092, Springer, Berlin and Heidelberg, pp. 247-259.
- Wikipedia (2018a). *Ghost in the Shell*. [online] Available at:
https://en.wikipedia.org/wiki/Ghost_in_the_Shell.
- Wikipedia (2018b). *List of multimedia franchises*. [online] Available at:
https://en.wikipedia.org/wiki/List_of_multimedia_franchises.
- Wilson, P. (1968). *Two kinds of power: An essay on bibliographical control*. Berkeley: Univ of California Press.
- Zapounidou, S., Sfakakis, M. and Papatheodorou, C. (2017), “Preserving Bibliographic Relationships in Mappings from FRBR to BIBFRAME 2.0”, in: Kamps J., Tsakonas G., Manolopoulos Y., Iliadis L., Karydis I. (eds), *Research and Advanced Technology for Digital Libraries*. TPD L 2017. Lecture Notes in Computer Science, Vol. 10450. Springer, Cham, pp. 15–26.

List of Publications

Part 1: Peer reviewed Academic Journals

Kiryakos, S., and Sugimoto, S. (in press, 2018a). “Building a bibliographic hierarchy for manga through the aggregation of institutional and hobbyist descriptions”, *Journal of Documentation*. doi: 10.1108/JD-06-2018-0089. Page count: 27.

Kiryakos, S., and Sugimoto, S. (in press, 2018b). “The Representation of a Multimedia Franchise as a Single Entity: Contrasting Existing Bibliographic Entities with Web-Based Superwork Portrayals”, *LIBRES – Library and Information Science Research*. Page count: 14.

Part 2: Peer reviewed Proceedings paper at an International Conference

Kiryakos, S., and Sugimoto, S. (2015), “A Linked Data Model to Aggregate Serialized Manga from Multiple Data Providers”, In Allen R., Hunter J., Zeng M. (eds) *Digital Libraries: Providing Quality Information. ICADL 2015. Lecture Notes in Computer Science*, Vol. 9469. Springer, Cham, pp. 120–131.

Kiryakos, S., Sugimoto, S., Nagamori, M., & Mihara, T. (2016), “Aggregating Metadata from Heterogeneous Pop Culture Resources on the Web”, In *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 2016, pp. 65–74.

Sugimoto, S., Kiryakos, S., Wijesundara, C., Monika, W., Mihara, T., & Nagamori, S. (2018) “Metadata Models for Organizing Digital Archives on the Web: Metadata-Centric Projects at Tsukuba and Lessons Learned”, In *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 2018. Page count: 10.

Lee, J. H., Jett, J., Cho, H., Windleharth, T., Kiryakos, S., Disher, T., & Sugimoto, S. (in press), “Reconceptualizing Superwork for Improved Access to Popular Cultural Objects”, *2018 Association for Information Science & Technology (ASIS&T)*, Vancouver, Canada. Page count: 8.

Part 3: Peer reviewed Poster at an International Conference

Kiryakos, S., Sugimoto, S., Lee, J.H., Jett, J., Cheng, Y., & Downie, J.S. (2017), “Towards a Conceptual Framework for Superworks”, In *Proceedings of the 7th Conference of Japanese Association for Digital Humanities*, September 11-12, 2017 pp.47-49. JADH.

Appendix I: Properties and values for entities shown in Figure 6.6.

rm1 – Resource Map	
dcterms:creator	http://mdlab.slis.tsukuba.ac.jp/
dcterms:modified	2018-8-01
ore:describes	http://mdlab.slis.tsukuba.ac.jp/skore/OnePiece/MW/aggregation*
a1 One Piece Work, Manga	
ore:aggregates	https://mediaarts-db.bunka.go.jp/mg/book_titles/107089
ore:aggregates	https://myanimelist.net/manga/13/One_Piece
dcterms:creator	https://viaf.org/viaf/19898231/
dcterms:creator	https://myanimelist.net/people/1881/Eiichiro_Oda
dcndl:transcription	オダエイイチロウ
dcterms:title	One Piece
dcterms:alternative	ワンピース
dcterms:subject	https://id.ndl.go.jp/auth/ndlsh/00564693
dcterms:subject	https://myanimelist.net/manga/genre/1/Action
dcterms:subject	https://myanimelist.net/manga/genre/2/Adventure
dcterms:publisher	https://www.shueisha.co.jp/
schema:publisherImprint	http://www.shonenjump.com/j/comics/
schema:publisherImprint	https://myanimelist.net/manga/magazine/83/Shounen_Jump_Weekly
dcterms:source	https://mediaarts-db.bunka.go.jp/mg/magazine_titles/11470
dcterms:audience	https://en.wikipedia.org/wiki/Shōnen_manga
dcterms:audience	https://myanimelist.net/manga/genre/27/Shounen
dcterms:format	https://comicmeta.org/vocab/Manga.html
dcterms:language	http://www.lexvo.org/page/iso639-3/jpn
dcterms:isPartOf	https://mediaarts-db.bunka.go.jp/mg/comic_works/81200
dcterms:issued	1997-07-22
dcterms:abstract	<p>Gol D. Roger, a man referred to as the "Pirate King," is set to be executed by the World Government. But just before his demise, he confirms the existence of a great treasure, One Piece, located somewhere within the vast ocean known as the Grand Line. Announcing that One Piece can be claimed by anyone worthy enough to reach it, the Pirate King is executed and the Great Age of Pirates begins.</p> <p>Twenty-two years later, a young man by the name of Monkey D. Luffy is ready to embark on his own adventure, searching for One Piece and striving to become the new Pirate King. Armed with just a straw hat, a small boat, and an elastic body, he sets out on a fantastic journey to gather his own crew and a worthy ship that will take them across the Grand Line to claim the greatest status on the high seas.</p>
dcterms:hasVersion	https://myanimelist.net/anime/21/One_Piece
schema:character	https://myanimelist.net/character/40/Luffy_Monkey_D
schema:isRelatedTo	https://myanimelist.net/manga/793/Wanted

*Note, sample URI

Appendix II: Brief descriptions and example website representatives for entities shown in Figures 5.4 (top) and 5.5 (bottom).

Entity Layer	Graph Entity	Description	Web Representation Example
Superwork	e1	Complete <i>Gundam</i> franchise representation.	https://en.wikipedia.org/wiki/Gundam
Series	e2	<i>Gundam SEED</i> series representation, all mediums.	https://en.wikipedia.org/wiki/Mobile_Suit_Gundam_SEED
	e3	<i>Gundam Wing</i> series representation, all mediums.	https://en.wikipedia.org/wiki/Mobile_Suit_Gundam_Wing
Work	e4	<i>Gundam SEED</i> television animation representation, all episodes.	https://en.wikipedia.org/wiki/List_of_Mobile_Suit_Gundam_SEED_episodes
	e5	<i>Gundam Wing</i> television animation representation, all episodes.	https://en.wikipedia.org/wiki/List_of_Mobile_Suit_Gundam_Wing_episodes
	e6	<i>Gundam Wing</i> manga representation, all volumes.	http://gundam.wikia.com/wiki/Mobile_Suit_Gundam_Wing_Endless_Waltz:_Glory_of_the_Losers
Object	e7	<i>Gundam SEED</i> television animation episode 1 representation.	http://gundam.wikia.com/wiki/False_Peace
	e8	<i>Gundam Wing</i> television animation episode 1 representation.	http://gundam.wikia.com/wiki/The_Shooting_Star_She_Saw
	e9	<i>Gundam Wing</i> manga volume 1 representation.	https://mediaarts-db.bunka.go.jp/mg/books/1165283

Entity Layer	Graph Entity	Description	Web Representation Example
Superwork	e1	Complete <i>Metal Gear</i> franchise representation.	https://en.wikipedia.org/wiki/Metal_Gear
Series	e2	Works belonging to the Mainline <i>Metal Gear</i> series.	https://en.wikipedia.org/wiki/List_of_Metal_Gear_media#Mainline
	e3	Spin-off works within the <i>Metal Gear</i> series.	https://en.wikipedia.org/wiki/List_of_Metal_Gear_media#Spin-offs
Work	e4	<i>Metal Gear Solid</i> game representation, all instances.	https://en.wikipedia.org/wiki/Metal_Gear_Solid
	e5	<i>Metal Gear Solid 2: Sons of Liberty</i> game representation, all instances.	https://en.wikipedia.org/wiki/Metal_Gear_Solid_2:_Sons_of_Liberty
	e6	<i>Metal Gear Acid</i> game representation.	https://en.wikipedia.org/wiki/Metal_Gear_Acid
Object	e7	<i>Metal Gear Solid</i> , PlayStation 1 release representation.	https://www.giantbomb.com/metal-gear-solid/3030-16140/releases/#toc-release-40830
	e8	<i>Metal Gear Solid 2: Sons of Liberty</i> , PlayStation 2 release representation.	https://www.mobygames.com/game/ps2/metal-gear-solid-2-sons-of-liberty
	e9	<i>Metal Gear Acid</i> , PSP release representation	https://mediaarts-db.bunka.go.jp/gm/syosai.php?id=392124000007&kbn=c12

Appendix III. Descriptions of expected attribute examples for each hierarchical entity listed in Chapter 5, subsection *Entity boundary intentions and hierarchy summary* (p. 46). Note that this section is not an exhaustive list of attributes expected to be included for each entity, but rather serves to provide guidance on the intended representation level for each entity within the hierarchy.

a. Superwork

i. Title – Superwork or franchise title. This may be an official title by the IP holder, or an unofficial but common franchise title used for reference by hobbyists or the general public.

ii. Original Creator – Agent(s) responsible for the creation of the progenitor of the franchise. Typically the original story author, illustrator, or character designer.

iii. Original Creation Date – Date associated with the beginning of the franchise. Original inception date be difficult or unavailable, in which case the date may be the first disseminated appearance or publication of the progenitor work.

iv. Original Medium – Medium or media type the franchise first appeared as, such as a manga, anime, game, piece of merchandise, and so on.

v. Genres / Tropes / Themes – Common genres, themes, tropes, archetypes, etc. that appear throughout the franchise. While each entity in the hierarchy may have its own and different values for this attribute, ones attached to the Superwork should be broad enough to be applied to each work under its umbrella.

b. Series

i. Title – Series or subseries title. This may be a unique official title, or identical to the Superwork title with an added qualifier to indicate various distinctions that the Series layer enables, e.g. *title* (Mainline) or *title* (Spin-off).

ii. Creator – Agent(s) responsible for series creation. In many cases, this may be identical to those referenced in the Superwork, though for larger franchises with numerous contributors, the creators may differ between individual Series.

iii. Original Release Date – Date associated with the release of the first instance of a given series. If a particular series exists across manga, anime, and video games, but was originally release as a manga, this date should be associated with that medium.

iv. Original Medium – Medium or media type the series first appear as. Depending on franchise, this may be the same or different than the equivalent attribute for the Superwork.

v. Existing Mediums – Mediums or media types the series contains creative works in. The utility of grouping these at the series level is to allow the straightforward identification of same-series derivatives and adaptations which may be more relevant to users than derivatives or adaptations from other series within a franchise.

vi. Universe / Timeline – Universe, timeline, story arc, or content-descriptor commonly used to group different distinct Series within a franchise.

vii. Canonicity – An indicator of whether the Series content is considered canon. While lower entities in the hierarchy would also utilize this attribute (as individual Works can be considered non-canon within a Series that is generally considered canonical), the use of this attribute at the Series level allows for the identification of fan-works or fan-fiction as their own ‘Series’.

c. Work*

i. Title – Work title. This may often be identical to the series title, as Work layer entities are separated by medium, and this medium indication is typically not present in the title.

ii. Creator – Agent(s) responsible for Work creation. May be identical to Series creators, particularly if the Work is the medium referenced by the **Original Medium** Series attribute.

iii. Medium – Medium or media type of the Work entity.

iv. Director – Director(s) of the animation Work. If multiple directors exist for different episodes or story-arcs, these may also be indicated here.

v. Original Network – Original television network responsible for the initial animation Work broadcast.

vi. Animation Studio – Animation studio responsible for the creation of the animation Work.

vii. Licensed Networks – Overseas networks licensed to broadcast the animation Work in their respective markets. The language and/or market may be indicated alongside each network.

viii. Original Broadcast – Start and, if the series is complete, end dates which the animation Work originally.

ix. Total Episodes – Total episode count, current or complete, for the animation Work.

Note that attributes iv – ix are example attributes that assume the Work entity describes a television animation. Dependent on the value of attribute **iii. Medium, said attributes would change to more medium-relevant ones, (e.g. in the case of a manga, Original Network becomes Original Magazine Appearance, Total Episodes becomes Total Volumes).*

d. Object

i. Title – Object title. In most cases, matches the Work title.

ii. Subtitle / Part Title – Subtitle or part title for the unique episode, volume, chapter, disc release, etc. being described.

iii. Creator – Agent(s) responsible for creative work within the Object. Typically will include those mentioned in the Work entity, though additional agents and roles may be unique to individual objects.

iv. Release Date – Object release date. As the Object entity is meant to be release / edition / publication agnostic, this may be the original release date and not the date of a unique later release, unless indicated.

v. Medium – Medium or media type of the Object entity. Though this can logically be inferred through its connection to the Work entity, individual Objects from different medium Works or different Series may be connected to indicated adaptations or other relationships, thus making the medium distinction at this level useful.

- vi. Format / Platform** – Used to indicate a unique format or platform the Object entity exists as. Though this may serve a limited purpose for some mediums, for others the distinction is more useful, e.g. distinguishing between Blu-ray and DVD versions of a home animation release, or between a Sony PlayStation and Nintendo GameCube versions of a video game release.
- vii. Part Indicator** – Value indicating the part of the Object entity, such as episode, volume, or disc number.

Appendix IV. Table with sample values for hierarchical entity attributes listed in Chapter 5, subsection *Entity boundary intentions and hierarchy summary* (p. 46) and Appendix III.

Entity	Attributes	Sample Values
Superwork	Title	Gundam
	Original Creator	Yoshiyuki Tomino
	Original Creation Date	April 7 1979
	Original Medium	Television Anime
	Genres / Tropes / Themes	Science Fiction, Mecha
Series	Title	Mobile Suit Gundam Wing
	Creator	Katsuyuki Sumizawa
	Original Release Date	April 7 1995
	Original Medium	Television anime
	Existing Mediums	TV Anime, OVA, Manga, Novels, Video Games, other
	Universe / Timeline	After Colony
	Canonicity	Canon, <i>Wing</i> Series
Work	Title	Mobile Suit Gundam Wing
	Creator	Katsuyuki Sumizawa
	Medium	Television Anime
	Director	Masahi Ikeda (1-26), Shinji Takamatsu (27-49)
	Original Network	TV Asahi
	Animation Studio	Sunrise
	Licensed Networks	Toonami (US)
	Original Broadcast	April 7 1995 – March 29 1996
	Total Episodes	49
Object	Title	Mobile Suit Gundam Wing
	Subtitle / Part Title	The Shooting Star She Saw
	Creator	Katsuyuki Sumizawa
	Release Date	April 7 1995
	Medium	Television Anime
	Format / Platform	Broadcast TV
	Part Indicator	Episode 1