

A report on ongoing research into the modeling of multimedia franchise hierarchies, the ‘Superwork’ entity, and the description of various pop culture resources using metadata aggregation

Senan Kiryakos
University of Tsukuba
Tsukuba, Ibaraki, Japan
senank@gmail.com

For several years, the University of Tsukuba, along with various international collaborators, has been conducting research into issues surrounding the formal modeling and bibliographic description of certain multimedia objects within popular culture, such as manga, anime, and video games. While the focus of this research has varied over time, several primary goals have been present throughout. These include the desire to improve bibliographic descriptions for ‘niche’ mediums, the representation of a multimedia franchise at its highest conceptual level, and the utilization of the wealth of user-created data on the web to assist in these goals, in addition to other related research objectives. In seeking to achieve these goals, research has been centred on two core concepts. The first is the utilization of metadata aggregation which attempts to improve the bibliographic description of multimedia objects by bringing authoritative data from cultural heritage institutions together with granular and semantically informative user-created data from various data providers on the web, such as fan sites or wikis. The second core concept – and the one in which most international collaboration has been centred on – is the creation of novel bibliographic descriptive models for various multimedia objects. Central to this has been the attempted creation, definition, and development of what has been referred to as a ‘Superwork’ entity – a conceptual resource meant to represent a total multimedia franchise and the network of creative works which together form that franchise. Over time, the concept and role of the Superwork has come to be an increasingly important considering when conducted further research, both in the areas of aggregation and bibliographic modeling and description. Accordingly, then, this paper presents a generally outline of past conducted research on both a bibliographic hierarchy for multimedia objects, efforts on aggregating institutional and hobbyist data, and the Superwork concept. Note that this paper is not meant to propose new ideas or results, but rather report on past work and further disseminate the concepts and conclusions encountered as a way to elicit thoughts, questions, and challenges from readers.

An Overview of Past Research

The goal most central to the popular culture and multimedia object research that our internal and collaborative efforts have focused on has been the overall betterment of representation and description of said objects, e.g. manga, anime, and video games, with “betterment” referring, generally, to the ability to

more adequately meet the needs of users of cultural heritage institutions that hold, maintain, preserve, and describe multimedia objects. While the spectrum of user needs can be vast, the desire to access more granularly informative descriptive data (when compared to the granularity of typical memory institution data) is frequently mentioned as being of high importance in various user analyses [1][2][3]. Outside of these user studies, a readily visible and accessible way to understand the information needs of users is to look at the data they have created and curated across various hobbyist resources on the web, such as fan websites or wikis. This user-generated web data is considerably more granular with regards to its descriptive data, containing useful information that institutional providers typically do not record, such as synopses, character information, and semantically informative genres or subject tags – data that users desire but that is frequently not recorded by traditional memory institutions. In pursuit of the goal of providing users with improved bibliographic representation and description, much past research attempted to access this user-generated granular web data and make it available for institutional use, with institutions also contributing their unique data (e.g. authoritative data created by libraries) where possible; metadata aggregation was and continues to be the main method of attempting to achieve this goal.

Aggregating Hobbyist and Institutional Metadata for Multimedia Objects

Although the idea of aggregating user-generated hobbyist data from the web with more formally created data produced by cultural heritage institutions appears straightforward, this is not the case. Early work [4][5] showed that the differences between the two data types resulted in problematic attempts at aggregation according to existing models and paradigms. Problems resulted primarily from traditional bibliographic models – the most often used in past research being the Functional Requirements for Bibliographic Records, or FRBR – being poorly suited to describing both the levels of granularity of web data, along with the entity levels that were being described. For example, while institutional data can be closely mapped to FRBR Group 1 Work, Expression, Manifestation, and Item entities, it is more difficult to do so with web data, as it is often not possible to identify web entities that align with entities such as the FRBR Item. The lack of common entities between institutional and hobbyist data combined with the inadequacy of existing ontologies in describing satisfactory levels of granularity meant that new models needed to be developed with which descriptive data for multimedia objects could be aggregated and attributed to. These models needed to better facilitate the description of entities for media objects as they were described on the web, support the dynamic nature of resource types like manga, anime, and video games (i.e. their growth and derivation over time), and support the aggregation of a wide range of data granularity to help serve diverse communities and their needs, be they institutional or hobbyist.

Because of the importance of web data to these aggregations, how individual and collective multimedia entities were described on the web guided much of thought process and model creation throughout the research. Apart from the differences in granularity and general entity representation when compared with institutional data, two core facets of hobbyist data were identified that models and aggregation methods needed to support. First, bibliographic data for pop culture objects is often described as a total multimedia franchise, e.g. *Star Wars*, *Bladerunner*, *Pokémon*, *Gundam* [6], rather than as a single creative work or series within that franchise. Second, relationships between resources, either within or external to a single franchise, are important for adding semantic context to resources. While considerable effort was spent attempting to make existing models adequately facilitate these facets of hobbyist data, in the end the choice was made to develop new bibliographic hierarchies and aggregation methods explicitly designed to better represent and describe the multimedia objects in question. Thus, latter portions of past research – continuing to the present – focused on the development of a new bibliographic hierarchy for multimedia objects and a method of aggregation to build that hierarchy, both of which sought to use existing resources and concepts from hobbyist resources on the web; the process of analyzing older models and developing new ones were disseminated as two prior publications, found at [7] and [8]. The remainder of this paper will focus on the bibliographic model and the Superwork concept rather than the aggregation model, based on OAI-ORE¹, which was developed in parallel. Further reading on this topic can be found in [7].

¹ Open Archives Initiative Object Reuse and Exchange; see <https://www.openarchives.org/ore/>

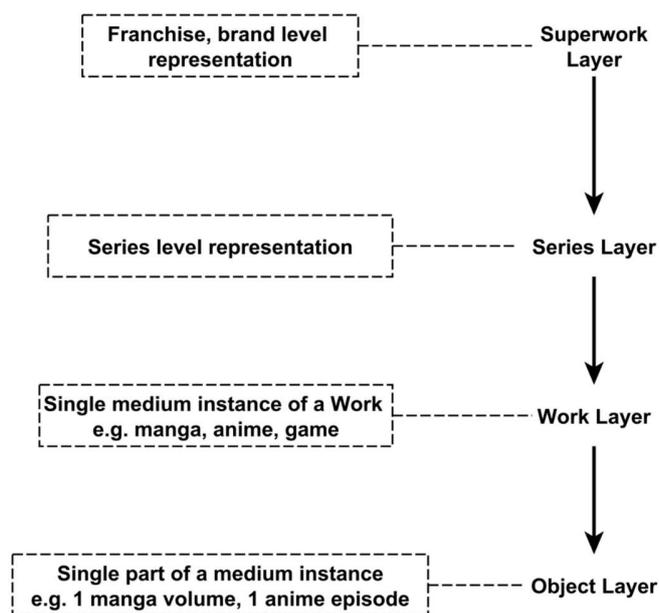


Figure 1. The bibliographic hierarchy developed for use with multimedia objects.

A Bibliographic Hierarchy for Multimedia Objects

Based on the need to better represent hobbyist web data, not only in regard to supporting increased granular descriptions, which could arguably be done by simply developing improved ontologies, but more importantly in supporting entities and relationships as they are understood and portrayed by users, a new bibliographic hierarchy was developed, an overview of which is shown in Figure 1. This hierarchy features core entities created based on how most web data exists across hobbyist and institutional providers:

1. The **Superwork** entity represents the idea of a complete multimedia franchise, agnostic of information that would only be attributable to a single series, creative work, or other instantiation within that franchise.
2. **Series** entities are used for a variety of levels of distinction, such as individual series within a franchise (e.g. *Gundam SEED*, *Gundam Wing*, both of which belong to the *Gundam* franchise), stand-out or prominent entries within a series or franchise (e.g. *Compilation of Final Fantasy VII*, a sub-series within the greater *Final Fantasy* mainline series), differing levels of canonicity (e.g. ‘Main’ series, ‘Spin-offs’, ‘Fan-fiction’), and others.
3. **Work** entities represent distinct medium instantiations of a given Series. For example, the *One Piece* franchise exists in mediums like anime and manga, and each of these distinct mediums is a unique Work entity.
4. Lastly, the **Object** entity is for a single part or instance of a given work. Using the same *One Piece* example, each episode of the TV animation would be its own object, as would each volume of the *One Piece* manga.

Though straightforward, the hierarchy here is better to able represent bibliographic data for a variety of multimedia objects when compared to existing models, and the core entities mirror the way that resources are portrayed and described on the web. Though not discussed in this paper, an equally important component of this research is the utilization of an OAI-ORE based aggregation model in building this

hierarchy, which enables a variety of functions, namely the addition of new entities should a community desire, the establishment of any number of relationships between entities both within and outside of a single franchise hierarchy, the use of any existing external vocabulary in order to support either specific institutional needs or a high level of descriptive granularity, and lastly the utilization of existing web data to act as aggregated resources for each of the entities within the hierarchy. What this last point means is that when the aggregation model is used to build the hierarchy in Figure 1, each of the four core entities is its own aggregation, and the resources that it aggregates are existing web resources describing that entity level. For example, the Wikipedia article for the *Gundam* franchise² would be an aggregated resource for the **Superwork** entity, the article for the *Gundam SEED* series³ would be an aggregated resource for the **Series** entity, the *Gundam SEED* anime article⁴ for the anime **Work** entity, and so on. The result of this is not only that most, if not all, descriptive data for multimedia objects on the web can be represented by an entity within the aggregated hierarchy, but that the entities themselves are based on how data exists across the web, rather than being arbitrary or abstract designations developed by the researchers. In summarizing the remainder of the research, institutional and hobbyist web data providers were analyzed in order to determine which of the hierarchical entities their descriptive data was describing. Sets of pages describing the same entity level were then used as aggregated resources for conceptual aggregations representing each of the four core hierarchical nodes. The descriptive metadata from these resources was then collected into a single resource (the OAI-ORE *Resource Map*, a web document which makes metadata available in various syntaxes such as RDF/XML and JSON-LD), resulting in aggregated entities containing a wealth of descriptive data, be it authoritative institutional data or granular hobbyist data. Lastly, relevant relationships were established between entities, both to form the aforementioned hierarchy and to connect relevant

² <https://en.wikipedia.org/wiki/Gundam>

³ https://en.wikipedia.org/wiki/Mobile_Suit_Gundam_SEED

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

entities within and across hierarchies, resulting in relationships such as sequels, crossovers, adaptations, and others to be described.

The Superwork Entity

While (in the opinion of the author) there are many interesting points within this research to discuss, such as the differences in representation and description levels among certain objects of popular culture, or the sophistication of user-generated bibliographic description and relationship networks on the web, perhaps the most novel concept – and the one with which the majority of the aforementioned international collaboration has been centred on – is that of the franchise-level Superwork. Differentiation between the Superwork and existing entities such as the FRBR Work, FRBR_{OO} Complex Work, and subject entities

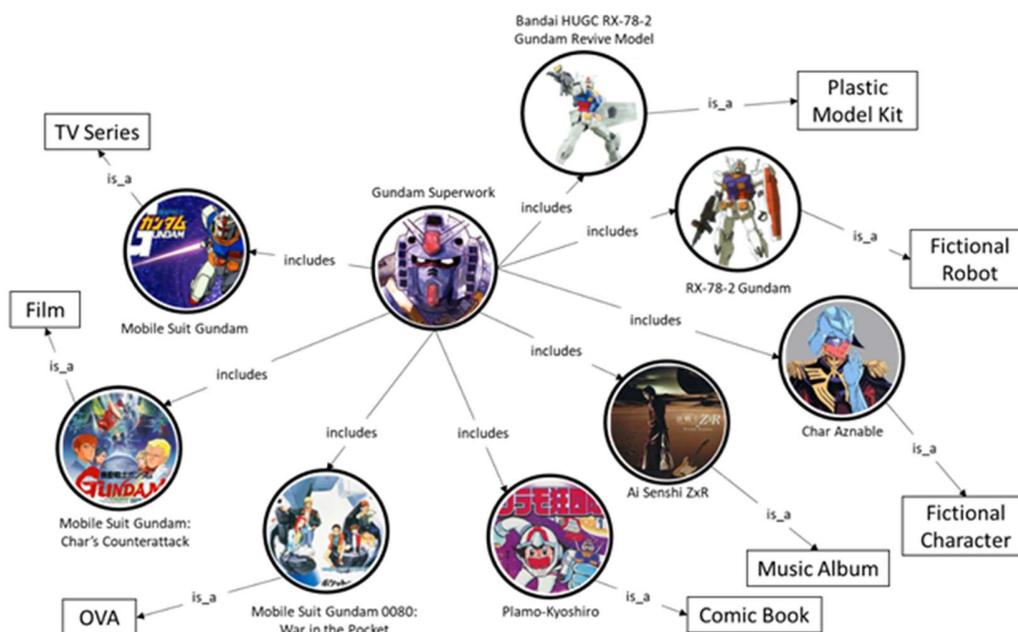


Figure 2. An example of the range of mediums connected to the *Gundam* Superwork; reprinted from [9].

have been difficult, though a significant distinction does appear to exist. Progress has also continued to be made on other issues, such as attempting to define the scope and boundaries of the Superwork, though this work is ongoing. It is clear, however, that the idea of a “franchise” concept is not well represented by a variety of existing bibliographic models, and that the concept is significant enough to investigate further. There are ever increasing creative and commercial incentives for the generation of multimedia franchises [10] and, importantly, it has been shown both in other studies [11] and on the web itself (e.g. Wikipedia) that users often understand and describe creative works within pop culture not only as singular entities, but in the context of the franchise to which they belong, and the network of relationships between other resources that are a part of. The significance of the franchise concept to users combined with the

inadequacies of existing models in sufficiently representing it has pushed the author to continue researching the Superwork entity, and the research presented here forms the foundation of work to be pursued in the near future. In addition to the initial goal of improving description levels to better meet user needs, which metadata aggregation has begun to address, future efforts will examine the importance of the discussed concepts in other areas, such as transmedia research, user-based modeling, and preservation of user-generated data through the holding of aggregated web data by institutions.

References

- [1] Fee, William. T. B. (2013), “Where Is the Justice... League?: Graphic Novel Cataloging and Classification”, *Serials Review*, Vol. 39 No. 1, pp. 37–46.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] For additional examples see Wikipedia (2018). *List of multimedia franchises*. Available at: https://en.wikipedia.org/wiki/List_of_multimedia_franchises.
- [7] 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*.
- [8] 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*.
- [9] 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.
- [10] see p. 207 in Langford, B. (2010), *Post-Classical Hollywood: Film Industry, Style and Ideology since 1945*. Edinburgh University Press.
- [11] 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.