Using Wikidata as Work Authority for Video Games

FUKUDA, Kazufumi
Ritsumeikan University,
Japan
fukudakz@gmail.com

Abstract

Video games have a short but rich history. Therefore, they have been gaining popularity as cultural heritage and research material. Several studies have analyzed the metadata and cataloging of video games. However, the research on its implementation is limited. Hence, we investigate the practice of cataloging video games at the Center for Game Studies, Ritsumeikan University (RCGS) in this study and examine the effectiveness of data utilization from Wikidata to construct an authority of works for video games. We accomplished this by associating the distribution package with Wikipedia and Wikidata. Consequently, records of works covering approximately half of the video games were created. However, the problem of uniformity of granularity and completeness was found in these data based on Wikipedia's culture and policies. Thus, data enrichment is difficult owing to the non-uniform granularity of bibliography with Wikidata. In contrast, the cost of data creation is effective. Furthermore, the external link ID is highly effective in enhancing the value of catalog as Linked Open Data (LOD). It is also evident that using published authority data is useful for data integration but Wikidata has some problems with its features. There is a need to consider the function and purpose of the catalog as linked data instead of a separate catalog. Thus, the adaptation of Wikidata for catalogs needs to be designed accordingly as linked data.

Keywords: video game; Wikidata; metadata; IFLA LRM; work authority

1. Background

The first video game in the world is believed to be Spacewar! (1961)[1], implying that video games have been in existence for almost 60 years. During this short period, video games have increasingly become popular. Statistically, more than 47,000 packages of video games have been published in Japan (Agency for Cultural Affairs n.d.). The number of online resources for video games published in the App Store and Google Play is expected to exceed 1 million. Through such diffusion, video games have come to be recognized as research materials and cultural heritage. Thus, the discussion on designing the metadata to improve accessibility for users and researchers is increasing.

Several studies have been conducted on video game metadata and cataloging. Jerome McDonough argued that traditional bibliographic descriptions are not suitable for interactive fiction including video games in the final report of Preserving Virtual Worlds Project (McDonough et al. 2010a). However, it is an early research on video game metadata. Furthermore, they investigated the potential of Functional Requirements for Bibliographic Records (FRBR) as a model to describe video games (McDonough et al. 2010b).

A close study on the metadata of video games is being conducted at the University of Washington, Seattle. They are creating a metadata schema that can describe the characteristic information of video games and interactive media. Jin Ha Lee developed the metadata schema in collaboration with the Seattle Interactive Museum (Lee et al. 2013). Sixteen core metadata elements (CORE 16) were selected in this research. Furthermore, they proposed 46 recommended elements

[1] Juul points out that its origin depends on the definition of the video game and there is a difference in recognition (Juul 2005). In short, the exact time of publication of first video games is controversial.

through user data survey and facet analysis (Lee et al. 2015). Jett proposed a conceptual model for video games and interactive media (Jett et al. 2016).

Groat analyzed the process of changing and standardizing the description of video games by describing the history of descriptive cataloging practices and MARC coding for video games (de Groat 2015). The Video Games Best Practice Task Force published the guide for video game cataloging (Online Audiovisual Catalogers, Inc. 2015). This document guides cataloging by RDA (resource description and access) that adopted the FRBR model. This proposal adopted MARC 21 to contribute to the standardization at libraries.

Various types of research on the cataloging and metadata development for video games have been developed as described above. These adopt the perspective of theoretical study. However, research accumulation based on the cataloging practice is insufficient.

We are proceeding research activities through by cataloging video games at the Ritsumeikan Center for Game Studies, Ritsumeikan University (RCGS). Metadata modeling, which is the premise of cataloging, is one of the most important research topics (Fukuda & Mihara 2018). The model is designed as an applied model of the IFLA Library Reference Model (IFLA LRM; Riva et al. 2017).

RCGS owns 10,000 video games and 4,000 related resources such as research books, magazines, strategy guides, official catalogs, and white papers. We are cataloging them by the metadata model. The bibliographic data published RCGS Collection (Ritsumeikan Center for Game Studies 2019). In addition, the data will be reused at the Media Arts Database produced by the Agency for Cultural Affairs (Agency for Cultural Affairs n.d.). This database is the only comprehensive database that provides access to popular cultural resources such as manga, anime, video game, and media arts. RCGS oversees the development of video games here.

The formulation of work records in this cataloging is a challenge. The definition of work in FRBR and IFLA LRM is highly abstract (IFLA Study Group on the FRBR 2009) because these models are designed as high-level conceptual models (Riva et al. 2017, p. 9.). It is assumed that the specific definition is left to the community and designer of the applied metadata model. It is one effective way to establish a strict definition and accordingly create a work record. Although this can function as a closed system, it is problematic in terms of connecting with external resources, granularity of records, and development cost.

2. Research Objective

This study examines the effectiveness of data utilization from Wikidata to construct an authority of works for video games.

Wikidata is an open platform that provides structured semantic data for Wikipedia and other applications (Erxleben et al. 2014, Mitraka et al. 2015, "Wikidata: Introduction" n.d.). The data is multilingual and highly interlinked. Furthermore, it has an external link such as identifiers of existing databases and authority controls (Vrandečić & Krötzsch 2014, p. 4-5). This feature is no exception for video game data. Video game records consist of IMDb ID, Metacritic ID, MobyGames ID, YouTube Gaming game ID, Twitch game ID, IGN game ID, and LC authority ID, among others.

In addition, Wikidata covers comprehensive cultural resources like Wikipedia. Therefore, it is assumed to be highly useful in the integrated handling of multiple related cultural resources. This is useful for integrating multiple related works in the Media Arts Database.

The bibliographical unit of video game articles in Wikipedia is not manifestation but close to work in FRBR. They focus on content features instead of physical features. An article often includes the original edition along with derivative editions such as limited edition, ported edition, localized edition, and remade edition. The tendency of the bibliographical unit is the same in Wikidata. For example, a record of one game has multiple platforms and ratings (FIG. 1). Thus,

the records of video games include related distribution packages in multiple standards, languages, and regions.

Wikidata has the potential to effectively function as a video game work authority. It is based on the bibliographical unit, harmonization, and integration of data.

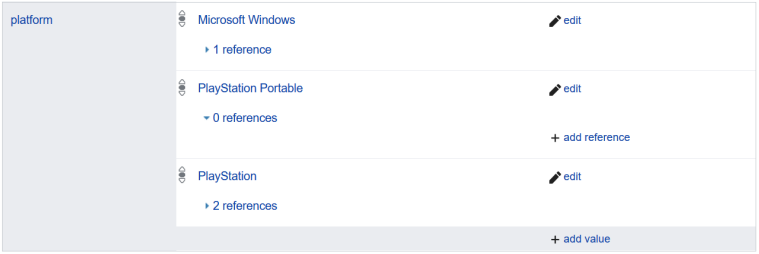

FIG. 1. Recorded multiple platforms of *Metal Gear Solid* in Wikidata ("Metal Gear Solid" n.d.)

In this research, we attempted to create bibliographic data with Wikidata as a work authority for video games. We examined the following research topics through this practice.

1. Comprehensiveness
2. Granularity
3. Completeness of records

The outline of the metadata model and the creation of data are described in the next chapter. We present the results of the above issues and evaluate this data creation based on relevant issues in chapter 4.

3. Metadata Model and Data Creation

We designed an interpretational model for the description of video games based on IFLA LRM (Fukuda & Mihara 2018). FIG. 2 indicates the main bibliographic entities in this model.

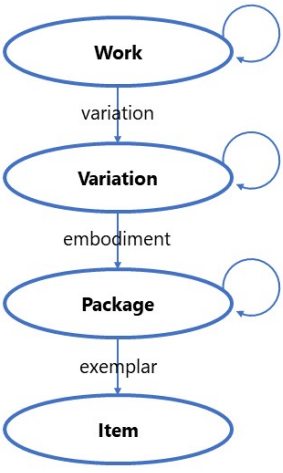

FIG. 2. Main entities of metadata model.

These entities correspond to work, expression, manifestation, and item (WEMI) of IFLA LRM.

In this model, work is a class that indicates game content based on individual creations. In this study, work records are created based on Wikidata. The data registered are instances of video games and video game series classes in Wikidata.

Variation is identified by video game content itself. It is an interpreted class of expression in IFLA LRM. They are recorded in a new record of variation when the languages and platforms of the games are different. If the special version has the same game content as the original version, they are recorded in one variation record. The bibliographical unit of variation can be said to be based on differences in computer programs. Variation helps in identifying similar content in the distribution package.

Package is identified by the distribution unit of video games. It is an interpreted class of manifestation in IFLA LRM. The versions such as bonus version, special version, and cheap version form new records of manifestation. In addition, career differences also form a new record, such as 5-inch floppy disks and 3.5-inch floppy disks. The difference between package and variation is whether they focus on their physical or content aspects.

Item is defined as "an object or objects carrying signs intended to convey intellectual or artistic content (Riva et al. 2017, p. 27)." It is a class that records individual actual objects. It is recorded based on the format managed by the collector.

Figure 3 indicates applying a case of "DARK SOULS 3" for the metadata model. 5 packages are listed in the figure. Two of these packages have the same content. One is a disk version and the other is an online resource version. The two packages are integrated to one variation. Furthermore, the four variations listed are integrated into a single work. The work record is created from Wikidata (Q20112508). The Wikidata record has several authority IDs. Those external links are inherited.

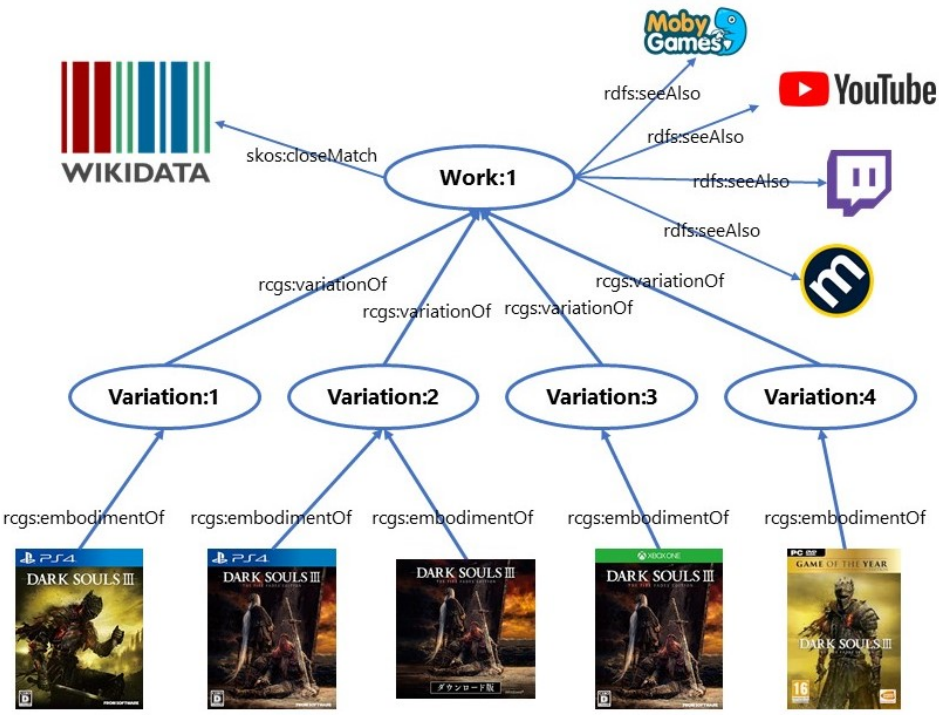

FIG. 3. Applying "DARK SOULS 3" on the metadata model

Work records were generated from Wikidata. The task of creating work by catalogers is proceeded by searching Wikipedia articles and Wikidata records that contain the relevant variation. Wikidata may be difficult to identify due to incompleteness of records. Thus, we refer to the Wikipedia description together. 16 staffs created variation records and linking with Wikidata as work at RCGS in Kyoto for four months.

All work records were obtained by the Wikidata Query Service using these Wikidata IDs. All work records were acquired by the Wikidata Query Service using these Wikidata IDs. We only use video games and video game series records. They are recorded the type on the "instance of (P31)".

4. Results

4.1 Comprehensiveness

According to Wikidata Query Service, 37,882 video games and 1,441 video game series have been registered as of June 8, 2019. This is not a small number, but it is insufficient to cover all video games. For example, MobyGames register more than 200,000 games (MobyGames n.d.)[2].

This cataloging created 3,474 work records. TABLE 1 lists the number of records cataloged for each class. The relationships of work to variation, variation to package, and package to item are listed in B. They can be denoted by 1-to-many. Expressions and manifestations are designed as an n:n relationship in FRBR / IFLA LRM. This study did not calculate the aggregation of expressions to describe compilations to describe many-to-many relationship (Riva et al. 2017, pp. 93-94).

TABLE 1. Number of records and relationships for each class of video games

Class	A. Number of records	B. Number of records having relationships with above class by one	B / A
Work	3,474	-	-
Variation	7,683	4,172 (with work)	54.3%
Package	9,383	9,104 (with variation)	97%
Item	10,930	10,930 (with package)	100%

Almost all records have relationships between variation, package, and item. Only a few packages have no relationship with variations, whose content is difficult to identify. These classes have rules for generating records, so the cataloger's effort will provide enough relationships for all resources. On the other hand, work records are generated from Wikidata in this study. Thus, we cannot create work records that are not registered in Wikidata.

As listed above, only 54.3% of the variations can be linked to Wikidata, implying that Wikidata could record only half of the video games.

4.2 Granularity

Wikidata is structured data that provides services based on Wikipedia formed by collective knowledge. Popular works are effectively recorded but not otherwise.

Wikipedia has no clear criteria for article granularity. One video game serial has multiple video games. The unit by which series are recorded is an important topic in creating abstract entity data. However, the Wikipedia guidelines in Japan state that it is entrusted to editors ("Project: Computer Game" n.d.). Additionally, the English version of Wikipedia has the following description in "Dealing with remakes" in the guideline for video game articles.

> If you can verify enough information to write a non-stub section about the distinct reception of a video game remake, as well as a non-stub section about its distinct game development or design, then the remake will qualify for its own article. However, having a separate article should not endanger the notability of the parent article. If there is not enough distinct information on the remake for a complete article, the few distinct aspects of the remake should be covered in the original game's article ("Wikipedia: Manual of Style/Video games" n.d.).

[2] This study did not adopt MobyGames for work authority because it relies on English games and the data form isn't linked open data.

The determination of article granularity is said to be delegated to editors. Therefore, the definition of granularity and bibliographic unit is not uniform.

In fact, this problem was frequently been found in the video game works created by us. Remakes are recorded on another record in popular works such as *Pokémon*. Remakes of almost all other types of video games are recorded on the original work record owing to the concentration of notability.

4.3 Completeness of records

The work class has the following properties in this cataloging:

- ID
- Label
- Title (by languages)
- Alternative title (by languages)
- Format
- Numbering of Part
- Date
- Place of Origin
- Intended Audience
- External ID
- Creator
- Production Company
- Game Play Genre
- Place in Game
- Character
- Precede Work / Succeed Work
- Series
- Other Distinguish Characteristics
- Note

We referred the properties of RDA and BIBFRAME to design. The details are published as Description Set Profile (Ritsumeikan Center for Game Studies 2019). It is expected to be recorded in these attributes for enough identification. But Wikipedia articles have a quality bias because it grew in the culture based on freedom and openness. In particular, the title and the label that consists of it are extremely important for identification.

821 records have no labels. These records are incomplete and there is no record for most of the items. In contrast, the popular game *Super Mario Bros.* has 359 property records.

4.4 Evaluation

As mentioned above, we examined the validity of Wikipedia as the work authority for several issues. Some issues were discovered in these results.

The coverage of work authority with variation is approximately half. It is assumed that it is possible to increase the number of links with work by making more effort by the staffs. However, it has the limits. It has become clear that minor games, e.g. pachinko simulated games, are not registered in Wikidata.

A description of the content of the work, such as subject or genre, is recorded in the work record. In other words, packages that are not related to the record of the work being embodied lack the description of the content. Furthermore, titles of the variations inherit from title of the works in this dataset. Therefore, about half records of variations don't have human-readable labels. This is serious problem for identification.

Also, its effectiveness cannot be denied from the viewpoint of being able to create data at a low cost. A bibliographic unit in Wikidata can also be positively regarded as a criterion that is notably based on collective intelligence. It is assumed that growth potential of Wikidata will gradually solve the issue of completeness in the future.

We must establish a plan to create a record of works that is not registered in Wikidata and register them in Wikidata. Records without labels are as well. Such activities can be more supportive of similar activities through Wikidata. But we faced challenges of granularity of bibliography we mentioned above in data enrichment. We must proceed a more detailed analysis of the game records on wikidata.

The external link ID is highly effective data that enhances the value of catalog as LOD. It is also evident that using published authority data is useful for data integration.

5. Conclusion and Future Works

In conclusion, adopting Wikidata as a work authority was found to be a valid method. The definition of work in IFLA LRM is "The intellectual or artistic content of a distinct creation". IFLA LRM positions it as a "high-level conceptual model". Therefore, it is assumed that such an abstract definition was adopted. The interpretation and creation of work for each are expected to create problems in the data integration. Wikidata provides solution of these problems for collaborative cataloging. However, its implementation is not easy. The adaptation of Wikidata features has some challenges. The adaptation of Wikidata for catalogs needs to be designed as linked data instead of a separate dataset.

Several research subjects have also been demonstrated in this study. The research topics in this study could reveal more explicit issues in a quantitative approach. Furthermore, it is assumed that it is necessary to develop the methodology of cataloging to continuously use public authorities. In addition, we need to arrange issues/articles to contribute to the community by registering enhanced metadata in Wikidata.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP19K20637, JP17K00512.

References

Agency for Cultural Affairs. (n.d.). Media Arts Database. Retrieved May 15, 2019, from https://mediaarts-db.bunka.go.jp/

Fukuda, K., & Mihara, T. (2018). A Development of the Metadata Model for Video Game Cataloging: For the Implementation of Media-Arts Database. In IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies in Session 75 - Audiovisual and Multimedia with Information Technology. (p. 11). Retrieved from http://library.ifla.org/2132/

de Groat, G. (2015). A History of Video Game Cataloging in U.S. Libraries. Cataloging & Classification Quarterly, 53, 135–156.

Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., & Vrandečić, D. (2014). Introducing Wikidata to the Linked Data Web. In Proceedings of the 13th International Semantic Web Conference - Part I (pp. 50–65). New York, NY, USA: Springer-Verlag New York, Inc. https://doi.org/10.1007/978-3-319-11964-9_4

IFLA Study Group on the Functional Requirements for Bibliographic Records. (2009). Functional Requirements for Bibliographic Records: final report. Retrieved from http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf

Jett, J., Sacchi, S., Lee, J. H., & Clarke, R. I. (2016). A Conceptual Model for Video Games and Interactive Media. Journal of the Association for Information Science and Technology, 67(3), 505–517. https://doi.org/10.1002/asi.23409

Juul, Jesper. Half-Real. Video Games between Real Rules and Fictional Worlds, 2005.

Lee, J. H., Tennis, J. T., Clarke, R. I., & Carpenter, M. (2013). Developing a video game metadata scheme for the Seattle Interactive Media Museum. Journal of Digital Library, 13(2), 105–117. Retrieved from http://mw2014.museumsandtheweb.com/paper/metadata-for-digital-video-games-at-the-seattle-interactive-media-museum/

Lee, J. H., Clarke, R. I., & Perti, A. (2015). Empirical evaluation of metadata for video games and interactive media. Journal of the Association for Information Science and Technology, 66(12), 2609–2625. https://doi.org/10.1002/asi.23357

McDonough, J. P., Olendorf, R., Kirschenbaum, M., Kraus, K., Reside, D., Donahue, R., & Rojo, S. (2010). Preserving Virtual Worlds Final Report. Library, 195. https://doi.org/http://hdl.handle.net/2142/17097

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. Digital Humanities Quarterly, 4(2). Retrieved from http://www.digitalhumanities.org/dhq/vol/4/2/000089/000089.html

Mitraka, E., Waagmeester, A., Burgstaller-Muehlbacher, S., Schriml, L. M., Su, A. I., & Good, B. M. (2015). Wikidata: A Platform for Data Integration and Dissemination for the Life Sciences and Beyond. In SWAT4LS. http://dx.doi.org/10.1101/031971

MobyGames. (n.d.). MobyGames Stats. Retrieved July 12, 2019, from https://www.mobygames.com/moby_stats

Online Audiovisual Catalogers Inc, Cataloging Policy Committee, & Video Game RDA Best Practices Task Force. (2015). Best Practices for Cataloging Video Games- Using RDA and MARC21. Retrieved from http://olacinc.org/sites/capc_files/GameBestPractices.pdf

Project: Computer Game - Wikipedia. (n.d.). Retrieved May 15, 2019, from https://ja.wikipedia.org/wiki/プロジェクト:コンピュータゲーム

Ritsumeikan Center for Game Studies. (2019). DSP for RCGS Collection. Retrieved August 15, 2019, from http://www.rcgs.jp/wp-content/uploads/2019/06/DSP_ver100.xlsx

Ritsumeikan Center for Game Studies. (n.d.). RCGS Collection -Prototype-. Retrieved from https://collection.rcgs.jp/

Riva, P., Le Boeuf, P., & Žumer, M. (2017). IFLA Library Reference Model, 101. Retrieved from https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A Free Collaborative Knowledgebase. Commun. ACM, 57(10), 78–85. https://doi.org/10.1145/2629489

Metal Gear Solid – Wikidata. (n.d.). Retrieved May 14, 2019, from https://www.wikidata.org/wiki/Q6582527

Wikidata: Introduction - Wikidata. (n.d.). Retrieved May 14, 2019, from https://www.wikidata.org/wiki/Wikidata:Introduction

Wikipedia: Manual of Style/Video games - Wikipedia. (n.d.). Retrieved May 15, 2019, from https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Video_games