

Developing a Mediated Vocabulary for Video Game Research

Tracy Hoffmann¹

Abstract: This paper presents a data-based approach to video game research and discusses its potentials and limitations. It introduces how the combination of several data sources, containing metadata describing the games, can be made productive for reconstructing the release history of video games. For this purpose, a mediated vocabulary is developed which can act as basis for the data integration process.

Keywords: Video Games; Release History; Vocabulary; Data Integration

1 Introduction

Compared to other media, video games have a comparatively short history. The earliest digital games were created in research institutions such as the Massachusetts Institute of Technology during the 1960s. In the 1970s commercial digital arcade games began to appear. Not much later, the home video game systems (console) market emerged (cf. [Lo09], [Ki06], [Ju13]). Since then, a wide range of game systems and many thousand video games have been released. New forms of play (mobile games, online games) and methods of distribution (digital platforms, „Games-as-a-Service“, micro-transactions) have been introduced. Throughout their history, video game contents and their character as products have been subject to fast and constant change.

Publications and initiatives from historians are showing a growing interest in the subject of video games (cf. Arbeitskreis Geschichtswissenschaft und Digital Spiele², [NGL19], [Ch16a], [Gi10], [WW16], [An15]). Like other popular mass-media e.g. music, film and literature, video games can be expressions for how a society sees itself and reflects its history. Therefore, video games can be approached within historical research from different points of views. For example video games with historical setting are used as case studies to explore the handling of historical themes in popular culture [Ch16a], [Gi10] or their usage in history education [An15].

However, the role of video games in culture, still provides further opportunities for research. They „represent culturally situated conceptions and ideas“ [WW16]. By considering the

¹ Universitätsbibliothek Leipzig, Digitale Dienste, Beethovenstraße 6, 04107 Leipzig, Germany tracy.hoffmann@uni-leipzig.de

² <https://gespielt.hypotheses.org/>

history of the medium itself, we can reflect history in a specific phase in a specific cultural or national context[An15]. For this, Webber sees a need to explore a „variety of cultural experiences of, and contexts for, the production, consumption and circulation of games.“[WW16]

Despite the growing academic interest and the cultural and economic impact, academic research data regarding the historical development of video games is still very limited. For questions like „The presentation of World War II in the genre of First-Person-Shooter in US-American productions in the first decade of the 21. century“[An15, pp. 399] it is necessary to identify the research subject in time and space and attribute it to the responsible actors. Additionally the spatial context in which the actor is situated (US-American production) is of importance. The metadata which is needed to define the scope of research is manifold. For instance, release dates, editions, responsible developers, genres or the relation of a game to a franchise. Presently, the only way to collect all information about one game³ or a series of interest is to manually collect data from multiple sites. As Lee et. al mention, this struggle is the same for gamers and library catalogers: „As a result, users often have to jump to multiple places to find and cross-check different types of information from these multiple sites.“[Le13]

To support video game research, the diggr (Databased Infrastructure for Global Game Culture Research)⁴ project aims to integrate and interlink datasets to construct an extensible knowledge graph for video games. This paper focuses on the conceptual issues and proposes a mediated vocabulary which can act as basis for the data integration process of online video game datasets. After discussing related work, we will present three video game datasets, which provide comprehensive game information for the knowledge graph. Then, we will discuss the video game vocabulary and show how it can be aligned with established upper-ontologies and finally point to future research possibilities.

2 Related Work

Linking heterogeneous databases is a common task and field of interest in many research domains. Depending on the domain and the specific data, different strategies are used. An approach with an aim similar to ours has been taken by Gawriljuk, Gleb et al. They build a comprehensive knowledge graph of artist information „from data spread over multiple data sources“[Ga16] to provide an integrated view on the data. To solve this task they are using specific linking techniques and align the data to a domain ontology.

Such a domain ontology is an important starting step in building a knowledge base. In the video game research domain, previous proposals for video game data models discuss the use of standards out of the archive, museum and library domain. Cultural heritage institutions have been dealing with video game collections for years and impressive

³ e.g. see the release history and introduction of Silent Hill 2 [Ne10, pp. 15-19]

⁴ <https://diggr.link>

collections are gathered in places all over the world⁵. However, there is no commonly accepted metadata standard to describe video games. Recent studies have focused on developing data models for a particular part of video game culture or use context, such as bibliographic information [Le13][Je16][Gr15][FM18], preservation [Mc11][Wi11] or in-game events (see <http://vocab.linkeddata.es/vgo/>).

Fukuda points out that „especially the FRBR model is the axis of these researches.“[FM18] The Functional Requirements for Bibliographic Records (FRBR) and now its successor IFLA Library Reference Model (LRM), is the most referred model when it comes to development of a video game data model. It is based on an entity relationship model and provides a conceptual model for the bibliographic universe. However, when it comes to domain-specific needs for video games, FRBR seems not adequate for all in this field. Jett et. al argues that FRBR is not suitable for video games „because video games arguably do not belong to a bibliographic universe“[Je16]. In conclusion, we cannot reuse an existing ontology for our use case because the modelling of video game data is still under development and no approach fulfills our requirements but we can reuse definitions.

3 Mediating Heterogeneous Data Models/Concept Spaces

The use of Linked Data provides a flexible technique to create a vocabulary, which integrates different concepts of a domain. It has the advantage of being open to future extensions with other concepts (e.g. game content actions, reception of games.). In addition, the use of the Resource Description Framework (RDF) helps to deal with inconsistency in data (e.g. naming and dates).

To build a vocabulary for our domain, we follow a bottom-up approach which is presented in figure 1. We first investigate the data structure and the overall concepts of the source databases. To work with the concepts of the source databases we create proxy vocabularies that reflect the data models. A proxy vocabulary consists of terms identified by proxy URIs⁶ for each concept in the data source. Following this, we create a domain vocabulary that reflects the common concept space of our domain as found in the source databases. This domain vocabulary reuses existing terms from related work as much as possible and makes adjustments where necessary. In a third step we integrate the proxy vocabularies by aligning them with our mediated domain vocabulary. Finally, the mediated vocabulary is aligned with an upper-ontology to provide relations to more abstract concepts. During the alignment process between the proxy and the mediated vocabulary, we have discovered biases and inconsistencies which will be discussed in the following sections.

⁵ Strong Museum of Play, Rochester, NY, USA;
Computerspielemuseum, Berlin, D;
National Videogame Arcade, Sheffield, UK

⁶ <http://patterns.dataincubator.org/book/proxy-uris.html>

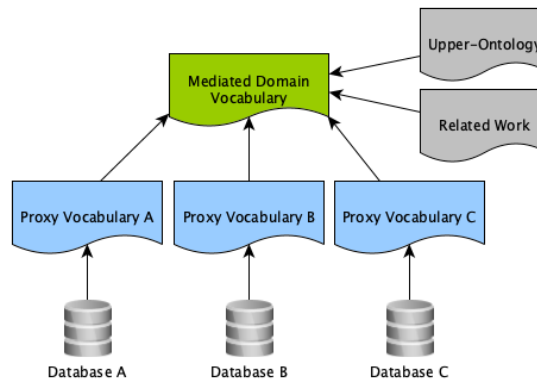


Fig. 1: The bottom-up strategy extracts the sources' data models as proxy vocabularies and aligns them with a mediated domain vocabulary. The mediated domain vocabulary is established by harmonizing the concepts from the proxy vocabularies and related work and is aligned with an upper-ontology.

4 Data Sources

First we had to identify suitable data sources. Nowadays, video game researchers can find an impressive amount of digital databases about video games. Some are very specific, some very broad, and some focus on specific areas of gaming culture. On one side, there are authoritative sources from cultural heritage institutions or age rating agencies, like *Unterhaltungssoftware Selbstkontrolle (USK)* in Germany. Such institutions, though, only provide data since their inception (1994 for USK, 2002 for CERO) and only for their specific region. They do not provide information about earlier video game history or the global distribution of games.

On the other side, there are fan-based databases, which provide much more detailed and specialized data about video games than the authoritative ones. However, none of the fan-based databases cover all release information. They neither provide a comprehensive view, which might be very important for researchers depending on the research question. In our research we focus on three video game databases which are promising in terms of covering most of the video game releases in question:

- Mobygames was founded in 1999 and grew to one of the biggest collections of video game data, and is supported by an active community. As Mobygames describe in their instructions: „Basically we're documenting how you can obtain a game [...]“⁷ they provide single data entries for each edition of a game (e.g. collector's edition, red edition).⁸

⁷ https://www.mobygames.com/info/standards#New_Entry

⁸ Not mentioned here: DLCs are also separate records in Mobygames

- The GameFAQs Website provides „game information, codes, walkthroughs, hints, message boards, save games files, and of course, FAQs“, on a platform level which means the information are separated between different hard or software platform versions of a game.
- The Agency for Cultural Affairs in Japan operates the Media Art Database. It is the only one comprehensive database including bibliographic records of video games in Japan and provides almost 100% coverage of all console game releases published in Japan since the 1980s. The Media Art DB uses the term game package and defines each entry as follows: „The download version and package version, budget version (The Best version etc.) and standard platform etc. are each allocated different GPIRs, and handled as separate titles“⁹. In conclusion this database lists every Japanese release as a single entry.

5 Mediated Vocabulary

As the short descriptions above indicate, each database has a different concept for their *game* records. Furthermore, a single game (e.g. Dark Souls, Metal Gear Solid V) is not only expressed differently in each database, relevant information can also be spread across multiple entries, without any clear indication of their relationship (different releases, or versions, of the same game). The mediated vocabulary must be capable of defining these relationships between them.

The mediated vocabulary is defined by four main classes:

Games The term *game* is one of the most abstract terms when we talk about video games. It can be compared with the entity work in IFLA LRM, which „is perceived through the identification of the commonality of content between and among various expressions“ [RLBŽ17]. The class *Game* comprises „characteristics that are typically recognized by users when they say ‘we played the same game’ “ [Je16] or „X is a remake of this game“.

All three data sources claim that they list video games but all provide different conceptual access points. As a result, we can not align any instance of the data sources to the game class. We have to construct an instance which represents this abstract class. By using other data sources we can provide instances of this class in the future or have to reconstruct one from the aggregated data.

Platform Realizations Many popular video games are developed for more than one platform (e.g. Sony PlayStation 2, Nintendo Switch). To provide access via the technical perspective to a game, like GameFAQs do, we define a *Platform Realization*

⁹ <https://mediaarts-db.bunka.go.jp/help/gm/help.html>

class which is the technical realization of a game for a specific hard or software platform. Each instance of this class has a relation to Platform.

Because GameFAQs describes their data on this level, each *GameFAQs Data* is also a subclass of *Platform Realization* (see figure 2).

Editions For marketing reasons, video games are often published in different editions (collectors edition, gold edition, etc.). Such editions do not differ with regard to the actual game, but they include additional merchandise items. Other editions include already published material like Add-ons, Downloadable Contents (DLCs) or other updates. A popular distribution strategy are Re-releases, Remastered editions or HD Editions for newer platforms which are revised or redesigned versions of a game.

Generic Edition is implemented as a superclass for grouping several creative or economic variations of a game together. Because of the aforementioned marketing strategies subclasses of the *Generic Edition* are *Special Editions*, e.g. Day One Edition or Gold Edition. With RDF, extensions are easy as we don't know what kind of specialized editions of games are created in the future. Nevertheless, *Edition* is a difficult term and complicated concept to comprehend.

As a result, instances of the class *Mobygames Games* has also have one of the *Edition* classes (see figure 2).

Local Releases Depending on video games success and fan-base, developers and publishers of video games follow different publishing strategies. For instance, many games have been released only in one country or region. Some are localized elsewhere only after years and with various changes in title, numbering, or even content. *Local Releases* are often edited to fit for a specific region or market. Jett et al. define a *Local Release* as an edition of a video game which is made „available and accessible in a particular region and in a particular language“ [Je16]. We see *Local Releases* as a grouping class which incorporates a specific *Edition* (Generic or Special) and a specific *Platform Realization* and is published in a particular region and with particular language options.

By this definition, all instances of the Media Art Databases are *Local Releases*. This results in a subclass of the *Local Release* class (see figure 2).

Our vocabulary expands the entities of Jett et al. by adding the concept of *Platform Realization*. With this step the hardware platform is no longer just a property. The platform has a huge impact on the representation, gameplay and reception of a game. This is especially true for old games. Use cases for this are mainly described in the field of game preservation, e.g. by Helen Stucky [St14] and James Newman [Ne12]. For our research it is also important to differentiate between *Platform Realizations* since this practice can lead us to different responsible actors. By adding the *Platform Realization* the *Local Release* needs a relation to this class and to *Edition* as well.

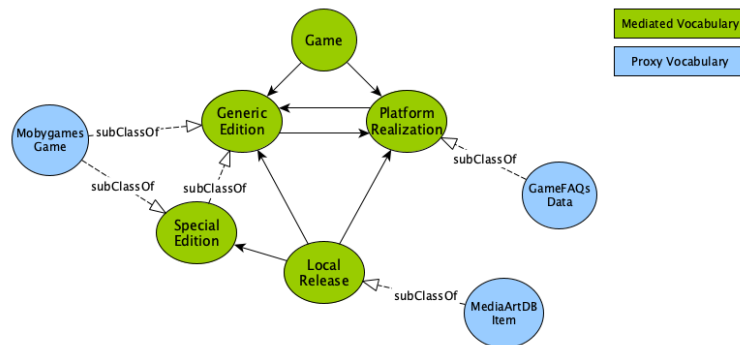


Fig. 2: The mediated vocabulary defines four main classes (Game, Generic Edition, Platform Realization and Local Release). The relation with the proxy vocabulary is provided by the use of `rdfs:subClassOf`.

6 Alignment with upper-ontologies

The alignment with an upper-ontology or top-level ontology can provide connectivity to other kinds of video game research or even other domains. They provide more general or abstract categories and concepts and increase the data interoperability.

As mentioned before, FRBR is the most frequently referenced conceptual model in related work. From a bibliographic point of view, it might be enough to use an entity relationship data model because it is adequate for static information. However, video games are not „static, ever-existing things that come from nowhere.“[Ch16b] Video games are results of mass production processes but also refer to immaterial content, in-game events and cultural artifacts like music and art. They incorporate many works from different actors, which can also be published and consumed outside the video game itself. Against this background, it appears sensible to align our video game vocabulary also with an ontology beyond bibliographic information.

The CIDOC Conceptual Reference Model (CRM) is used in the museum domain and increasingly in other contexts, especially as a way of modeling knowledge and resources in research projects. It is an object-oriented model and has an event based character. It works as a high-level ontology and is complex and abstract. With domain specific extensions with respect to the underlying concepts of CIDOC CRM, it is possible to provide an interchangeable model with the aim of facilitating widespread usage. An example for this approach is the symogih.org ontology[BR16] that emerged in the historical research domain.

The focus on events is useable for the domain of video games. A lot of video game related information is event-based, for instance, a release or the announcement of a game for a specific platform in a specific country on a specific date. Gameplay and narrative are also rich in events (see [CY08]). Based on these considerations, we could use the best

of both worlds. We use FRBR, respectively its successor IFLA LRM with the focus on published material and CIDOC CRM with its event-based character and its openness for a wider historical view. FRBRoo is an object-oriented Model of FRBR resulting from a harmonization with CIDOC CRM. Currently, we suggest an alignment of the top concepts: Game to F15 Complex Work, Platform Realization to F22 Self-Contained Expression, Generic Edition to F24 Publication Expression and Local Release to F3 Manifestation Product Type. This approach needs additional research and evaluation, since the complex nature of video games and FRBRoo would lead us fast to potential mapping issues. We are also aware of LRMoo, the successor of FRBRoo, which we will take into account in the future.

7 Conclusion

In this article we have proposed an approach to a video game vocabulary capable of describing video games sufficiently for multiple contexts, and of integrating heterogeneous video game databases. Our vocabulary allows us to integrate complementary information on specific video games from multiple sources, thus expanding the contextual information available to researchers substantially. As soon as we are interested in the differences between releases, editions, platform realizations, or want to gain an overview of the release histories of multiple titles or specific genres, the various online databases and their complementary information become pivotal resources. Our vocabulary provides a way of accessing and combining these resources, thus opening new paths for video game research.

The example in figure 3 shows an excerpt of the release information from the video game Dark Souls aligned with our conceptual approach. Dark Souls provides an example for a game which has been released on different platforms in several editions, also with additional content over time. Recently a remastered edition years after the original game was released. The main developer, the Japanese company FromSoftware, has launched a successful series with this game, which has led to a genre-defining (*Souls-like*) work. In this case, GameFAQs provide the most accurate release information about Dark Souls but don't provide much further data, for instance data about companies involved in production of the game. By combining this source with data from Mobygames and Media Art Database, we get an almost comprehensive view about the production, distribution and release history of the game. By investigating the series, we can recognize a constant internationalization of involved companies over time which shows expressions for modern production and distribution practices.

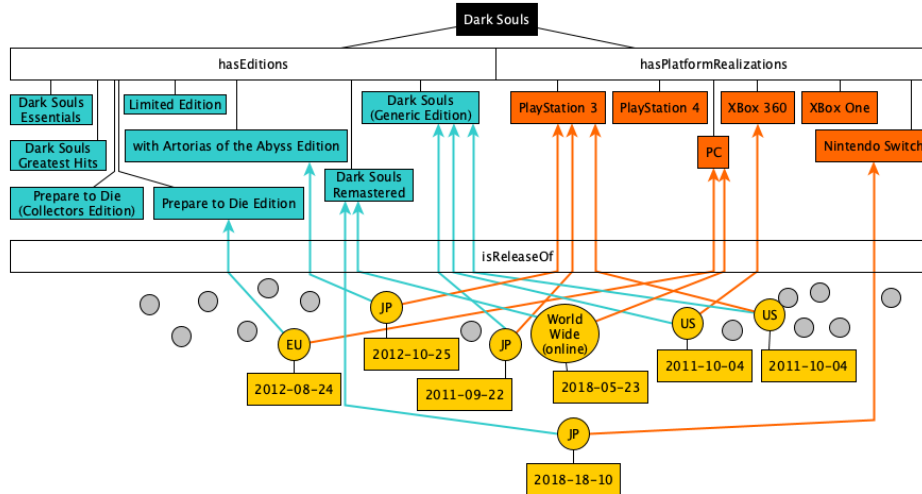


Fig. 3: After applying the conceptual approach to the release information about the video game Dark Souls across the three data sources, we can see the conceptual differences of the distribution.

There are many challenges to be considered, for instance dealing with duplicated or contradicting data (different release information on different platforms). Future research must also consider other concepts not mentioned in this paper like series or franchise. As described above, the alignment to upper-ontologies must be evaluated and further developed as well.

Nonetheless, we hope this approach contributes fruitfully to the ongoing discussion regarding video game data. The mediated vocabulary, called diggr Video Game Vocabulary¹⁰, currently comprises 12 classes and 27 properties. It is published on Github under CC0¹¹ in order to facilitate collaborative development opportunities. We also have released a human-friendly documentation under <https://diggr.github.io/diggr-video-game-vocabulary/>. Likewise, the proxy vocabularies are published on Github, one vocabulary per source.

Doing so, we hope to contribute to an ontology of video games which can be extended with other concepts of different video game research domains, and to building a knowledge graph for the history of video games in a structured and semantic way.

¹⁰ <https://github.com/diggr/diggr-video-game-vocabulary>

¹¹ <http://creativecommons.org/publicdomain/zero/1.0/>

References

- [An15] Angela Schwarz: Game Studies und Geschichtswissenschaft. In (Klaus Sachs-Hombach, Jan-Noel Thon, ed.): *Game Studies : aktuelle Ansätze der Computerspielforschung*, pp. 398–447. von Halem, Köln, 2015.
- [BR16] Beretta, Francesco; Riechert, Thomas: Collaborative Research on Academic History using Linked Open Data: A Proposal for the Heloise Common Research Model. *CIAN-Revista de Historia de las Universidades*, 19(0):133–151, 2016.
- [Ch16a] Chapman, Adam: *Digital Games as History*. Routledge, New York, NY: Routledge, 2016. | Series: Routledge advances, 2016.
- [Ch16b] Chryssoula Bekiari; Martin Doerr; Patrick Le Bœuf; Pat Riva: Definition of FRBRoo: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism. 2016.
- [CY08] Chan, Jupiter T. C.; Yuen, Wilson Y. F.: Digital game ontology: Semantic web approach on enhancing game studies. In: *Conceptual Design (CAID/CD)*. pp. 425–429, 2008.
- [FM18] Fukuda, Kazufumi; Mihara, Tetsuya: A Development of the Metadata Model for Video Game Cataloging: For the Implementation of Media-Arts Database. In: *IFLA WLIC*. Kuala Lumpur, 2018.
- [Ga16] Gawriljuk, Gleb; Harth, Andreas; Knoblock, Craig A.; Szekely, Pedro: A Scalable Approach to Incrementally Building Knowledge Graphs. In (Fuhr, Norbert; Kovács, László; Risse, Thomas; Nejd, Wolfgang, eds): *Research and Advanced Technology for Digital Libraries*. Springer International Publishing, Cham, pp. 188–199, 2016.
- [Gi10] Gish, Harrison: Playing the Second World War: Call of Duty and the telling of history. *Eludamos. Journal for Computer Game Culture*, 4(2):167–180, 2010.
- [Gr15] Greta de Groat; Eric Kaltman; Marcia Barrett; Christine Caldwell; Glynn Edwards; Henry Lowood; Noah Wirdrip-Fruin: , Core Metadata Schema for Cataloging Video Games Version 1: Game Metadata and Citation Project (GAMECIP) Tech Report 1, 2015.
- [Je16] Jett, Jacob; Sacchi, Simone; Lee, Jin Ha; Clarke, Rachel Ivy: A conceptual model for video games and interactive media. *Journal of the Association for Information Science and Technology*, 67(3):505–517, 2016.
- [Ju13] June, Laura: For Amusement Only: The life and death of the American arcade. <https://www.theverge.com/2013/1/16/3740422/the-life-and-death-of-the-american-arcade-for-amusement-only>, 2013. Accessed: 01.03.2019.
- [Ki06] Kirriemuir, John: Understanding Digital Games. In (Rutter, Jason; Bryce, Jo, eds): *A History of Digital Games*. SAGE Publications Ltd, pp. 21–35, 2006.
- [Le13] Lee, Jin Ha; Tennis, Joseph T.; Clarke, Rachel Ivy; Carpenter, Michael: Developing a video game metadata schema for the Seattle Interactive Media Museum. *International Journal on Digital Libraries*, 13(2):105–117, 2013.
- [Lo09] Lowood, Henry: Videogames in Computer Space: The Complex History of Pong. *IEEE Annals of the History of Computing*, 31, 2009.
- [Mc11] McDonough, Jerome P.: Packaging videogames for long-term preservation: Integrating FRBR and the OAI reference model. *Journal of the American Society for Information Science and Technology*, 62(1):171–184, 2011.

- [Ne10] Neitzel, Britta: 'See? I'm real...': Multidisziplinäre Zugänge zum Computerspiel am Beispiel von 'Silent Hill', volume 4 of Medien'Welten. Lit, Münster, 3., unveränd. Aufl. edition, 2010.
- [Ne12] Newman, James: Best Before: Videogames, Supersession and Obsolescence. Routledge, London, 1st edition edition, 2012.
- [NGL19] Nooney, Laine; Guins, Raiford; Lowood, Henry: Introducing ROMchip. ROMchip, 1(1), 2019.
- [RLBŽ17] Riva, Pat; Le Boeuf, Patrick; Žumer, Maja: IFLA Library Reference Model. International Federation of Library Associations and Institutions, 2017.
- [St14] Stuckey, Helen: Exhibiting The Hobbit: A tale of memories and microcomputers. Kinephanos: Journal of Media Studies and Popular Culture, pp. 90–104, 2014.
- [Wi11] Winget, Megan A.: Videogame preservation and massively multiplayer online role-playing games: A review of the literature. Journal of the American Society for Information Science and Technology, 62(10):1869–1883, 2011.
- [WW16] Wade, Alex; Webber, Nick: A future for game histories? Cogent Arts & Humanities, 3(1):1, 2016.