

Domain Analysis for a Video Game Metadata Schema: Issues and Challenges

Jin Ha Lee*, Joseph T. Tennis, Rachel Ivy Clarke

Information School, University of Washington
Mary Gates Hall, Ste 370, Seattle, WA 98195
{jinhalee, jtennis, raclarke}@uw.edu

Abstract. As interest in video games increases, so does the need for intelligent access to them. However, traditional organization systems and standards fall short. Through domain analysis and cataloging real-world examples while attempting to develop a formal metadata schema for video games, we encountered challenges in description. Inconsistent, vague, and subjective sources of information for genre, release date, feature, region, language, developer and publisher information confirm the importance of developing a standardized description model for video games.

1 Introduction

Recent years demonstrate an immense surge of interest in video games. 72% of American households play video games, and industry analysts expect the global gaming market to reach \$91 billion by 2015 (GIA, 2009). Video games are also increasingly of interest in scholarly and educational communities. Studies across various scholarly disciplines aim to examine the roles of games in society and interactions around games and players (Winget, 2011). Games are also of interest to the education community for use as learning tools and technologies (Gee, 2003). Thus we can assert that video games are entrenched in our economic, cultural, and academic systems.

As games become embedded in our culture, providing intelligent access to them becomes increasingly important. Effectiveness of information access is a direct result of the design efforts put into the organization of that information (Svenonius, 2000). Consumers, manufacturers, scholars and educators all need meaningful ways of organizing video game collections for access. Current organizational systems for video games, however, are severely lacking. What organizational challenges emerge due to the unique nature of video games? How does the lack of standardization affect access to these games? A collaborative domain analysis for the development of a metadata schema specifically for video games reveals issues inherent in this domain.

2 Challenges and Critical Literature Analysis

Current models of video game organization come from two divergent sources: the field of knowledge organization which specializes in arranging, describing, and presenting metadata for information objects and collections, and video game information from commercial systems on the internet.

Describing non-book artifacts—like video games—with knowledge organization standards has long been problematic. Hagler (1980) observed that imposing book-based characteristics on non-book materials creates inapplicable and unusable standards. Leigh (2002) notes this approach often leaves materials described by form rather than content. Even newer models like *Functional Requirements for Bibliographic Records* (FRBR) do not cover all types of materials and works: work, expression, manifestation, or item cannot be determined easily in a classic computer game (McDonough et. al., 2010a). Attributes derived from context, like mood or similarity to other objects—perhaps significant for video games—are not represented in the FRBR model (Lee, 2010). Other existing standards are similarly problematic. *Library of Congress Subject Headings* (LCSH) contains only 214 headings for describing video games by name (e.g., *Halo*, *Legend of Zelda*), with notable series missing (e.g., *Final Fantasy*, *God of War*). LCSH includes only 5 terms for computer game genre, limiting the ability to describe and therefore search or browse by genre.

Recent interests in video game preservation suggest metadata description as a preservation strategy (Winget 2008; McDonough et. al., 2010b). Besides emphasizing preservation rather than description, these projects focus on domain analysis from a data- or creator-centric point of view, rather than end users. Currently the only systematically designed game-specific descriptive framework comes from Huth (2004). However, this schema only addresses historical game systems, and like the previous examples, does not provide for the needs or behaviors of users. This limited understanding and focus of domain analysis of video games impedes development of useful information systems that meet the needs of real users.

Video game organization and description also comes from commercial systems on the internet. The web contains massive information about video games, scattered across many sites and sources. Websites such as Amazon, GameStop, GameFly, etc. are generally geared toward purchase decisions and mostly provide basic elements like title, genre, platform, release date, and publisher. Other sites provide abundant descriptive information, but it is often unstructured, cumbersome to navigate, and unverified. Users may have to visit multiple sites to find and cross-check information. All these challenges indicate the need for a more formal and standardized representation of video games based on a user-centered domain analysis approach.

3 Domain Analysis

At the University of Washington Information School, we are collaborating with the Seattle Interactive Media Museum (SIMM) to develop a metadata schema for describing all aspects of video games for improved organization, access, and preservation.

The SIMM aims to contribute to the aggregation, research, preservation and exhibition of interactive media culture and the physical, digital, and abstract artifacts therein. In 2011, the authors, SIMM colleagues and selected students participated in a special topics course “Video Game Metadata” at the Information School. This course offered opportunities for students interested in organizing video games to collaborate with the authors and the SIMM founders to get hands-on experience creating a metadata schema that will be used in real life.

The bulk of the course focused on document- and user-based domain analysis activities to determine metadata elements crucial for describing video games. First, 5 different personas epitomizing the most common types of game players and consumers potentially interested in the SIMM were developed to represent the needs, behaviors, and goals of that particular user group (Cooper, 1999): Player, Parent, Collector, Academic, and Game Developer/Designer. Once these personas were described in detail, we recorded metadata elements essential to each persona and compiled them into one list. From this, the class distilled a set of 16 core elements perceived to be most useful to all 5 personas. The CORE included Title, Edition, Platform, Format, Developer, Retail Release Date, Number of Players, Online, Special Hardware, Genre, Series/Franchise, Region, Rating, Language, and UPC. We report on the schema in more detail elsewhere (Lee et al. under review). Here we highlight and discuss problems that arose during our domain analysis.

4 Discussion

After deciding upon the CORE elements in our metadata schema, the class spent several weeks cataloging video games to test the schema’s usability and the domain analysis. As we worked, we identified several challenges for description, some unique to video games and others shared by other non-textual information objects.

4.1 Inconsistent, Vague, and Undefined Genre Labels

Genre is one of the few elements that describes content of a game rather than descriptive features (e.g., title, platform). Therefore it seems immensely useful for browsing a video game collection as well as finding new games to play. As we investigated hundreds of labels from different sources offering genre classification, it became evident that the genre metadata across these websites significantly vary with regards to the types and granularity of the terms. Most websites did not provide definitions for the genre labels, and those that did do not match across other sites. For instance, on Mobygames, both *Super Mario Bros.* and *Grand Theft Auto* are classified as “action” although most people would agree that they significantly differ. We found these current labels too broad and vague to be of use.

Establishing a controlled vocabulary for video game genres is an iterative process. We started with field-testing the cataloging process. We established a controlled list of genre and style labels taken from a number of websites related to games. We established instructions allowing multiple labels in an attempt to provide more specific

information about game content. But this too was problematic: not only did it not solve the issue of label ambiguity, it introduced a new issue of how to order the multiple genre labels in a meaningful way. Due to this, we are pursuing further work in this area by working on a faceted scheme for video game genres.

4.2 Lack of Reliable Source for Retail Release Date Information

A game's release date was agreed to be important for all the user personas. However, as we cataloged the games, a lack of reliable source for this information became evident. The only date information obtainable from the game itself is the copyright date. Using copyright information for the release date is problematic, especially for games that belong to a series, because the copyright date typically indicates the date the series was first published and does not apply to the later games in the series.

We explored different ways to obtain this information. First, we reviewed the websites mentioned in 4.1. Using these multiple sources to find and cross-check the release date for the game worked for some cases, but we did occasionally find conflicting information. For instance, the release date for the North American version of *Shenmue* on Wikipedia is November 6, 2000, as opposed to November 7 on GameSpot, and November 8 on Allgame. While the difference in date might be insignificant for average users, it poses problems for identifying and preserving games from an organizational point of view like the SIMM's. The most reliable source of release date information came from game companies' websites, although many did not carry information about every game that they published. We contacted some companies (e.g., ATLUS) and were told that there is no single person managing that information. We suspect this may be a common issue because many game companies are short-lived or merge with other companies.

4.3 Inconsistent and Marketing-Oriented Description of Features

Game features was a highly debated metadata element and eventually excluded from the CORE elements. While we agreed that game features were a valuable addition, obtaining consistent information made cataloging difficult and time-consuming. Commercial websites do include feature descriptions, but the source of this information is often unknown. Some websites, like Allgame, have their own list of features whereas others do not list feature information at all.

During our cataloging exercise, most of the class used the features element for a variety of information that was potentially useful but unable to be represented in any other element. Thus the features element ended up more like a traditional "notes" field. Through faithfully transcribing features, we learned that many sources contain text geared toward marketing rather than objective description (e.g., "Unleash over 100 mind-blowing spells" from *Disgaea*). We concluded that either this element needs controlled vocabularies from which features can be chosen, or it should be uncontrolled to include any information that catalogers think would be useful for users.

4.4 Unclear Boundaries for Region and Language

Region information is necessary for players because most console games are locked to particular geographic areas. Some games, such as smartphone apps, are free of regional restrictions, but can still be targeted for particular language-speaking audiences. Thus it can be unclear how to describe the “region” of a game. In some cases, a game is released in a country without being localized, meaning a Japanese game can be released in Korea without being translated into Korean. If so, should the region information include Japan as well as Korea? There are also cases where a game is available in multiple languages although it is still locked to particular region: for instance, a game originally released in Japan and later published in North America may have an option for Japanese subtitles and/or voice track. In this case, should the main language be Japanese or English? These cases suggest a need for detailed rules to describe language and region information.

4.5 Difficulties in Distinguishing Developer vs. Publisher

Game containers usually have various logos representing companies involved in production. Unless other sources were consulted, we found it difficult to determine which companies represent publishers vs. developers. This is further complicated because some companies are publishers as well as developers. Sometimes this information can be found in the manual, but this was not consistently true for all cases. For older games, some companies have dissolved, making it difficult to find any information. In addition, there are multiple ways of describing a company (e.g. Nintendo, Nintendo Corp., Nintendo US), implying a need for better analysis of companies and a controlled vocabulary of organization names.

4.6 Other Issues

There were several other issues in describing games: mismatching titles and numbering of games released in multiple regions (e.g., *Final Fantasy VI* in Japan released in North America as *Final Fantasy II*); multiple titles and other names by which a game is known (e.g., *The Legend of Zelda* vs. *Zelda*); denoting actual differences among different versions/editions of games (e.g., Special, Classic, Limited, etc.); difficulty of determining series information unless the cataloger is familiar with the game, to name a few.

5 Conclusion and Future Work

The issues described in this paper emerged from our first step in creating a formal metadata schema for describing video games and interactive media. Through domain analysis and cataloging, we encountered several challenges, many unique to video games. These confirm the need for a standardized description for games, including metadata element definitions, instructions for description, and controlled vocabular-

ies. We plan to further develop our schema by extending the CORE set of elements by defining a larger “recommended” set of potential use to gamers and developing controlled vocabularies for particular elements such as genre and publisher. Additionally, we plan to conduct systematic user studies to discover which information elements are perceived as useful and necessary for end-users such as gamers or parents of young gamers. We can also conduct quantitative analysis of metadata element frequencies to complement the user studies (cf. Tennis, 2003). We believe that our end results will be useful for any game related organizations: not only libraries, archives, and museums with video games in their collections, but also commercial enterprises like game developers, manufacturers, and distributors. Improving organization and access will enhance people’s gaming experiences and also have substantial commercial and cultural consequences.

6 References

1. Cooper, A.: *The Inmates are Running the Asylum*. Sams, Indianapolis (1999)
2. Gee, J.P.: *What Video Games Have to Teach Us about Learning and Literacy*. Palgrave Macmillan, New York (2003)
3. Global Industry Analysts. *Video Games—A Global Strategic Business Report*. http://www.strategyr.com/Video_Games_Market_Report.asp (2009)
4. Hagler, R.: *Nonbook Materials: Chapters 7-11*. In: Clack, D.H.(ed.) *The Making of a Code: The Issues Underlying AACR2*. ALA, Chicago (1980)
5. Huth, K.: *Probleme und Lösungsansätze zur Archivierung von Computer Programmen—am Beispiel der Software des ATARI VCS 2600 und des C64* (Unpublished master's thesis). Humboldt Universität, Berlin (2004)
6. Lee, J.H.: *Analysis of User Needs and Information Features in Natural Language Queries Seeking Music Information*. *JASIS&T* 61(5), 1025-1045 (2010)
7. Lee, J.H., Tennis, J.T., Clarke, R.I., Carpenter, M.: *Developing a Video Game Metadata Schema for the Seattle Interactive Media Museum*. Submitted to *International Journal on Digital Libraries* (under review)
8. Leigh, A.: *Lucy Is “Enceinte”*: The Power of an Action in Defining a Work. *Cataloging & Classification Quarterly* 33, 3-4 (2002)
9. McDonough, J., Krischenbaum, M., Reside, D., Fraistat, N., Jerz, D.: *Twisty little passages almost all alike: Applying the FRBR model to a classic computer game*. *Digital Humanities Quarterly* 4(2) (2010a)
10. McDonough, J., Olendorf, R., Krischenbaum, M., Kraus, K., Reside, D., Donahue, R., Phelps, A., Egert, C., Lowood, H., Rojo, S.: *Preserving Virtual Worlds Final Report*. <http://hdl.handle.net/2142/17097> (2010b)
11. Svenonius, E.: *The Intellectual Foundation of Information Organization*. MIT Press, Cambridge (2000)
12. Tennis, J.T.: *Data Collection for Controlled Vocabulary Interoperability—Dublin Core Audience Element*. *Bulletin of the ASIS&T* 29.
13. Winget, M., Murray, C.: *State of the Archive: A Review of Video Game Archives within the United States*. In: *Proceedings of the Annual Meeting of the ASIS&T*. ASIS&T, Columbus (2008)
14. Winget, M.A.: *Videogame Preservation and Massively Multiplayer Online Role-Playing Games: A Review of the Literature*. *Journal of the ASIS&T* 62(10), 1869-1883 (2011)