School of Information Studies - Faculty Scholarship

School of Information Studies (iSchool)

3-2013

Developing a Video Game Metadata Schema for the Seattle Interactive Media Museum

Rachel I. Clarke
Syracuse University

Jin Ha Lee
University of Washington

Joseph T. Tennis
University of Washington

Michael Carpenter
University of Washington

Recommended Citation

Clarke, Rachel I.; Lee, Jin Ha; Tennis, Joseph T.; and Carpenter, Michael, "Developing a Video Game Metadata Schema for the Seattle Interactive Media Museum" (2013). *School of Information Studies - Faculty Scholarship*. 169.
https://surface.syr.edu/istpub/169

Developing a Video Game Metadata Schema for the Seattle Interactive Media Museum

Description/Abstract

As interest in video games increases, so does the need for intelligent access to them. However, traditional organizational systems and standards fall short. In order to fill this gap, we are collaborating with the Seattle Interactive Media Museum to develop a formal metadata schema for video games. In the paper, we describe how the schema was established from a user-centered design approach and introduce the core elements from our schema. We also discuss the challenges we encountered as we were conducting a domain analysis and cataloging real-world examples of video games. Inconsistent, vague, and subjective sources of information for title, genre, release date, feature, region, language, developer and publisher information confirm the importance of developing a standardized description model for video games.

Keywords

Video games, metadata schema, multimedia, interactive media, cultural artifacts, Seattle Interactive Media Museum

Disciplines

Cataloging and Metadata | Library and Information Science

Additional Information

Creative Commons License

Developing a Video Game Metadata Schema for the Seattle Interactive Media Museum

Jin Ha Lee, Joseph T. Tennis, Rachel Ivy Clarke, Michael Carpenter

Information School, University of Washington, Mary Gates Hall, Ste 370, Seattle, WA 98195, USA

Phone: 206.685.0153

Fax: 206.616.3152

E-mail: jinhalee@uw.edu; jtennis@uw.edu; raclarke@uw.edu; carp121@uw.edu

URL: http://ischool.uw.edu/

Abstract: As interest in video games increases, so does the need for intelligent access to them. However, traditional organizational systems and standards fall short. In order to fill this gap, we are collaborating with the Seattle Interactive Media Museum to develop a formal metadata schema for video games. In the paper, we describe how the schema was established from a user-centered design approach and introduce the core elements from our schema. We also discuss the challenges we encountered as we were conducting a domain analysis and cataloging real-world examples of video games. Inconsistent, vague, and subjective sources of information for title, 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 in 2010,the game industry generated $25.1 billion in revenue (ESA, 2011). Industry analysts expect the global gaming market to reach $91 billion by 2015 (GIA, 2009). This increased pervasiveness inspires design and development as well as consumer consumption, amplifying the power of the video game market in the global economy. Video games are also increasingly of interest in scholarly and educational communities. Studies of games across disciplines like computer science, communication and media studies, arts and humanities, and social sciences aim to examine the roles of

games in society and interactions around games and game players (Winget, 2011). Games are also of significant interest to the education community, with a focus on how games can be used as learning tools and technologies (Gee, 2003). Video games are entrenched in American economic, cultural, and academic systems.

As games become more embedded in our lives and culture, providing intelligent access to these forms of interactive media becomes increasingly important. Effectiveness of information access is a direct function of the intelligence put into 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 due to the challenges rooted in the unique nature of video games as cultural artifacts and the lack of efforts for standardization. As a result, many systems use different labels for metadata elements as well as different vocabularies to describe games. The objective of our study is to create a metadata schema that can capture the essential information about video games and interactive media in a standardized way. This will allow for better navigation through a game collection as well as improved interoperability across multiple organizational systems. Improving organization and access will not only enhance people's gaming experiences, but also have substantial commercial and cultural consequences.

2. Challenges and Critical Literature Analysis

Current models of video game organization come from two divergent sources. First is the contemporary field of knowledge organization, a subset of library and information science (LIS) that specializes in arranging, describing, and presenting metadata for information objects and collections. Historically, these collections focus mostly on books and similar documents, treating artifacts like video games as products of popular culture and therefore of less scholarly value.

Describing non-book artifacts with LIS standards has long been problematic. Hagler (1980) observed that imposing book-based characteristics on non-book materials leaves these items in the lurch: unlike books, video games do not come with title pages, so traditional library standards based on title pages are unusable.

An ongoing lack of principles for describing non-book items results in description based on physical form, rather than intellectual content (Leigh, 2002). This becomes especially problematic with the exponential increase in born-digital items, since their physical form itself is debatable. Items such as software, contemporary digital art, and video games are now created specifically in and for a digital, electronic environment. Descriptions based on physical form are no longer applicable to these items. Digital media are defined as much—or even more—by their performativity and interactive nature than by any physical characteristics (Reinhart, 2007).

A significant attempt to overcome some of these obstacles, *Functional Requirements for Bibliographic Records* (FRBR), was developed to be a generalized view of the bibliographic universe independent of any cataloging code or implementation (Tillett, 2004). FRBR is a conceptual model representing the entities and relationships of the bibliographic universe, where attributes and relationships of bibliographic entities are defined and described based on main generic user tasks in searching and using national bibliographies and library catalogs (IFLA Study Group on the FRBR, 1998; Kruth, 2001). In FRBR, there are four different levels of bibliographic entities: work (intellectual/artistic creation), expression (work realized in the form of notation, sound, image, etc.), manifestation (physical embodiment of an expression), and item (an exemplar of a manifestation). However, applying the FRBR model to video games presents fundamental problems: McDonough et al. (2010) tried to apply the FRBR model to a classic computer game but could not easily determine work, expression, manifestation, or item. Despite FRBR's attempt at a comprehensive set of attributes, there are still limitations due to missing characteristics germane to interactive media. Attributes derived from the context of a cultural object, like a user's reaction to an object (e.g., mood), or similarity-based relationships (e.g., similar games)--which can be significant in the context of video games--are not represented in the FRBR model (Lee, 2010). Winget and Murray (2008) also argue for the importance of collecting information related to the "context of use" for video games.

Other existing content and object description standards are similarly problematic, and illustrate the glaring lack of innovation in game description. Unlike other digital media, games lack specific controlled vocabularies for subject and genre. For instance, digital art is sufficiently served by established topical art vocabularies like the Getty's *Art & Architecture Thesaurus* and the *Thesaurus for Graphic Materials* in conjunction with the *Library of Congress Subject Headings* (Hanlon & Copeland, 2001). Video games have no such specialized indexing language, and the general-purpose *Library of Congress Subject Headings* contains only 219 headings for describing different video games by name (e.g., *Dead or Alive*, *Halo*, *Legend of Zelda*), with many notable series missing (e.g., *Final Fantasy*, *Dragon Quest*, *Mass Effect*, *God of War*). There are a mere 5 headings with regards to genre (Computer adventure games, Computer baseball games, Computer flight games, Computer war games, and Computer word games), clearly limiting the ability to describe and therefore search or browse games by genre.

Recently, the LIS community demonstrated increased interest in the preservation of video games, notably the "Preserving Virtual Worlds" project (McDonough et al., 2010) which identifies several challenges for preserving virtual worlds and suggests metadata description as a preservation strategy. Winget (2011)'s review of video game preservation literature reveals a focus on games as artifacts thus limited to traditional preservation challenges of hardware, software, emulation, and scope. In addition to an emphasis on preservation rather than description, both projects focus on game information from a data- or creator-centric point of view, rather than that of the end user.

Currently, the only systematically designed game-specific descriptive framework comes from a German master's thesis by Huth (2004), who drew existing elements from OCLC Metadata Elements (OCLC, 2003), the Dublin Core Metadata Element Set (DCMI, 2003) and added new elements. The metadata is organized into five groups: Representation, Reference, Provenance, Fixity, and Context (Anderson, 2010). But this schema only addressed early game systems not reflective of today's gaming environment, especially with regards to newer innovations like online real-time games involving multiple users. Huth's

approach, like those above, reveals a focus on historical preservation with little provision for users' needs and desire for access from their own perspectives and their behaviors. This limited understanding of video games and their users in the LIS community is an impediment to developing useful information systems that meet the needs of real users.

The second source of video game organization and description comes from commercial systems, mainly on the Internet. Although the web contains massive information about video games, it is scattered across many sites and sources. Websites such as Amazon.com, Moby Games, allgame.com, Giantbomb, IGN, GameFAQs, GameSpot, etc. are generally geared toward gamers making purchase decisions and so provide only basic descriptive elements like title, genre, release date, and publisher. Other informational websites such as Wikipedia provide a large amount of descriptive information, but it is often unstructured, cumbersome to navigate, unvetted and unverified. As a result, users often have to jump to multiple places to find and cross-check different types of information from these multiple sites.

The metadata across these websites is also uncontrolled, meaning that there is no accepted standard for describing games in a consistent manner. Without clear, comprehensive descriptions it can be challenging to collocate similar games or generate recommendations for new games based on what a user enjoyed in a previous playing experience. Many commercial game websites use their own vocabularies for describing information such as game genre. These genre labels are not formally designed according to established standards or principles, and often do not match or crosswalk across different sites. For instance, a "platform" game (where players navigate the game by making characters jump from one platform to another) is classed as a sub-genre under "action" on allgames.com, but identified as its own separate genre by IGN Entertainment (ign.com), while Moby Games fails to include it at all. Furthermore, the genre labels tend to be general and therefore too broad and vague to be of any use. The "role-playing" category on Moby Games retrieves 3,727 results, making the category impossible to browse.

Even searching for a known game is difficult because a set of primary access points for games is not clear or consistently agreed upon. With other cultural objects such as books, music, or movies, the name of the artifacts' creator, composer/performer, or director are commonly provided as primary access points, respectively. With games, however, the group of people involved in creating a game is typically very large, making attribution to a single artist or creator difficult. It is also unclear if users even know or remember the names of designers, or if they are even interested in finding games this way. In addition, other feasible access points like characters, music tracks, or motifs have not been adequately explored on the user's behalf. Indeed, it is currently unknown which elements of video games would make for the most useful access points. All these challenges indicate the need for a more formal and standardized representation of video games based on a user-centered approach.

3. Study Design

3.1. Method

At the University of Washington Information School, we have been collaborating with the Seattle Interactive Media Museum (SIMM), recently established by Andrew Perti and Michael Carpenter, to develop a new metadata schema for describing all aspects of video games for improved organization and access. While the SIMM is interested in the preservation of video games and related materials, their objective also includes aggregation, research and exhibition of interactive media culture and the physical, digital, and abstract artifacts therein, therefore implying a need for robust, media-specific metadata to serve a variety of use cases. As an emergent organization, the SIMM provides an optimal crucible for creating such a schema.

In the Autumn quarter of 2011, the authors, colleagues from the SIMM, and graduate students participated in a special topics course "Video Game Metadata" at the University of Washington Information School. The course was designed to offer interested students the opportunity to collaborate with the authors as well as the creators of the SIMM in order to get hands-on experience in creating a metadata schema for use in a real-world application.

The majority of the course focused on user- and document-based analyses in order to determine metadata elements crucial to describing video games. First, different personas epitomizing the most common types of game players and potential SIMM patrons were developed. A persona is an archetype representing the needs, behaviors, and goals of a particular group of users and using personas enables a goal-directed design of a system (Cooper, 1999). The 6 personas that emerged were Player (Jeffrey, a Junior High Student), Parent (Marcia, a Classroom Assistant and a mother of 3), Collector (Sam, a Copywriter for Amazon.com), Academic (Dr. Russell, an Economics professor), Game Developer/Designer (Debra, a Game Designer), and Curator/Librarian (Nancy, an Academic Librarian). Full descriptions of the six personas and the use scenarios are included in Appendix I. Based on these personas we created several use scenarios for the SIMM website which helped us in selecting metadata elements that would be useful for each user group.

After creating these personas and use cases, we compiled a master list of metadata elements from a number of major commercial, hobby, and review websites related to video games including Mobygame, Giantbomb, Allgame, Amazon, Gamefaqs, Wikipedia, etc. This constituted the primary form of domain analysis. This follows the method highlighted in Hjørland (2002). In his work, Hjørland identified ways domains can be studied and understood in order to create metadata. We used the extant sources of information organization listed above in order to see how the domain was shaped, what was listed, and where there were lacunae. We also brainstormed several additional elements that might be useful for the specified personas and other people interested in video games. We ended up with a list of 61 different information features and went over them one by one, trying to determine if they were necessary or potentially useful from the perspective of each persona. The following table shows what we determined to be the relative importance of each element for different personas. The solid circle denotes the features that were regarded as highly important for the persona and the unfilled circle denotes features that would potentially be helpful but not necessary. The table only shows the select elements that were deemed important for multiple personas, not the full list of 61 elements.

Table 1. The Relative Importance of Metadata Elements for the Five Personas

3.2. Limitations

Personas can be a very useful tool in design, although they are not free from limitations. Previous literature notes challenges in defining the right personas and verifying that personas accurately reflect user data (Grudin and Pruitt, 2002; Chapman and Milham, 2006). Generally, persona-based approaches for design can be most successfully used when the designers have a good understanding of the persona types that will be using the system and thus be able to see the world from the personas' points of view (i.e., perspective taking) (Bagnall et al., 2005). The fact that this research was carried out in a class environment enabled us to have at least one or more persons who actually fit into each persona type .This allowed us to make reasonable assumptions about the users' motivations and behaviors. Another limitation is that the persona type can be too broad and may not accurately represent the goals, needs, desires, and knowledge of a particular class of users, especially when inferences are made based on personas' high specificity (Chapman and Milham, 2006). Considering these limitations, the authors do not intend to solely rely on personas for creating our metadata schema. In conjunction with domain analysis from numerous game-related websites, personas can serve as a powerful tool for establishing the initial set of core metadata elements. However, this is part of a bigger project that aims to establish a more complete set of metadata elements in multiple stages. For our future steps, we will be conducting in-depth user interviews as well as a large scale survey in order to obtain more user data that would help us improve our selection and evaluation of metadata elements. Our broader intention is to take multiple user-centered design methods to maintain and continue to improve our schema over time.

4. The Core Metadata Elements

Based on the investigation of personas and different use scenarios as well as the information in Table 1, we decided to focus on establishing an initial set of core elements that should be described by any system attempting to organize video games and interactive media. Our final CORE set consists of 16 elements that were deemed useful for a range of user groups: Title, Edition, Platform, Format,

Developer, Publisher, Retail Release Date, Number of Players, Online Capabilities, Special Hardware, Genre, Series/Franchise, Region, Rating, Language, and UPC. We have borrowed and modified the definitions from existing standards such as FRBR (for Edition, Developer, Retail Release Date, and Language) and CIDOC CRM (for Title) in order to maintain some degree of interoperability for common elements. However, for other elements, we had to create definitions based on the group discussion on what each of the elements referred to, since we could not find reusable definitions from existing standards. Details of each element are described in Table 2. Through discussion, we decided that the element "system requirements" in Table 1 should be split into "online" and "special hardware" in order to reduce ambiguity.

Table 2. The 16 CORE Metadata Elements

After we decided upon the 16 CORE elements in our metadata schema, the remainder of the quarter-long class was dedicated to testing the usability of the schema. We collected 30 games and attempted to catalog them according to the schema. Games for this exercise were carefully selected to present a variety of genres, platforms, creators, and editions. Additionally, we selected games with widely varying packaging and documentation elements in order to test the efficacy of each chief source of information. The full list of games tested in the exercise is included in Appendix II.

5. Discussion

As we progressed through the cataloging exercise, many challenges for describing video games emerged. Some of the problems are unique to video games and others are commonly encountered when describing and organizing other non-textual information objects.

5.1. Information Sourcing Issues

While each CORE element includes a specific definition and chief source of information, sourcing this information still proved problematic. It is important to note that under the direction of the SIMM, we strove to describe each game at the Manifestation level of the FRBR model of description. This means that each game is described at what is widely considered the "edition" level in the domain of

video games. The reasoning behind this decision is closely related to what we intend this core set of metadata to be. CORE16 were selected as the most general and essential information about games that should be recorded for describing any video game collection. Item level description such as the condition and provenance of the item will certainly be critical in some environments (e.g., for a video game museum curator), but not in others (e.g., online database for video games; commercial websites for video games). This decision was also partly due to the fact that we are designing our schema from a user-centered approach, and our assumption is that for general users, the item level description would not be sought as often as the manifestation level description. The item level description will most likely be included as we move onto the second phase of the project which focuses on establishing a recommended set of metadata elements, an expansion of the CORE16.

Because of this stipulation from the SIMM with regards to the focus on the manifestation level, rather than describing games at the more granular Item level of FRBR (e.g. an actual game cartridge or optical disc) we identified and defined a chief source of information that would encompass many FRBR Items. Thus, for each element, the most commonly cited sources of information include the housing of material of a game, such as the box, manual, and/or cartridge. However, as previously noted, this traditional view focused on physicality becomes problematic very quickly. Some contemporary games are born digital and available to users via download, meaning they have no boxes, no cartridges, and only rarely have manuals (usually in the form of on-screen instructions rather than printed materials). This immediately challenges our established designations of chief source of information for many of our CORE elements. Many of these games come without a reliable physical source; without this, we have no access to critical metadata except by playing the game itself. For some games, we were able to find a different version released for another platform in a game box with a manual (e.g., *Plant vs. Zombies* released for Microsoft Windows). However, finding a physical counterpart is not always possible, as some games are developed exclusively as direct downloadable apps for tablets and smartphones (e.g. *Chaos Rings* for iOS/Android). Other desirable descriptive information deemed important to users could not be sourced from the games themselves, but

was only available via secondary sources. Even when secondary sources of information were found, the quality of the information varied immensely.

5.1.1. Inconsistent, Vague and Undefined Source Information

For example, "retail release date" is an element that all participants agreed to be important for all the user personas. The release date information of a game can provide a lot of contextual information about the game to users. For instance, for an RPG game, the release date can shed some insights into what kind of visual style, battle system, and the level of difficulty may be expected. The release date can also heavily affect the purchase decision of gamers, be useful for historical analysis of video game trends for scholars, and will be important for curators for preserving the accurate information about the artifacts. However, as we started cataloging examples of actual games, it became evident that there is in fact no reliable source of this information. The only date information we can obtain from the game itself is the copyright date. Using copyright date information for the release date is problematic, especially for games that belong to a particular series. Copyright date typically indicates a date when the first manifestation of the series was published and thus does not apply to any of the later manifestations. Also for early games developed prior to the early 1990s, this information is not well-documented, which makes it difficult to determine the exact date, especially for games that have multiple versions released (McDonough et al., 2010). We also had an extended discussion on how specific this information should be – in other words, is the publication year sufficient, or should we include month information, or do we need an exact date? Our final decision to preserve the exact date was due to the fact that for most current games, data at this level is usually obtainable without too much difficulty, and it is better to preserve more information than less. We acknowledge that for older games, we might only be able to obtain and record the publication year.

We explored different ways to obtain this information. First, we looked at different websites including Wikipedia, Amazon, GameSpot, GameFaqs, etc. Using multiple sources to find and cross-check the release date for the game seemed to work for some cases, but often we found conflicting information on these multiple sites. For instance, the release date for the North American version

of the game *Shenmue* on Wikipedia is November 6, 2000, as opposed to November 7 on Gamespot, and November 8 on Allgame.com. While the difference in date might be perceived as insignificant for average gamers, it does pose a problem for identifying and preserving these games from an organizational point of view, such as that of the SIMM. While specific game company websites turned out to be the most reliable source of release date information, most did not carry information about all the games that they published. This is especially problematic for games published by now-defunct companies. We contacted some game companies such as ATLUS and SquareEnix, and were told that there is no single person who manages such information to whom they could point us. We believe this is probably a common issue across game companies, especially because many are short-lived or merge with other companies.

The "genre" element also suffered from inconsistent information sources. Genre is one of the few elements that describes the content of a game rather than descriptive features. Therefore, it was perceived as the most useful information for browsing a video game collection as well as discovering new games to play. As we investigated hundreds of genre labels from different sources of genre classification for video games, it became evident that genre metadata is uncontrolled, meaning that there is no accepted standard for describing games in a consistent manner, or even correctly. On most websites, we could not even find definitions for the genre labels. Many commercial game websites use their own terms for describing information like genre, with local definitions that do not match across different sites: for instance, on Mobygames.com, both *Super Mario Bros*. and *Grand Theft Auto* are classified as "action" although most people would agree that they are very different. Most of us agreed that these current labels are general and too broad and vague to be of any use. We recommend taking a new approach in describing the genre information of games which is discussed further in section 5.3.

5.1.2. Subjective Source Information

Another issue with source information is the potential lack of objectivity. "Features" was a highly debated metadata element that was ultimately excluded from the CORE elements. The problem with this element is that it is impossible to

obtain consistent information, even with a designated chief source of information, which makes it difficult and time-consuming for catalogers and searchers. Commercial websites such as Amazon often include the description of features although it is unclear as to where this information is derived from. Some websites such as Allgame.com have their own list of features whereas others do not list any feature information at all.

During our cataloging activities, most of us ended up entering a wide variety of information for this element that could potentially be useful but not represented in any other field. Thus it ended up taking the role of a traditional "notes" field. Deciding to faithfully transcribe the features listed on the designated chief source of information (i.e., the game box in the case of the "features" element) allowed us to maintain some consistency in our data entry. However, we learned that many games contain text that is heavily geared toward marketing rather than objectively describing the features of the game (e.g., "Unleash over 100 mind-blowing spells" from *Disgaea*; "The Fun-Dead Game of the Year" from *Plants vs. Zombies*). Our discussion ended with three divergent suggestions: 1) there needs to be a list of controlled vocabularies from which catalogers can choose features; 2) this field should be left similar to the notes field where catalogers can decide to leave any information that they think would be useful for the system users; 3) this element should be populated with a verbatim transcription from the box, marketing hyperbole and all, in order to preserve the historical accuracy of the chief source of information.

5.2. Unclear Conceptual Boundaries

Several elements thought to be useful to all personas as well as the SIMM suffered from unclear conceptual boundaries. Despite clear element definitions, teasing out the differences in descriptive information from the sources of information is challenging at best. This is perhaps due to incompatibility of established descriptions, definitions and concepts with those emergent from the specific domain of video games. Video games clearly diverge from the established FRBR conceptual model; however, many bibliographic description elements are still linked to this conception and so reflected in the CORE elements. It is for this reason that we believe we can contribute to content description standards. Both

the Resource Description and Access (RDA) standard and the Cataloging Cultural Objects (CCO) standard do not explicitly address the conceptual model of video games as an entity for description. In the case of RDA, video games are considered containers of moving images and nothing more (Canadian Library Association et al., 2010). In fact there is no definition of video game in RDA. CCO (Baca et al., 2006) is likewise agnostic with regard to the conceptual boundaries and definition of video games. Cultural objects, the purview of CCO, are left to the cataloguer and his or her institution to decide. This leaves room for a more specific conceptualization of metadata required for video games, like what we offer here.

5.2.1. Region and Language

Region information is necessary for players because most of the console games are locked via hardware restrictions, to a particular region such as North America (NTSC U/C), Japan and Asia (NTSC-J), and Europe and Oceania (PAL). Some games, like smartphone apps, are free of those regional restrictions, but can still be targeted for audiences speaking particular languages. In cases like such, it can be unclear as to what to describe as the "region" of the game. There can also be cases where the game is released in a particular 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? Also there are cases where the 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 can have an option for Japanese subtitles and/or voice acting. In this case, should the main language be Japanese or English? All these cases suggest that it is necessary to have a fairly detailed rule on how to describe the language and region information.

5.2.2. Developer vs. Publisher

The box of the game usually has different names and logos representing the companies involved in producing the game. The challenge we encountered was that without consulting other online sources, it was often difficult to determine which company represents the publisher versus the developers. The problem is further complicated by the fact that some companies actually can be publishers as

14

well as developers of the games. Sometimes this information can be found inside the manual but this was not consistently true for all cases. For older games, some of the companies have already dissolved and it was difficult to find any information about the particular organization based on a company name, logo, or acronym. In addition, there are many different ways of describing the company (for instance, Nintendo, Nintendo Corp., Nintendo US) so there must be a controlled vocabulary listing the preferred form of these names of organizations.

This problem is not unfamiliar to archivists writing administrative history of a fonds (body of records created by an organization). Both Cook (1993) and Millar (2002) have described the mercurial nature of organizations. The records of an organization are meant to reflect the ordinary course of business. However, when business changes, so too does the structure of records creation. We see that manifest in this context with the changing organizational structure of game companies. A full archival contextual analysis of game companies would help make this aspect of metadata more robust and meaningful, but also more complex as we would need to represent the change in the organization over time.

5.3. Need for Better Subject Access

In most video game descriptions that are currently available on various websites and catalogs, the only prominent access point provided for any kind of subject access to the games is genre information. For our cataloging exercise, we established a preliminary controlled list of genre and style labels from which catalogers could select terms. The instructions we established allowed for selecting multiple labels in an attempt to provide more specific information about the content of the game. However, this did not solve the issue of label ambiguity, and it introduced another problem: how to order the different genre labels in a meaningful way. Examining the genre labels also made us realize that the genre element is not strictly about the gameplay or style--it is overloaded with a range of different types of information. In order to tease out these subtleties we recommend creating a faceted scheme for video game genres. Facet analysis is the process of examining a subject field and dividing it into fundamental categories, each of which represents an essential characteristic of division of the subject field (Spiteri, 1997). Some of the dimensions we identified so far include gameplay

(e.g., action, RPG, strategy), style (e.g., platformer, MMORPG, tower defense), the purpose of games (e.g., educational, party), target audience (e.g. adult, early childhood), presentation (e.g, 2D, anime/manga), temporal aspect (e.g., real-time, turn-based strategies), point of view (e.g., first-person, third-person), theme (e.g., fantasy, sci-fi), mood/affect (e.g., horror, mystery), setting (e.g., futuristic, space), and so on. We believe that by harnessing these particular characteristics, we will be able to develop systems that reveal or suggest similar games with significantly improved results. For instance, using this faceted scheme, the genre facets of a game such as *Final Fantasy XIII* can be described as follows:

> Gameplay (RPG); Style (Action RPG); Purpose (Entertainment); Target audience (Teen – ESRB); Presentation (3D); Point of view (Third-person); Theme (Fantasy); Mood/Affect (Mystery; Inspirational); Setting (Futuristic); Temporal aspect (Real-time); Type of ending (Circuitous); Visual style (Photorealism – Illusionism)

Future work will report further on this development.

5.4. Other Issues: Names, Versions, Series, and Platforms

There were several other issues in describing video games. The naming of the games was one of them. Sometimes there are mismatching titles and numbering of games that are released in multiple regions (e.g., *Biohazard* in Japan was released in North America as *Resident Evil*; *Puzzle Bobble* in Japan was based on the arcade game *Bubble Bobble* and was released in North America and Europe as *Bust-a-Move*; *Final Fantasy IV* in Japan was released in North America as *Final Fantasy II*), and there is also an issue of multiple titles and other names by which the game is known (e.g., *The Legend of Zelda* vs. *Zelda; Super Mario Bros.* vs. *Mario*). When the old game is ported into a new platform it can be given a different name (e.g. *Tales of Graces F* released for Playstation 3 in North America vs. *Tales of Graces* released for Wii in Japan). Denoting the actual difference among different versions/editions of the games (e.g., Special, Classic, Limited, Collector's, Deluxe, Super, Premium, Gold, Platinum) can also be challenging without conducting additional research on each item. Sometimes the same games are packaged and sold differently in multiple ways (e.g., *God of War Saga*

Collection vs. *God of War: Collection* vs. *God of War: Origins Collection*). We are currently discussing different options for dealing with this naming problem such as employing attributes or specifying the relationship types among different titles.

Determining the series information can also be difficult. Sometimes the numbering after the title can help, but the first published game of the series of course does not typically have any numbering associated with it. For other games within a series, there is no numbering that can directly connect the games (e.g., *Katamari Damacy*, *We Love Katamari*, *Katamari Forever*, *Beautiful Katamari*). Some games belong to multiple series: for example, *Persona 4* belongs to *Shin Megami Tensei* series (Parent) as well as *Persona* series (Child), and *Tales* games have various titles belonging to the main series (e.g., *Tales of Vesperia*, *Tales of Symphonia*) as well as the spinoff series (e.g., *Tales of the Tempest*, *Tales of VS.*).

There is also the issue of the coherence of a series. Some games make up a series because there is actually a continuation of the story that is told across multiple games (e.g., *Halo* series) whereas others are not connected in any way story-wise, but do share a similar theme or gameplay format and thus constitute a series (e.g., *Final Fantasy* series). To make things even more complicated, there are examples such as *Shadow Hearts* series – *Shadow Hearts: Covenant* (the second game in the series) is a continuation of the story told in *Shadow Hearts* (the first game), but *Shadow Hearts: From the New World* (the third game) is a completely new story featuring a similar gameplay format and battle system as the previous games. Also some spinoff series feature particular species or characters from another series (e.g., *Chocobo Racing* featuring chocobos from *Final Fantasy* series), therefore they are not connected directly with regards to the story, theme, or general gameplay format of the original series. Determining all this information related to series will require significant amount of time researching the background information of each game for the cataloger. We are considering the separation of elements Series and Franchise, or adopting another element called Universe to represent these different types of series.

Additionally, the concept of platform can be confusing to some users as the software and hardware needed to play the game are completely integrated for some consoles (e.g., Sony Playstation) but in other cases, they are separate (e.g., games developed for iOS can be play used in any devices that runs iOS including iPad, iPhone, etc.).

6. Conclusion and Future Work

The efforts described in this paper were a first step in creating a formal metadata schema for describing video games and interactive media. Through the process of exploring personas and use scenarios, selecting metadata elements, and cataloging actual games based on those elements, we encountered several challenges, some of which are unique due to the nature of video games. These challenges confirm the importance of having a standardized way of describing games including definitions of metadata elements, instructions for description, and controlled vocabularies, as well as conceptualizing a new model, specific to the video game domain. We plan to further develop our schema by continuing the following efforts: 1) extending the CORE set of elements by selecting and defining a larger "recommended" set that can potentially be useful for users of video games, and 2) developing controlled vocabularies for particular elements such as genre, publisher, etc. This second version of the schema will not only contain a larger number of metadata elements, but also incorporate hierarchical and faceted structures for some of the elements (e.g., genre, plots, visual style). Additionally, we are conducting a series of systematic user studies involving in-depth interviews in order to discover which information elements are perceived as useful and necessary for end-users such as gamers or parents of young gamers. The information we obtain from these interviews will further help us verify the importance of including particular metadata elements as well as improve the definitions and instructions provided for each element. After the completion of the recommended set in addition to the CORE16, we plan to do a more extensive evaluation of these schemas by creating a database of metadata records for a sample game collection and conducting a usability test of this database. All of these efforts will ideally move us closer to understanding the universe of games more fully, and potentially lead us to new domain-specific conceptual models that more accurately reflect and represent this space. We believe that our end results

will be useful for any game related organization: not only libraries, archives, and museums with video games in their collections, but also commercial enterprises like game developers, manufacturers, and distributors.

Acknowledgements: We would like to thank the class participants of INFX 598 Video Game Metadata and especially Andrew Perti at the SIMM for their valuable contributions to the project.

References

1. Anderson D, Delve J, Pinchbeck D (2010) Toward a workable emulation-based preservation strategy: Rationale and technical metadata. New Review of Information Networking 15(2): 110-131

2. Baca M, Visual Resources Association (2006) Cataloging Cultural Objects: A guide to describing cultural works and their images. American Library Association, Chicago.

3. Bagnall P, Dewsbury G, Sommerville I (2005) The limits of Personas. Proceedings of the 5th Annual DIRC Research Conference: 39-40.

4. Canadian Library Association, Chartered Institute of Library and Information Professionals (Great Britain), Joint Steering Committee for Development of RDA, American Library Association (2010) RDA toolkit: resource description & access. American Library Association, Chicago.

5. Chapman CN, Milham R (2006) The personas' new clothes: Methodological and practical arguments against a popular method. Proceedings of Human Factors and Ergonomics Society Annual Meeting: 634-636

6. Cook T (1993) The concept of the archival fonds in the post-custodial era: Theory, problems and solutions. Archivaria 35: 24-37

7. Cooper A (1999) The inmates are running the asylum. Sams, Indianapolis

8. Dublin Core (2003) Dublin Core metadata element set, Version 1.1. http://www.dublincore.org/documents/dces/. Accessed 20 March 2012

9. Gee JP (2003) What video games have to teach us about learning and literacy. Palgrave Macmillan, New York

10. Global Industry Analysts (2009) Video Games—A Global Strategic Business Report. http://www.strategyr.com/Video_Games_Market_Report.asp. Accessed 20 May 2012

11. Grudin J, Pruitt J (2002) Personas, participatory design and product development: An infrastructure for engagement. Proceedings of the Participatory Design Conference: 144-161

12. Hagler R (1980) Nonbook materials: Chapters 7-11. In: Clack DH (ed) The making of a code: The issues underlying AACR2. American Library Association, Chicago

13. Hanlon A, Copeland A (2001) Using the Dublin Core to document digital art. Journal of Internet Cataloging 4(1-2): 149-161

14. Hjørland B (2002) Domain analysis in information science. Eleven approaches - traditional as well as innovative. Journal of Documentation 58(4): 422-462

15. Huth K (2004). Probleme and Lösungsansätze zur archivierung von computer programmen—am beispeil der software des ATARI VCS 2600 und des C64. Unpublished master's thesis, Humboldt Universitat

16. IFLA Study Group on the FRBR, International Federation of Library Associations and Institutions, Section on Cataloging, Standing Committee (1998) Functional requirements for bibliographic records: Final report. K.G. Saur, München

17. 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)

18. Kruth M (2001) An interview with Barbara B. Tillett. Cataloging & Classification Quarterly 32(3): 3-24

19. Lee JH (2010) Analysis of user need and information features in natural language queries seeking music information. Journal of the American Society for Information Science and Technology 61(5): 1025-1045

20. Leigh A (2002) Lucy is "Enceinte": The power of an action in defining a work. Cataloging & Classification Quarterly 33: 3-4

21. McDonough J, Olendorf R, Kirschenbaum M, Kraus K, Reside D, Donahue R, Phelps A, Egert C, Lowood H, Rojo S (2010) Preserving Virtual Worlds Final Report. http://hdl.handle.net/2142/17097. Accessed 17 January 2012

22. Millar L (2002) The death of the fonds and the resurrection of provenance: Archival context in space and time. Archivaria 53

23. OCLC (2003) Digital Archive Metadata Elements. Online Computer Library Center, Dublin

24. Rinehart R (2007) The Media Art Notation System: documenting and preserving digital/media art. Leonardo 40(2): 181-187

25. Spiteri LF (1997) The use of facet analysis in information retrieval thesauri: an examination of selected guidelines for thesaurus construction. Cataloging & Classification Quarterly 25(1): 21-37

26. Svenonius E (2000) The intellectual foundation of information organization. MIT Press, Cambridge

27. The Entertainment Software Association (2011) Essential facts about the computer and video game industry. http://www.theesa.com/facts/pdfs/ESA_EF_2011.pdf. Accessed 30 March 2012

28. Tillett BB (2004) What is FRBR? A conceptual model for the bibliographic universe. http://www.loc.gov/cds/FRBR.html. Accessed 2 October 2011

29. Winget M, Murray C (2008) State of the archive: A review of video game archives within the United States. In: Proceedings of the annual meeting of the American Society for Information Science & Technology. ASIS&T, Columbus

30. Winget MA (2011) 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

Appendix

I. Five Personas and Use Scenarios

- **Player persona**

Name: Jeffrey Cunningham

Occupation: Junior High Student

Gender: Male

Education: As a junior high student, Jeffrey excels in art, wood shop, and typing classes. He plays basketball on the JV team.

Computing and Web experience: Owned his first laptop computer at age 10. Although he's not into programming, he does have considerable Web 2.0 skills. He regularly updates two separate blogs, participates in fantasy football and baseball leagues, and has 1432 friends on Facebook. Plays lots of online video games.

Personal Web behavior patterns: Likes to browse game web sites. Has special affinity for gamefaqs.com and gamerankings.com. Uses these sites mostly to inform video game purchases and obtain gameplay information.

How they will use the site: Jeffrey comes to thesimm.org mostly as a recreational activity. He enjoys the ability to view hi-resolution box art for his favorite games. He also browses the tiff files of classic game magazines. Recently he has taken an interest in researching the histories of local gaming companies he might one day apply to.

Any additional site-specific demographics: Jeffrey can't wait to get an iPad in order to peruse game magazines on the go.

- **Parent Persona**

Name: Marcia Strom

Occupation: Classroom Assistant

Gender: Female

Education: BA in Psychology

Computing and Web experience: Primarily uses a desktop at home for email. Will browse the web to shop on occasion and is a whiz at using travel sites to find

good airplane fares.

Personal Web behavior patterns: Marcia does not use a large variety of websites. Once she is comfortable with a certain interface it takes a lot to pull her away to another one. She found out about thesimm.org through her son who went on a field trip to the SIMM.

How they will use the site: Once Marcia's son Joey showed her thesimm.org, she immediately saw a link connecting the exhibit Joey saw on video game music design to Super Mario Brothers 3, a game she played while in college. Once the floodgates to Nostalgia Land opened, Marcia spends free time watching gameplay videos and listening to music files of the games she played with her dad as a young girl.

Any additional site-specific demographics: Marcia has 3 children and plans on taking them to the SIMM to see the exhibits in person.

- **Nostalgia/Collector Persona**

Name: Sam Schneider
Occupation: Copywriter for Amazon.com
Gender: Male
Education: BA in Communications
Computing and Web experience: Very web savvy, especially Web 2.0. Has designer friends and colleagues and he definitely "speaks the language."
Personal Web behavior patterns: Uses a tablet for most of his non-writing computing needs. Keeps tabs on his world with heavy use of RSS feeds. Facebook lurker.
How they will use the site: Sam is a true game geek and spends a fair amount of time browsing numerous game sites. He has recently adopted thesimm.org as his go-to site for information on classic video games. He surfs here because the associative interface keeps attracting his attention to items he has not seen in years. These include gameplay videos, screenshots, and hi-resolution promotional art he loved when he was younger.
Any additional site-specific demographics: Gamer with wife and kids, but lots of expendable income. He has a man den that includes current generation systems and a few classic ones too.

- **Academic/Scholar Persona**

Name: Dr. Clancy P. Russell

Occupation: Economics professor

Gender: Male

Education: PhD in Social Economics

Computing and Web experience: Is seasoned when it comes to internet research in library databases. Has good luck with Google. Does not create much on the web; he mostly has his teaching assistants create his web content.

Personal Web behavior patterns: Dr. Russell is very old-school when it comes to using the internet. He is skilled at searching library databases but needs assistance from library staff to hone in on and find some articles. He keeps two email accounts: one on Yahoo! and the other at the university. The only web news he sees in on the Yahoo! front page.

How they will use the site: Dr. Russell is teaching a class in social economics. One of his modules will focus on the phenomenon of microtransactions on the Xbox Live Marketplace and the Playstation Network. He uses thesimm.org in order to find links to the most current research and scholarship regarding both the games chosen for the study, and microtransactions in general. He uses the creator information to contact the producers of the games in hopes of setting up a webinar for his students.

Any additional site-specific demographics: Dr. Russell does not do a lot of research on video games outside of library databases, so he chose thesimm.org because he is confident the information is both accurate and presented clearly.

- **Game Developer/Designer Persona**

Name: Debra Gurvitz

Occupation: Game Designer

Gender: Female

Education: BS Computer Science

Computing and Web experience: Expert

Personal Web behavior patterns: Surfer, Publisher, and Critic

How they will use the site: Debra is constantly looking for new ideas for games based on themes, mood, and characters of older games. She is also always interested in trivia related to games so she can try to design with that in mind.

Any additional site-specific demographics: Debra knows the thesimm.org is a robust database where she can find arcane bits and pieces to put into her designs. She also knows she can look at a wide range of games displayed by different criteria. This allows her to make accurate references to past games in the new ones she creates.

- **Curator/Librarian Persona**

Name: Nancy Henderson

Occupation: Academic Librarian

Gender: Female

Education: MS Library and Information Science

Computing and Web experience: Nancy is quite tech-savvy and always up-to-date on new IT devices and applications.

Personal Web behavior patterns: As a librarian, she is extremely skilled at searching library databases and the Web for any type of information. She also spends a lot of time on the Web, and especially on social media websites.

How they will use the site: At Nancy's library, they have decided to build a video game collection for students as well as scholars who are interested in game research. She uses thesimm.org in order to learn more about how games are organized, and get inspired on how she can organize the games in their new collection as well.

Any additional site-specific demographics: She is not a gamer herself, so she has to look up information on video games on various website in order to help plan the game related library events and programs.

II. Games Used for Testing

Table 3. Games used for testing the metadata schema