

User-centered Approach in Creating a Metadata Schema for Video Games and Interactive Media

Jin Ha Lee, Hyerim Cho, Violet Fox
University of Washington
Mary Gates Hall, Suite 370
Seattle, WA 98195
+1 206.685.0153
{jinhalee, chohr113, vfox}@uw.edu

Andrew Perti
Seattle Interactive Media Museum
305 Harrison St
Seattle, WA 98109
+1 518.653.5864
andrew.perti@thesimm.org

ABSTRACT

Video games and interactive media are increasingly becoming important part of our culture and everyday life, and subsequently, of archival and digital library collections. However, existing organizational systems often use vague or inconsistent terms to describe video games or attempt to use schemas designed for textual bibliographic resources. Our research aims to create a standardized metadata schema and encoding scheme that provides an intelligent and comprehensive way to represent video games. We conducted interviews with 24 gamers, focusing on their video game-related information needs and seeking behaviors. We also performed a domain analysis of current organizational systems used in catalog records and popular game websites, evaluating metadata elements used to describe games. With these results in mind, we created a list of elements which form a metadata schema for describing video games, with both a core set of 16 elements and an extended set of 46 elements providing more flexibility in expressing the nature of a game.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries – Standards, K.8.0 [Personal Computing]: Games

General Terms

Design, Standardization, Theory

Keywords

Video Games, Interactive Media, Cultural Artifacts, Metadata, Description, Seattle Interactive Media Museum

1. INTRODUCTION

Video games and interactive media have become an important artifact in our culture and everyday life in many aspects including entertainment, business, research, and so on. As video games, smartphone/tablet game apps, and other novel forms of interactive media become more popular with all types of users and communities, providing intelligent access also becomes increasingly important. Yet we still lack an organizational system that can effectively and thoroughly describe these video games in a standardized way. Current systems and standards fall short,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
JCDL '13, July 22–26, 2013, Indianapolis, Indiana, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2077-1/13/07...\$15.00.

whether originating from traditional library settings (e.g., FRBR [Functional Requirements for Bibliographic Records], RDA [Resource Description and Access], or LCSH [Library of Congress Subject Headings]) or from commercial settings such as game-related websites. They tend to focus only on providing basic descriptive information about games such as title, platform, publisher, etc.; their metadata elements and indexing terms are often not defined or explained at all [14], [22] (e.g., what does “genre” mean and what accounts for “strategy” games?) Moreover, it is often unclear where these organizational systems are sourcing their metadata from (i.e., is it the packaging such as the box or the manual? Or is it from playing the game itself or another secondary source?). In addition, it is not difficult to find examples of game representations that have misinformation or contradicting information as much of the metadata is unverified.

The objective of our research is to create a standardized metadata schema and encoding scheme (i.e., controlled vocabulary) that can be used to describe video games and interactive media across multiple organizations in different domains. This project started in 2011 as a joint effort between the University of Washington Information School and the Seattle Interactive Media Museum (SIMM). The objectives of the SIMM include aggregation, research, exhibition, and preservation of interactive media culture and the physical, digital, and abstract artifacts therein; thus robust and media-specific metadata is needed in order to serve a wide variety of use cases. Creating a metadata schema for the video game domain is a challenging task because it is almost impossible for anyone to have a thoroughly comprehensive understanding of games across multiple platforms, genres, and time periods. In order to ensure that we are reflecting as much domain expertise and knowledge as possible, we decided to conduct this research project in the form of a graduate level course. A total of 21 people participated in Phase I of the project and 18 in Phase II of the project. In Phase I, we created the basic backbone of the schema, CORE16, which lists 16 core elements that should be included in any organizational systems describing video games and interactive media [15]. Phase II of our project focused on expanding this core list by developing a larger “recommended” set of metadata elements. In this paper, we report on our progress made in Phase II and introduce a list of recommended metadata elements in our schema. We will also describe how the schema was designed, followed by discussing some of the challenges we encountered throughout the process.

2. RELATED WORK

In the Library and Information Science (LIS) field, the description and organization of video games traditionally has not been given much consideration. Although interest in video games has grown

considerably in commercial, cultural, and academic domains, this deficiency still hampers the development of effective video game organization systems. Current library and archival collections focus mostly on books and similar text-based documents, often treating artifacts like video games as products of popular culture and therefore of less scholarly value [15].

In the field of knowledge organization, the problems of describing non-book artifacts have been frequently discussed. Lubetzky [17] was the first to seriously discuss integrating non-book items into the library catalog, offering the concept of the work [16]. Lubetzky recognized that a book is an edition of a particular work, which is closely related to other editions and translations of the work. The problems of applying book-based features to non-book items were noted by Hagler [6] and Leigh [16]. The lack of principles for describing non-book items resulted in description based on their physical form, rather than intellectual content [16]. We identified evidence of this issue for video games in LC subject headings used for WorldCat records (e.g., computer games, electronic games, mobile games, application software).

The limited applicability of traditional content description standards on video games has also been noted by McDonough et al. [18]. The authors attempted to apply the Functional Requirements for Bibliographic Records (FRBR) model [9] to classic computer games, but could not easily determine work, expression, manifestation, or item of video games. Moreover, the FRBR model is missing some potentially useful characteristics of multimedia objects including music or games; for example, mood or similarity to other objects which can be beneficial for users' discovery of new music or games [13].

As Winget [22] pointed out describing video games for preservation is a truly demanding task, as is illustrated by Massively Multiplayer Online Role-Playing Games (MMORPG). They are not only digital, but also consist of software systems that have significant hardware, peripheral, and network dependencies which are difficult to collect and formally represent. The more we understand the characteristics of video games, the harder it becomes to apply current organization standards and rules to describe them. In order to overcome these problems, several scholars have started exploring new ways to organize and preserve video game information (e.g., [18], [22]). However, these attempts tend to emphasize preservation rather than description of games and employ data-centric methods rather than considering the end user's point-of-view.

Currently, the only systematically designed descriptive framework for games can be found in Huth's master's thesis work [8] which was reviewed extensively in Anderson et al. [1]. Huth designed his framework by using existing elements from OCLC Metadata Elements and Dublin Core Metadata Set and adding some new elements [22]. However, in this study he focused on describing games for older platforms such as Commodore 64 and the Atari 2600 systems. With a limited focus on very early game systems, Huth's work does not take into account the current game environment, which includes experiences such as online real-time games involving multiple users, game mobile apps, etc.

In order to create a descriptive framework that can effectively describe games across multiple platforms and time periods, it is crucial to improve our general understanding of the video games themselves as well as users' information needs and behaviors related to video games. Our study aims, through a systematic domain analysis and user studies, to contribute to developing a

theorized understanding of how people seek information related to games. This will inform our design of the metadata schema as it will help us select metadata elements that reflect the information needs of real users.

3. STUDY DESIGN

3.1 Methods

Metadata development processes are often relatively informal and draw upon professional experience rather than a more deliberate process [20]. However, as we developed our schema we strove for clarity, both in the process of creating the schema as well as presenting our result. Our methodological approach in designing the metadata schema can be described as a combination of domain analysis [7], in-depth interviews, and a user-centered design incorporating personas [3]. In approaching our metadata design from a user-centered perspective, we sought to address what we considered the missing component of many of the currently available organizational systems as pointed out in previous literature [23]. Figure 1 summarizes our whole design process (Phase I and II), following an iterative process.

In Phase I, we started by conducting an extensive domain analysis focusing on analyzing empirical data on how games are currently being described in the video game community as described in Hjørland [7]. Hjørland identified ways domains can be studied and understood in order to create metadata. We used the existing sources of information organization in order to understand how the domain was shaped, what was listed, and where there were gaps. We ended up collecting 61 existing metadata elements from a wide variety of video game-related websites as well as catalog records (Mobygames, Giantbomb, Allgame, Gamefaqs, Gamespot, IGN, Wikipedia, WorldCat, Amazon, to name a few). Afterwards, we developed 6 different personas epitomizing the most common types of people who would be interested in games: 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). We found the use of personas (archetypes representing the needs, behaviors, and goals of a particular group of users [3]) useful for our goal-directed design of this system. Based on these personas and several use scenarios, we evaluated 61 elements and identified the core elements that were deemed to be most useful for the all the user groups. These elements were evaluated and further revised based on our cataloging activity of sample games. More detailed information on our research activity in Phase I can be found in Lee et al. [15].

For Phase II, we started by exploring a larger number of game-related websites to collect additional terms to add to our inherited master list of 61 elements from Phase I. At the same time, we also conducted a facet analysis [19] of video game genre terms in order to tease out the multiple dimensions represented in these terms and included some of these dimensions as new metadata elements (e.g., theme, style, setting). Duplicates were removed and some similar elements were merged (e.g., platform, console, and system; achievements, awards, and trophies). As a result, we ended up with a total of 135 metadata elements.

In order to understand which information is valued and sought by real users, we conducted in-depth interviews of 24 gamers. The interview data helped us improve our understanding of what video

games mean to users, as well as how they obtain and organize their games, how they search and browse video games, and what kinds of information are useful for them for these tasks, etc. [Further discussion in Section 4.1].

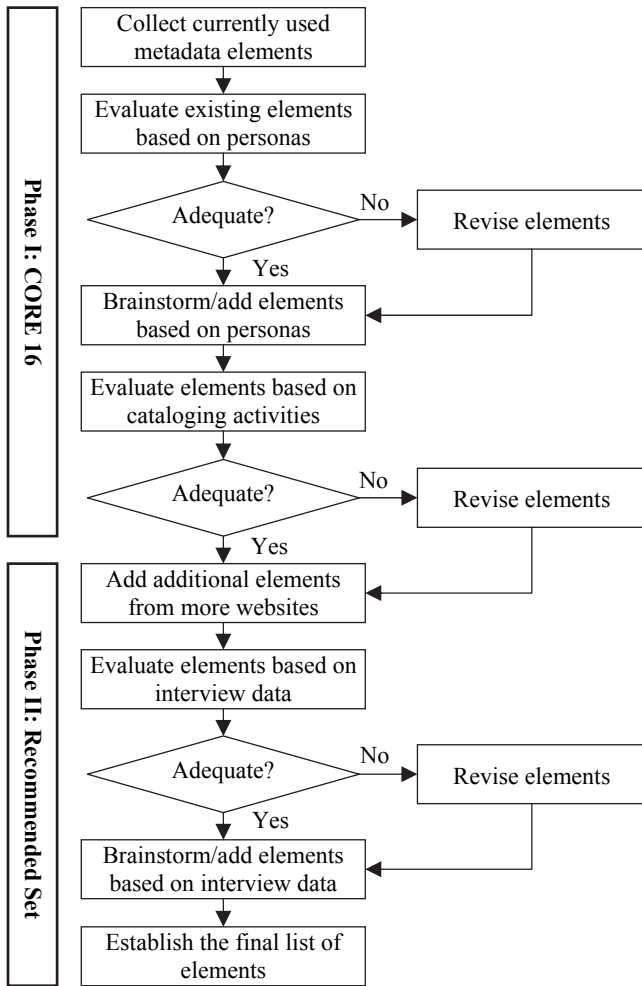


Figure 1. Metadata Design Process

Afterwards, each of the 135 metadata elements was evaluated based on our understanding of user needs as reflected in the interview data. We particularly focused on the elements that were mentioned as useful information by multiple interviewees. For the elements that were not specifically mentioned by interviewees, we tried to evaluate the value of the particular information from the users’ perspectives based on what they told us about their needs and behaviors related to video games. We also considered the following factors to determine if we should include the element in the recommended set:

- Source of information: Can the metadata be obtained from the physical manifestation of the game (e.g., box, manual) or the game itself? Or does it need to be obtained from a secondary source (e.g., game websites)?
- Ease of acquisition: How easy it is to obtain this particular metadata? What is a reasonable estimation for the amount of time and effort necessary for capturing the data?
- Cost-benefit: Is it worth describing this data considering the cost needed and the value added?

- Objectivity: Is it possible to obtain relatively objective metadata or do we need to rely on subjective judgments or opinions?
- Stability: Is the information for the particular metadata element relatively stable over time? Or will it keep changing?

As a result, we ended up with a total of 46 metadata elements (including the CORE16) as a recommended set.

3.2 Limitation

The authors acknowledge that constructing a standard such as a metadata schema is not a trivial, nor a one-time task. It should be noted that our goal is not to attempt to construct a “complete” schema. Rather we seek to construct a comprehensive schema that reflects the needs of real users. We hope that this will serve as starting point for a discussion in the game research community on how to build a useful organizational system for these artifacts for multiple stakeholders involved, and a basic framework that the community can collaboratively build upon. We plan to continue working on this project through much iteration, maintaining a partnership with the Seattle Interactive Media Museum, in order to further revise and refine the schema. We also plan to conduct a more systematic evaluation of the schema by applying it to create metadata records of sample video game collections provided by SIMM, identifying potential problems and issues through that process.

4. DATA AND DISCUSSION

4.1 User Interviews

In order to obtain more information about users who play video games and seek games for their own use or for other people, we conducted in-depth interviews of 24 gamers, ranging from casual to hardcore. Some of these gamers also identified themselves as collectors, parents, and educators. Most of them were male (21 out of 24). 11 participants were in their 20s, another 11 in their 30s, and 2 in their 40s. In our interview, we asked a variety of questions with regards to video games including their experience with video games, their game collections, game organization, game-related information needs and seeking behaviors, and the meaning of games to them. Due to limited space, in this paper we focus on reporting our findings on their game-related information needs and seeking behaviors. Several themes emerged as user requirements for video game metadata and organizational systems from analyzing the interview data. In the following text, we use codes to refer to participants (e.g., P1, P2) in order to protect their privacy.

4.1.1 Need for Subject Access Points

A number of users wanted information about the content of the game that could help them make purchase decisions. Metadata such as genre was one of the primary access points that were being used for search and discovery (P4, P5, P11, P16). However, several participants also mentioned the need for more detailed description or categorization of genre information. This is because genre itself “*doesn’t say a whole lot*” (P17) and “*genre alone can be somewhat misleading about how the game is actually designed to be played*” (P3) referring to the vastly different styles of games under the same genre. Users also recognize that genre is a significantly overloaded term and the multiple dimensions of genre need to be fleshed out. One participant (P17) said that there are two categories of genres: “structural” and “thematic.” Structural categories are the common categories used by video

game websites, and thematic categories cover plot and emotional content. Another participant (P6) used different terms to convey a similar idea: “elemental genre” for terms like “science fiction,” “horror,” “fantasy” that describe the details of the game universe, and “gameplay genres” such as RPG or FPS (first-person shooter). Other participants (P8, P13) used the term “setting” to refer to the same concept of the “elemental genre”. The fact that users referred to the same idea using multiple terms illustrates the importance of establishing a standard way of describing game information. For gamers who preferred games that have a strong story component (such as RPGs), the plot/narrative was essential information:

“It had great plot twists ... betrayal, love, excitement, it had side stories which were great, character development. It really felt like a very complicated story, even more so than a book or a movie.” (P3)

“Storyline is really I think the most important to me. Like it’s gotta have something that’s gonna draw me in.” (P11)

“That’s why I like the action-adventure, because there’s usually a story to be told, it’s not just “here’s a gun, go kill that guy” kind of thing.” (P22)

In addition, information such as the theme was deemed useful by participants who said they wanted to seek (P6) or avoid (P4) particular types of themes such as sex or violence. One user (P17) felt that the ESRB tags provided with the rating are useful but are not comprehensive enough to describe the themes.

In our metadata schema, we incorporated a number of subject access points to meet the needs mentioned by our users such as genre/gameplay, style, plot/narrative, theme, setting, and mood/affect. These elements, when taken together, should provide much more detailed information about the content of the game compared to simple genre labels.

4.1.2 Divided Reaction on User Ratings and Reviews

Some users (P4, P7, P10) highly valued ratings and review information from other users and also from “professional reviewers” (P7). Information such as scores provided by websites such as Metacritic or user reviews/ratings seemed to have heavy impact in making the purchase decision for those users.

“I read the review (to find out) if it’s worth to buy, worth to pay... since I would be a general user too, I would prefer to read general users’ reviews first. Professional, they have their own point of view to the game or the way of playing and that’s very biased. I believe the general user is right.” (P4)

Conversely, some users (P3, P5, P8) had negative opinions about this type of information, viewing it as too subjective, too crowded, and simply useless.

“I think allowing people to post videos all the time without restriction can get -- it just floods the market and overwhelms the senses. People creating like their own Top 10 lists and it just gets frustrating. Reposting the same videos and trying to compete with each other for certain amounts of views is also - it’s just a horrible idea for the consumer.” (P3)

“There are some games that get terrible ratings by other people that I absolutely love and some that get amazing ratings by other people that I’m not a big fan of.” (P5)

“As far as reviews go, I don’t put any stock in them.” (P8)

Participants like P22 specified that balanced reviews (i.e., not all positive or negative) discussing different components of the game were helpful. P19 said he only goes to two specific sources (IGN, GameTrailers) he trusts for reviews and nothing else. After much discussion, in our metadata schema, we decided not to include elements that solely rely on user-generated content such as user ratings or reviews. This is due to the high subjectivity of information which does not align well with organizations such as SIMM’s goal of providing highly authoritative and accurate information. Rather, we decided to provide a variety of different metadata that can help users more effectively determine the value of the game themselves such as box art, screenshots, trailers, and gameplay videos.

4.1.3 Visual Information on Games

Although most metadata standards tend to focus on providing textual description of the information objects, the interviews with our users revealed that visual information about the games was highly sought after. Artistic or visual style of the game seemed to heavily affect how users feel about games (P3, P5, P6, P13). It can determine how immersive the game is, and it was sometimes mentioned as the very reason for selecting certain games as their favorites.

“My favorite game is an old game for the Playstation 2, it’s called Shadow of Colossus, very good game, very innovative. It’s sort of -- it was really focused on artistic direction and wasn’t really focused on making a huge commercial success. I thought it was more true to the experience of gaming than a lot of games at the time.” (P3)

“I can forgive the frustrating game mechanics in order to keep playing because the aesthetics are good, the storyline is good, the goal is compelling.” (P6)

A number of users (P3, P4, P5, P8, P11, P16) also said they highly value the images of the box, screenshots showing the characters and the environment, trailers, and gameplay videos because they immediately help them determine if they will enjoy the game or not. P4 said “character is the most memorable thing for the video game’s appearance” and therefore he definitely wants pictures of the characters. P16 preferred games with Japanese RPG visual styles whereas P8 wanted games with realistic graphics and wanted to avoid “cartoon-y” graphics.

“Make it so it gives me enough detail on the box. If I see a game that looks good, I will occasionally go and look up a rating for it, see what other people are saying about it. More often I’ll try to look at the box and generally I like to see that there’s a lot of ways to customize my character.” (P5)

“Mainly watching game play video is the most important just because I’ve been playing games for so long, I have a very good idea about what I like and what I would enjoy. Watching a video is enough.” (P3)

In our schema, we decided to include an element visual style in order to describe the different styles of visual aesthetics which can hopefully be used in collocating similar games based on their appearance. Work by Järvinen [10] and Kyprianidis et al. [12] on artistic styles and techniques for images/videos provided the basis for creating a controlled vocabulary for this element.

4.1.4 Companies and Franchise

For video games, information on the companies involved in creating and distributing the games seemed to be highly valued by

users, maybe more so than other media such as books. We conjecture that this may have to do with the unique nature of games which generally involves a large group of people in the creation process, making it difficult to attribute them to a single artist or creator. Therefore users tend to remember the publishers or developers of games more readily than the individual directors, designers, or composers. P16, for example, was generally positive towards any games published by Square Enix and even had a separate folder to organize their games on his iPhone. P10 said knowing who developed the game is extremely important to gauge the quality of the game. P5 showed strong loyalty to particular franchises as this information helped him discover new games similar to those he previously enjoyed. These major game franchises serve as recognizable brands associated with particular playing experiences and properties of a game and can serve as useful information for users who are seeking familiar games [11].

We added the distributor element to the recommended set in addition to publisher and developer in CORE16 since this information is becoming increasingly important, especially for digitally downloadable games. We also worked on creating a controlled vocabulary of game companies to be used for these elements [further discussion in Section 4.3.6].

4.1.5 Multiple Stages of Information Needs

P16 described his multiple stages in searching information about games as follows:

- a) After announcement: When he receives advertisements about an upcoming game, he looks at the official description of the game as well as the images and videos to determine if the game interests him further.
- b) Before purchase: He sets up email reminders to keep track of the release date and region information.
- c) At the time of purchase: He looks at the screenshots and gameplay videos to make the final purchase decision.
- d) Playing: For RPGs specifically, he avoids reading game cheats, walkthroughs or articles describing the game because of potential spoilers.
- e) After the gameplay: If he is satisfied with the game, he will look for similar games by subscribing to email/RSS feeds.

P19 also mentioned that he tends to read reviews before purchasing a game and goes to wikis or walkthrough websites once he starts playing the game. Gamers' behavior of seeking different information as they proceed through the purchase path (e.g., publisher-released assets during pre-launch, advancement information during launch, and extension content during post-launch) was also noted in previous literature [4]. This implies that the organizational systems should provide options for hiding some of the detailed metadata elements (e.g., plot/narrative, theme, characters) in order to avoid revealing too much about the game content which may reducing the playability of the game.

4.1.6 Balance of Gameplay, Story, and Aesthetics

Multiple participants (P3, P6) mentioned that another important aspect is how a game balances the multiple components such as gameplay, story, and aesthetics. Depending on the user's taste, one of these components might be significantly more important than the others. For example, P8 said complicated storylines appeal to him and he searches for evidence of depth and developer's attention to the story. This also depends on the types of games as well. For example, puzzle or party games would typically focus on the gameplay whereas RPGs generally have a

strong story component. It may be useful to employ a feature where the organizational system can calculate and display the ratio of how well these different components were represented in each game based on their metadata records and use such information for generating recommendations of games.

4.1.7 Background Information on Game

Multiple participants (P10, P11) wanted to know more about the background or context of the game. P11 said she greatly enjoys hearing about and sharing "behind-the-scene" stories from and about the creative teams behind the game. In order to satisfy those needs, organization and preservation efforts on video games will need to move beyond the object in hand and explore different ways of capturing and preserving the context of how the game was created and used. Although this is outside of the scope of this paper, the importance of this issue must be noted, especially if we are concerned about the future generation truly understanding how certain games were accepted by contemporary gamers and why they were perceived to be so significant at the time of release.

4.1.8 Other User Needs and Requirements

Character customization options were especially important for MMORPG users (P5, P7, P13). Narrative types such as "linear," "branching," or "open" were also mentioned as important information by P13 who likes MMORPGs and strategy games. Some participants (P5, P6) mentioned that their budget affects their video game purchasing and sharing behaviors, thus they put a high value on price information. Point-of-view was also mentioned as useful information as participants like P6 who strongly prefers the first-person, highly immersive point-of-view in his games. Some users (P4, P9) talked about wanting to know how long it takes to complete a game. Interviewees who were also parents or teachers (P1, P2, P9) discussed learning objectives and difficulty of the games. Somewhat contrary to what we initially expected, all the parents or teachers we interviewed seemed less determined to always try to find games that have certain educational values and were more willing to have their children or students play games to have fun (P1, P2, P9, P23, P24).

(Answering the question asking about the importance of a game having educational values) "Not so important. I believe that most games, the more involved games, whether or not the educational value is inherent or obvious, I think there's a lot of learning going on." (P23)

"I think there's a value for just having fun. I think in today's world we have to kind of qualify everything. There is definitely some need for if it's being utilized in an educational sense, to have a goal. However, there's also importance placed on the fact that children just also need to play games." (P24)

Some participants also talked about information overload, implying the importance of an organizational structure that is well thought out and displayed:

"I think the downfall of a lot of websites is actually having too much content and not organizing the content well enough that they have." (P3)

"I don't want to get on a website and then have to feel like it's an 8-to-5 job trying to find information on one game. I want to be able to go in and within 2 or 3 clicks on the website, I want to be able to find what I'm looking for." (P8)

4.2 Metadata Elements

The list of metadata elements we selected for the recommended set is provided in Table 1. The CORE16 elements are noted with asterisks next to the name of the element. We strove to adopt definitions of these elements from existing standards (e.g., FRBR, CIDOC CRM, Getty Art & Architecture Thesaurus) whenever possible, in order to maintain some degree of interoperability. However, due to the unique nature of the artifact, most definitions

had to be newly created as it was impossible to identify reusable definitions from existing standards and vocabularies.

As part of our effort, we also created various encoding schemes (i.e., controlled vocabularies) for many of our elements. However, due to limited space, they will not be fully discussed in this paper. Another publication reporting the full encoding schemes is under preparation.

Table 1. Definitions of the Metadata Elements in the Recommended Set

Element	Definition
Title*	Proper names that are used to refer to a video game, assigned by the creator. (modified from CIDOC CRM, 2011, p.16)
Alternative title	Alternative titles of the video game, including the abbreviations.
Edition*	A word or phrase appearing in the manifestation of the game that normally indicates a difference in either content or form between the manifestation and a related manifestation previously issued by the same publisher/distributor (e.g., second edition, greatest hits), or simultaneously issued by either the same publisher/distributor or another publisher/distributor (e.g., collector's edition, limited edition). The edition designation pertains to all copies of the manifestation produced from substantially the same master and issued by the same publisher/distributor or group of publishers/distributors. (modified from FRBR, 2009, p. 41)
Format*	The distribution medium or method that provides the executable code of a video game.
Series *	The names referring to the creative universe from which the manifestation of the game derives.
Franchise/Universe	Proper names to refer to a cluster of video games that is characterized by the shared use of characters, or other component of games such as a similar gameplay or battle system.
Platform*	The hardware and firmware required to realize the game.
Developer*	An organization or group of individuals and/or organizations responsible for creation and/or realization of a game (i.e., writing the code of the game) (modified from FRBR, 2009, p.25)
Publisher*	An organization or group of individuals and/or organizations responsible for the manufacture, marketing, and distribution of a particular product manifestation.
Distributor	An organization or group of individuals and/or organizations responsible for distributing a game that has exclusive rights to market (within certain regions).
Special hardware*	A hardware that is required or recommended for playing the game in addition to the main platform/console.
Online capabilities*	The capabilities for playing the game online and/or downloading the game or additional features online.
System requirements	Software (or OS/firmware) requirements to realize the game in a particular platform.
Game credits	The intellectual contributors of the game and their roles as specified in the manual/end-game credits.
Official website	A URL and an archived copy or representation of the webpage as available from one of the officially associated companies on a specific date.
Price/MSRP	The suggested manufacturer's retail price (MSRP) at time of initial release in the region where the game was released.
Retail release date*	The date of the public/commercial release of the manifestation. (modified from FRBR, 2009, p.42)
Controls	The main default control buttons/keys and the corresponding main game actions.
Packaging	All items included in the original packaging of the game.
Number of players*	The number or range of the number of players the game can accommodate either separately or concurrently. Examples: 1, 1-2, 1-8, 1-Many (for massively multiplayer online games)
Rating*	The ratings designed to provide information about the content in video games for consumers' informed purchase decision by organizations such as the Entertainment Software Rating Board (ESRB). (modified from ESRB Game Ratings & Descriptor Guide) Examples: MA-13 Parental Discretion Advised. Mature Audiences; Everyone. E. (ESRB). Note that some older games do not have this rating information.
Purpose	The reason for why the game exists as intended by the game designer(s)/developer(s).
Customization options	The settings which can be modified by the player for personalized experience.
Difficulty levels	The options specifying the levels of challenge as defined internally in the game options (god mode, expert, etc.) and whether or not a user can freely change the difficulty setting during the game.
Achievements/Awards/Trophies	Title, description, requirements, and emblems of all unlockable in-game notifications.
Region*	The names used to refer to a place, region or territory where a game is designated as playable.
Language*	The language in which the work is expressed. (from FRBR, 2009, p. 36)
Identifier*	An alphanumeric code uniquely assigned to a manifestation of the game. When available, use the Universal Product Code (UPC) which is a 12 digit number representing a barcode. For some games (e.g., downloadable game app), this information may not be available. Note: this element used to be UPC in an earlier version of CORE16.

Box art/cover	The image featured on the front of the box/packaging or the officially released image which virtually represents the downloadable games.
Screenshots	Still images taken from the video output during the main gameplay.
Trailers	Video footage released and/or endorsed by the developer/publisher of the game for promotional purposes.
Gameplay videos	Video footage of actual, in-game play (excluding introductions, cutscenes, or trailers).
Genre*/Gameplay	A category of games determined by the overall nature of the experience based on a pattern of interactions and game rules. Note: definition revised from an earlier version of CORE16.
Style	A particular distinctive characteristic, mode of action, or manner of a gameplay.
Plot/Narrative	A story or sequence of events, the central conflict, or other plot elements of the game. (modified from Getty Art & Architecture Thesaurus)
Theme	An underlying intent or meaning in the game; also a common thread, motif, subject, or idea that recur in the game. (modified from Getty Art & Architecture Thesaurus)
Setting	The surroundings or environment (spatial and temporal) in which the game takes place.
Mood/Affect	The pervading atmosphere or tone of the video game which evokes or recalls a certain emotion or state of mind.
Temporal aspect	The methods by which time passes in the game and/or manner in which events take place.
Presentation	The manner or style of game display.
Point of view	Perspective from which the player views the gameplay.
Character names	The given full name (if available) and role of the main game characters.
Character types	Classification of game character types based on physical or metaphysical embodiment.
Link to historical events	Event(s) from history that are the setting for (or play a significant role in) the game. This may be a named event, a historical period, or a time and geographic location. (e.g., U.S. Civil War or Ancient Rome).
Type of ending	The method by which the player is led to gameplay culmination.
Visual style	A cohesive and unifying visual aesthetic.

4.3 Problems and Challenges

4.3.1 Source of Information

One of the main challenges we encountered while developing this metadata schema was deciding which source should be used to obtain the game information. Many descriptive elements can be filled by observing and transcribing information from the physical object itself (e.g., title, edition, platform, number of players). However, a large number of elements require additional research on the game using secondary sources or by playing the game itself (e.g., price, release date, achievements/awards/trophies, theme, setting). This suggests that clearly specifying the chief source of information for each element is extremely important.

Moreover, the general trend of moving towards digital downloadable format of games aggravates this problem as many of the recent games are now published without any packages or manuals. The expectation is that the manuals are now embedded in the game itself or accessible from an associated company's website. While one can actually play the game or the website is still alive, it remains possible to obtain this information. However, in a long term, there is no guarantee that all the games today will be playable. New versions of video game consoles are rapidly being developed and released, often without backward compatibilities. We may come to a point in the future where we have a collection of old games that are simply not playable on any devices that are available at that time, which will mean the source for obtaining the most authentic metadata is also subsequently lost.

4.3.2 Level of Description

The recommended metadata elements presented in this paper serve to provide a set of descriptions which satisfy the requirements for a majority of users: consumers, academics, gamers, parents, and others. Based on the FRBR model of group 1 entities, most of the user requirements we identified fell within the manifestation level of description. This is not to say that item

level of description (e.g., condition, provenance, location) will not be necessary, especially for organizations such as SIMM. In any kinds of organizations that deal with a video game collection, various administrative metadata elements will be necessary in order to manage the items in the collection (e.g., archivists need to describe the condition of a singular copy of a game, reference librarians need to know the exact location of a game in order to assist in retrieval of the item). Research for ascertaining the administrative metadata requirements will be conducted and explored in Phase III of this study as we work on establishing a larger "full" set of metadata for video games.

4.3.3 Background Knowledge of Games

In our metadata schema, we added several metadata elements that focus on describing the content or aboutness of the game rather than descriptive metadata (e.g., genre/gameplay, style, plot/narrative, theme). Many of these elements required establishing controlled vocabularies in order to ensure that catalogers can describe these elements in a consistent manner. However, the process of creating a controlled vocabulary for each subject metadata element not only takes a significant amount of time and effort, but also requires the creators to have extensive background knowledge of the video game domain. Consider "temporal aspect" as an example: without having played or researched numerous video games, it would be extremely difficult to understand and enumerate different ways that time passes and events take place in video games (e.g., real-time, turn-based, time manipulation, multiple game clocks, time travel, time action, calendar-based game clock).

4.3.4 Efficiency vs. Accuracy in Generating Metadata

Any controlled vocabularies we established for our metadata elements will inherently always be incomplete as there is a limitation on how many games can be considered as we establish these lists. Maintaining controlled vocabularies is particularly challenging in the video game domain because of its rapidly

changing nature. Therefore it could be very helpful to involve users of the organizational system or other game enthusiasts to help generate and maintain these vocabularies. However, organizations such as SIMM may be reluctant to adopt this method for generating their data since the key objective of their organization is to provide and preserve the most accurate and authoritative information about video games. This conflict between efficiency versus accuracy in generating the metadata must be somehow resolved in the organizational system. We are currently discussing several possibilities including setting up an approval process for accepting user-generated content or distinguishing verified vs. non-verified information in the metadata records.

4.3.5 *Subjectivity of Information*

There were some elements that we initially determined useful but ended up not including because of the difficulty of obtaining objective data. One prime example is the “estimated time of completion” for games mentioned by multiple users (P4, P9, P22). This information was valued by users like P4 and P22 because they were often playing games on a bus ride and did not want games that take a long time to complete, and P9 who wanted to find games for his students to play during the class. However, depending on how the game is played, the time for completion can widely vary. This is especially the case for more complex games with a number of side quests. We discussed options of providing a range of hours or asking users’ opinions and averaging the numbers. However, none of the options seem as a particularly satisfying solution as an individual’s way of how to play a game is a very personal and subjective matter. User rating/review information is another example of metadata element that was excluded due to its inherently subjective nature.

4.3.6 *Challenges for Particular Elements*

There were also several challenges we encountered as we were defining and establishing controlled vocabularies for particular metadata elements.

- *Character Names*

Providing information about character names was judged to be a valuable way to collocate similar games, but quickly brought up questions about how we should define “character.” What constitutes being designated a “main character” in a game? Does he/she/it have to be playable? Or can a game have a non-playable character that has a significant role in the narrative (such as the major villains/bosses)? Other games have very large character casts, such as the Pokémon series or games based on recruiting different party members (e.g., *Chrono Cross* with 45 possible members, *Radiata Stories* with 176 possible members). All these examples confirm that it is critical to provide very specific instruction on how to describe each of these elements in order to maintain the consistency in the granularity of data across all records.

- *Developer/Publisher/Distributor*

Compared to subject metadata elements such as genre or theme, developing a controlled vocabulary for video game companies used for developer, publisher, and distributor elements may be perceived as a tedious but relatively straightforward task. However, a closer investigation revealed the complexities of company evolution, the unpredictability of shifting markets, and

the endless tangle of relationships involving numerous mergers, bankruptcies, and acquisitions. Sorting through all the different possible relationships is difficult: abbreviations and variations of names for equivalence relationship; acquisitions, subsidiaries, and regional branches for hierarchical relationship; and partial ownership, poly-hierarchical, and mergers are only a few examples of the many possible relationships that can exist among these companies. In addition, obtaining reliable information about the history of these companies is extremely challenging, especially for companies that are older or not US-based.

Even when a game has accompanying materials such as a box and/or manual, identifying this information is not a trivial issue. Many games simply show a logo representing a particular company on the packaging without specifying the specific role of the company [15]. Since one company can take multiple roles, it is often unclear how to accurately determine this information even with researching secondary source materials.

- *Plot/Narrative*

None of the game-related websites or organizational systems we reviewed had a standardized element encapsulating plot, even though a number of users we interviewed mentioned the importance of story or plot for selecting and enjoying a particular game and seeking similar games. Our exploratory attempt to create such an element uses a subject-verb-object structure to describe the most prominent storyline(s) in the game.

We began our work in this area by looking at different approaches in categorizing basic plots in literature including Foster-Harris (1959), Booker (2004), and Uther (2004). Compared to the plots of literature, the plots in video games seem less varied. This is perhaps due to the fact that games generally have to incorporate certain story elements, such as a major conflict between some entities or an extreme situation (e.g., the end of the world), in order to justify critical game components such as a battle system and character development. Our preliminary attempts at assigning plot confirmed the challenges of finding the most effective level of granularity in describing the plot. Describing a game’s plot as the story of a hero saving the world would apply to an overly large number of games, rendering it useless; on the other hand, an extremely specific plot description such as “a tycoon builds roller coasters” (as in the *RollerCoaster Tycoon* series) would not allow for the ability to find similar games.

Separating the concept of gameplay from plot sometimes proved problematic. Certain features clearly describe gameplay without shedding light on the story (e.g., casting spells), but others are more ambiguous. For example, *Myst* is known for its clever puzzles which enable the player to explore the game’s environment, but the plotline involves discovering the secrets behind mysterious books and their owners. In playing the game, the plot seems incidental to the compelling atmosphere and challenge of solving puzzles. In game such as *L.A. Noire*, the plot is complex and integral to the game’s appeal. At the other end of the spectrum, some games have no plot (e.g., *Tetris*) or have a more or less irrelevant plot (e.g., *Mortal Kombat*). Another important consideration involves giving away too much of the plot; users have varying degrees of sensitivity to spoilers.

4.4 *Application of the Schema*

In order to demonstrate how this schema can be applied to capture the metadata of video games, the authors attempted to describe

three sample games from different eras using our proposed schema [Table 2]. Some of the metadata elements (screenshots, trailers, gameplay videos) are not listed as it is impossible to embed the metadata in the paper, the content simply takes up too

much space (controls, box art, game credits), or it is not applicable to all three examples (distributor, system requirements, achievements/awards/trophies, links to historical events).

Table 2. Application of the Metadata Schema for Sample Games

Metadata Element	Sample game 1	Sample game 2	Sample game 3
Title*	Super Mario Bros.	The Legend of Zelda: Ocarina of Time	Shin Megami Tensei: Persona 3 Portable
Alternative title	SMB; Super Mario Brothers; Mario	TLOZOOT; OOT	P3P; SMT P3P
Edition*	Retail	Retail	Persona 3 Portable; Enhanced remake of Shin Megami Tensei: Persona 3
Format*	Cartridge	Cartridge	UMD
Series *	Super Mario Bros.	The Legend of Zelda	Persona
Franchise/Universe	Mario	Zelda	Shin Megami Tensei
Platform*	Nintendo Entertainment System	Nintendo 64	PlayStation Portable
Developer*	Nintendo R&D4	Nintendo EAD	ATLUS
Publisher*	Nintendo of America, Inc.	Nintendo of America, Inc.	ATLUS USA
Official website	-	http://www.zelda64.com/	http://www.atlus.com/p3p/
Price/MSRP	Unknown	\$69.95	\$39.99
Retail release date*	1985	November 21, 1998	July 6, 2010
Packaging	Shrink Wrap, Box, Cartridge, Manual, Poster, Foam Block	Shrink Wrap, Box, Cartridge, Manual, Nintendo Power Catalog, Safety Information Guide	Shrink Wrap, Box, UMD, Manual
Number of players*	2 Alternating	1	1
Rating*	N/A	E	Mature 17+; Blood, Partial Nudity; Sexual Themes; Violence (ESRB)
Purpose	Entertainment	Entertainment	Entertainment
Customization options	N/A	N/A	Choice of male or female protagonist
Difficulty levels	Normal/Difficult; Difficult mode available after game ends	-	Beginner/Easy/Normal/Hard/Maniac; Cannot change once the game starts
Region*	North America	North America	North America
Language*	English	English	English
Identifier*	045496630140	045496870041	4984995900469
Genre*	Action	RPG	RPG
Style	Platformer	Action RPG	JRPG
Plot/Narrative	Protagonist-explores-world. Protagonist-defeats-bosses. Protagonist-rescues-princess.	Protagonist-explores-world. Protagonist-fights-evil. Protagonist-rescues-princess.	Protagonist-develop-friendship. Characters-confront-self. Characters-prevent-apocalypse.
Theme	Fantasy – Princess	Fantasy – Princess, Supernatural – Fairies	Supernatural – Demons
Setting	Spatial – Virtual Worlds, Nature	Spatial – Nature, Deserts, Tundra, Ocean	Spatial – Schools, Asian Temporal – Modern
Mood/Affect	Quirky	Humor, Dark, Mysterious, Adventurous	Dark
Character names	Mario (protagonist); Luigi (second player); Princess Toadstool; Bowser (main antagonist)	Link (protagonist); Zelda; Ganon (main antagonist)	Minato Arisato (protagonist); Female protagonist (name unknown); Junpei Iori; Ken Amada; Akihiko Sanada; Shinjiro Aragaki; Yukari Takeba; Mitsuru Kirijo; Fuuka Yamagishi; Aigis; Koromaru; Metis; Nyx (main antagonist)
Character types	Natural-Human; Supernatural-Other monster	Supernatural-Elf; Supernatural-Other monster	Natural-Human; Supernatural-Demon
Type of ending	Finite	Finite	Circuitous, Branching
Temporal aspect	Real-time	Real-time, Time manipulation, Time travel	Turn-based, Calendar-based
Presentation	2D, Side-scrolling	3D	2D, 3D
Point of view	Third person	Third person	Third person
Artistic style	Retro	Anime/Manga	Anime/Manga

5. CONCLUSION AND FUTURE WORK

In this paper, we provided an overview of our current schema for describing video games developed from a user-centered approach. As we developed a metadata schema and encoding schemes for video games and interactive media, a number of challenges and potential problems emerged. Although we do not currently have solutions for many of these challenges, we hope that our findings will serve as a catalyst for discussion in the digital libraries community. Many of the issues discussed here are also applicable to other artifacts from popular culture such as apps, video clips, comics, webtoons, etc. Without a standardized way for organizing and describing these artifacts, it will be impossible to search and discover that app you want, or the video clip you once watched and want to revisit.

In our future work, we plan to focus on revising the structure of this metadata schema: adopting attributes for particular elements (e.g., title, rating) and organizing certain elements (e.g., genre) in a hierarchical and faceted structure. We will also continue to work on detailed instructions for obtaining and transcribing metadata and providing examples to help catalogers understand how to apply the schema. Our work on establishing controlled vocabularies will be ongoing. In addition, we plan to conduct a large-scale user survey to test our current schema and achieve a deeper understanding of users' information needs to be reflected in later revisions of the schema. The authors hope that this work will augment existing standards in the LIS field (e.g., FRBR, CCO, LCSH) as well as assisting organizations such as SIMM by providing a more formal metadata schema and encoding schemes that can be used across academic and commercial domains.

6. ACKNOWLEDGMENTS

We thank Gregory Bem, Dr. Joseph Tennis, the SIMM, and all the participants of INFX 598 for their valuable contributions to the project. We also thank the University of Washington's Office of Research for their financial support for this research.

7. REFERENCES

- [1] Anderson, D., Delve, J., and Pinchbeck, D. 2010. Toward a workable emulation-based preservation strategy: Rationale and technical metadata. *New Review of Information Networking* 15, 2, 110-131.
- [2] Booker, C. 2004. *The seven basic plots: Why we tell stories*. Continuum, London.
- [3] Cooper, A. 1999. *The inmates are running the asylum*. Sams, Indianapolis.
- [4] Google. 2012. *Understanding the Modern Gamer: What Google search data says about gamer behavior*. Google Whitepaper.
- [5] Foster-Harris, W. 1959. *The basic patterns of plot*. University of Oklahoma Press, Norman.
- [6] Hagler, R. 1980. "Nonbook materials: Chapters 7-11." In *The Making of a Code: The Issues Underlying AACR2*. ALA, Chicago.
- [7] Hjørland, B. 2002. Domain analysis in information science. Eleven approaches - traditional as well as innovative. *J Doc*, 58, 4, 422-462.
- [8] Huth, K. 2004. *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.
- [9] IFLA Study Group on the FRBR, IFLA Section on Cataloging, Standing Committee. 2009. *Functional requirements for bibliographic records: Final report*. K.G. Saur Verlag, München.
- [10] Järvinen, A. 2002. Gran Stylissimo: The audiovisual elements and styles in computer and video games. In *Proceedings of Computer Games and Digital Cultures Conference*. Tampere University Press, Tampere, 113-128.
- [11] King, D., Delfabbro, P. and Griffiths, M. 2010. Video game structural characteristics: A new psychological taxonomy. *International Journal of Mental Health and Addiction* 8, 1, (January 2010), 90-106.
- [12] Kyprianidis, J., Collomosse, J., Wang, T., and Isenberg, T. 2012. State of the art: A taxonomy of artistic stylization techniques for images and video. *IEEE T Vis Comput GR*, 2012, 1-20.
- [13] Lee, J. H. 2010. Analysis of user needs and information features in natural language queries seeking music information. *J Am Soc Inf Sci Tec*, 61, 5, 1025-1045.
- [14] Lee, J. H., Tennis, J. T., and Clarke, R. I. 2012. Domain analysis for a video game metadata schema: Issues and challenges. *Theory and Practice of Digital Libraries, Lect Notes Comput SC*, 7489, 280-285.
- [15] Lee J. H., Tennis, J. T., Clarke, R. I., Carpenter, M. 2013. Developing a video game metadata schema for the Seattle Interactive Media Museum. *International Journal on Digital Libraries*, 13, 2, 105-117.
- [16] Leigh, A. 2002. Lucy is "Enceinte": The power of an action in defining a work. *Cataloging & Classification Quarterly*, 33, 3-4.
- [17] Lubetzky, S., Connell, T. H., and Maxwell, R. L. 2000. *The future of cataloging: Insights from the Lubetzky symposium*. American Library Association, Chicago.
- [18] 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.
- [19] Spiteri, L. 1998. A simplified model for facet analysis. *Can J Inform Lib Sci*, 23, 1-30.
- [20] Tonkin, E. 2007. A user-centered approach to metadata design. In *Proceedings of the 11th European Conference on Digital Libraries (ECDL)*, Budapest, Hungary.
- [21] Uther, H.-J. 2004. *The types of international folktales: A classification and bibliography based on the system of Antti Aarne and Stith Thompson*. Academia Scientiarum Fennica, Helsinki.
- [22] Winget, M. A. 2011. Videogame preservation and massively multiplayer online role-playing games: A review of the literature. *J Am Soc Inf Sci Tec*, 62, 10, 1869-1883.
- [23] 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.