

12-1-2015

Empirical Evaluation of Metadata for Video Games and Interactive Media

Rachel I. Clarke

Syracuse University, rclark01@syr.edu

Jin Ha Lee

University of Washington

Andrew Perti

Seattle Interactive Media Museum

Follow this and additional works at: <http://surface.syr.edu/istpub>

 Part of the [Cataloging and Metadata Commons](#), and the [Interdisciplinary Arts and Media Commons](#)

Recommended Citation

Clarke, Rachel I.; Lee, Jin Ha; and Perti, Andrew, "Empirical Evaluation of Metadata for Video Games and Interactive Media" (2015).
The School of Information Studies: Faculty Scholarship. Paper 172.
<http://surface.syr.edu/istpub/172>

This Article is brought to you for free and open access by the School of Information Studies (iSchool) at SURFACE. It has been accepted for inclusion in The School of Information Studies: Faculty Scholarship by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Empirical Evaluation of Metadata for Video Games and Interactive Media [preprint]

Jin Ha Lee (corresponding author)
Information School, University of Washington
Mary Gates Hall, Suite 370, Seattle WA 98195
Phone: 206.685.0153,
Email: jinhalee@uw.edu

Rachel Ivy Clarke
Information School, University of Washington
Mary Gates Hall, Suite 370, Seattle WA 98195
Phone: 206.685.0153,
Email: raclarke@uw.edu

Andrew Perti
Seattle Interactive Media Museum
305 Harrison St, Seattle WA 98109
Phone: 518.653.5864
Email: andrew.perti@thesimm.org

Abstract

Despite increasing interest in and acknowledgement of the significance of video games, current descriptive practices are not sufficiently robust to support searching, browsing, and other access behaviors from diverse user groups. In order to address this issue, the Game Metadata Research Group at the University of Washington Information School, in collaboration with the Seattle Interactive Media Museum, worked toward creating a standardized metadata schema. This metadata schema was empirically evaluated using multiple approaches—collaborative review, schema testing, semi-structured user interview, and a large-scale survey. Reviewing and testing the schema revealed issues and challenges in sourcing the metadata for particular elements, determining the level of granularity for data description, and describing digitally distributed games. The findings from user studies suggest that users value various subject and visual metadata, information about how games are related to each other, and data regarding game expansions/alterations such as additional content and networked features. The metadata schema was extensively revised based on the evaluation results, and we present the new element definitions from the revised schema in this paper. This work will serve as a platform and catalyst for advances in the design and use of video game metadata.

Introduction

Video games are of increasing importance to American society as objects of economic stimulus as well as cultural heritage. Due to this increased interest in games for consumer entertainment as well as historical, cultural, and scientific study, many cultural heritage institutions have established collections of video games and related media. The Library of Congress collects, preserves, and offers access to two main types of video games: educational games that support the Library's initiatives and controversial games collected to support legislation related to sex and violence in video games (Owens, 2012). In Great Britain, the National Videogame Archive (NVA)

focuses on collecting and preserving hardware, original software, design documents, and marketing materials (Newman, 2009). Recently, the British Library began collaborating with the NVA to archive video game related websites, including screenshots and walkthroughs of games (Crookes, 2012). This related information is valuable to scholars studying games in a social or historical context. Other organizations, such as the Strong National Museum of Play in Rochester, New York, do not collect video games exclusively, but recognize the importance of sharing historical video game information. Their recent exhibit “Atari By Design: From Concept to Creation” used materials on loan from the library and archives at the International Center for the History of Electronic Games, which collects printed materials related to the history of video games and the ways they affect how people learn, play, and connect. The exhibit placed design and development materials like sketches and storyboards alongside original arcade game cabinets to let museum visitors experience the entire process from ideation to gameplay.

These and many other organizations collect, preserve, and circulate video games and interactive media, yet each institution has a unique way of describing them. Public and academic libraries shoehorn video games into their local catalogs, using workarounds in metadata records to differentiate games from other media so patrons might find them more easily. In the Library of Congress, games are cataloged in the Library’s moving images database using metadata designed for motion pictures and sound recordings; however, this solution is still in transition, and raises questions about challenges researchers might face trying to find records for games in a database designed primarily for motion pictures that is only available on-site in the Library’s reading room. Additionally, current Library of Congress genre headings designed for literature and film are insufficient for video games as they do not adequately describe games in a recognizable way for players or researchers (Owens, 2012). Archives that focus on hardware will also have different metadata needs than collections focused on software or print materials. And all organizations, regardless of focus, face the challenges of collecting, preserving, and offering access to digitally distributed games—games that have no physical component but exist only in electronic bits accessed through download or streaming from “the cloud.” Libraries, archives and museums need a common, shared metadata standard that covers all aspects of video games – for players, researchers, and curators – that can describe all types of video games, from historic arcade games in cabinets to game apps downloaded on smartphones, in descriptive vocabulary terms that are relevant to games and understandable to users.

In order to address this issue, the GAMER (GAME METadata Research) Group at University of Washington Information School, in partnership with the Seattle Interactive Media Museum (SIMM), has been working on a research project to build a standardized metadata schema and controlled vocabularies for video games since 2011. The objective of our research is to create a robust, media-specific metadata schema intended to describe a variety of games, from historical to contemporary, and serve a variety of use cases to meet the needs of users of the SIMM or other similar cultural heritage organizations. In particular, we seek to answer the following research questions:

- 1) What kinds of information about video games must be provided to support users’ information seeking activities from their perspective?
- 2) Is the proposed metadata schema capable of describing a wide variety of games across different genres, formats, platforms, and time periods?

This paper reports our efforts to evaluate how well the schema represents and reflects the information needs, behaviors, and language of the user groups for whom it was designed, and whether the current version of proposed metadata schema is capable of describing all games within the video game domain.

Relevant Work

Organization and Preservation of Video Games

In library and information science, a few projects share similar objectives of improved organization and preservation of video games and interactive media. The Preserving Virtual Worlds project was a collaborative research project conducted by the Rochester Institute of Technology, Stanford University, the University of Maryland, and the University of Illinois at Urbana-Champaign as part of Preserving Creative America, an initiative of the National Digital Information Infrastructure and Preservation Program at the Library of Congress (McDonough et al. 2010b). This project focused on preserving older video games and software, and establishing best practices and strategies for game preservation. While this project laid preliminary groundwork for basic metadata standards, the final report specifically calls for future work in establishing relationships and entities, and states that the project barely scraped the surface for standardized ontologies in this domain (McDonough et al. 2010b). Currently the project is in its second phase, focusing on determining significant properties for educational games. Our project involves a wider range of games, from older games to recent digitally downloadable games and game apps on tablets and smartphones, as well as games for entertainment, and incorporates user behaviors and needs beyond preservation. The motivation for our project also stems from a very practical problem: how to best organize and provide access to SIMM's game collection. Thus, one of our future goals is to create a set of metadata records for a large and diverse game collection.

Stanford University Libraries and the National Institute of Standards and Technology, funded by the National Software Reference Laboratory, have also launched a project on the digital preservation of 15,000 software titles, including games, in the Stanford University Libraries¹. Their goal is to preserve these titles for use by academic scholars for research purposes. Our project, however, targets a wider range of user personas representing stakeholders beyond just researchers (Lee et al., 2013b). Twenty-four user interviews from earlier phases of this research investigated game information needs and search/browse behaviors of real users such as players, parents of player, collectors, developers, and curators. The results directly informed the development of our initial metadata schema and encoding schemes. Stanford University Library, in partnership with the University of Santa Cruz, also recently received funding from IMLS to develop a metadata scheme for digital games, as well as a system for citing in-game events and game states². The authors expect their findings will augment our previous and current research efforts.

Megan Winget led an IMLS-supported video game preservation project³ centered on studying the creation process and artifacts of the video game industry through observation of and interviews with game developers. The project specifically focused on supporting the collection and preservation of massively multiplayer online (MMO) games, whereas our work involves all genres of video games for a variety of platforms (i.e., console, handheld, mobile, PC). Rather than focusing on the creation process of a game, our work emphasizes the user experience of the game and focuses on describing the game information that are most relevant to users. This is evidenced by several metadata elements such as mood, visual style, plot, theme, etc. that describe the content or subject of the games.

Additionally, the International Center for the History of Electronic Games is currently testing the functionality of approximately 7,000 games in their collection and video recording old games, supported by IMLS⁴. The International Game Developers Association Game Preservation Special Interest Group has also advocated for and published on the importance of preserving video games since 2004 (Newman, 2009). We believe that our research effort will amplify the impact of these preservation efforts by greatly enhancing access to collections through robust metadata.

Evaluation of Metadata Schemas

Metadata evaluation is a longstanding topic in library and information science, yet perspectives vary on how to perform such evaluation. Most evaluation seeks to identify and improve metadata quality—yet few attempts to concretely define what “quality” entails. Without established conceptual and operational definitions, it is difficult, if not impossible, to evaluate metadata quality (Moen, Steward and McClure, 1997). In the absence of definitions of quality, scholars and organizations offer different principles that constitute quality metadata. Bruce & Hillman (2004) compiled seven criteria for metadata quality: completeness, accuracy, provenance, conformance to expectation, logical consistency and coherence, timeliness, and accessibility. NISO (2007) identifies “good” metadata as metadata that conforms to community standards; supports interoperability; uses authority control and other content standards; includes clear conditions for and terms of use; supports long-term curation and preservation; and possesses the qualities of authority, authenticity, archivability, persistence, and unique identification. After reviewing multiple studies, Park (2009) notes that completeness, accuracy, and consistency are the most commonly identified criteria for metadata quality. Others do not enumerate lists of criteria but rather summarize quality as “fitness for purpose,” (Guy, Powell, & Day, 2004) that is: does the metadata accomplish the purpose it was designed to achieve? For traditional bibliographic collections, the purpose of metadata is typically to allow users to find, identify, select and obtain materials of interest (IFLA, 2009), and therefore good quality metadata must accomplish these goals.

Yet methods for measuring metadata quality based on these criteria vary. Many studies focus on evaluation at the metadata creation stage, with a special focus on how to improve quality at the point of entry. Currier et al. (2004) found that inaccurate data entry and inconsistent subject vocabularies affected resource discovery. Simple metadata input tools, such as drop-down menus, can significantly improve consistency and therefore improve metadata quality (Greenberg et al., 2001). In order to evaluate metadata quality after creation, other methods may be used. One typical method is to assess quality through expert review. However, this raises the question of who qualifies as an expert. For many mature metadata standards, such as those in traditional bibliographic description, longstanding communities such as the Program for Cooperative Cataloging (PCC) and the Online Audiovisual Catalogers (OLAC) have come to consensus over time about definitions of quality (Hillman, 2008). Such communities also benefit from longevity as experts develop over time. Unfortunately, nascent metadata standards in early stages of development by their very nature do not allow for such resources. Additionally, regardless of length of participation or amount of experience, a reviewer’s expertise may still be considered subjective. Another typical method to evaluate metadata quality after creation is to check for consistency in application among multiple metadata creators. Commonly referred to as “inter-indexer consistency” due to its origins in indexing studies, this method quantitatively measures the degree of agreement between different indexers indexing the same document (Hughes & Rafferty, 2011). When used to evaluate metadata creation, the measurement may not be limited to consistency of subject indexing term application, but address the values for any element(s) in a schema. Historical studies find that agreement ranges vary considerably, possibly due to a lack of agreed upon definition of what constitutes an exact value match and also the use of different measurement calculations (Hughes & Rafferty, 2011). While consistency was put forth earlier as a criteria of metadata quality, and inter-indexer consistency has been looked upon as an indicator of quality (Funk et. al, 1983), consistency may also be detrimental (Bloomfield, 2001). After all, what good is consistently applied metadata if the actual value is incorrect?

It should be noted that these evaluation techniques—like most methods of metadata evaluation in general—were designed to evaluate descriptive metadata itself; that is, recorded descriptive values as opposed to elements of a schema or a schema as a whole. Even less is written

about how to evaluate a schema as a holistic entity. One technique to evaluate a schema is to compare it with other schemas, such as Beak and Olson’s (2011) comparative analysis of a standard bibliographic metadata schema with a metadata schema specifically designed for children. Such an analysis was able to highlight missing elements as well as elements directly supporting children’s information-seeking behavior, thus increasing access points for children in system implementation. Other methods evaluate the entirety of a metadata schema by implementing it in a functional system and testing it in situ. For example, Shukar et al. (2013) developed a federated metadata schema for e-government and sought user feedback on the schema through its deployment via a user interface portal. Although the study reports positive user responses, it is difficult to determine if the success was due to the schema itself or the design of the user interface through which it was presented. As no metadata schema is likely to be used without some sort of interface layer, the effect of the interface design must always be taken into consideration.

In this study, we evaluate our metadata schema by employing two methods: 1) schema testing by creating sample metadata records of a test collection of video games, and 2) user evaluation of the schema through multiple data collection methods. They are further discussed in the Study Design and Methods section.

Current Status of the Project

Since the inception of the video game schema development by GAMER and SIMM in 2011, we have attempted to incorporate the strengths of all of the previously described evaluations methods where applicable. In order to ensure that we are reflecting as much domain expertise as possible, we conducted part of this research project in the form of a graduate level course. Considering it is almost impossible for any single person to have a thoroughly comprehensive understanding of video games and interactive media across multiple platforms, genres, and time periods, it was essential that we involve as many video game experts and enthusiasts as possible in the research process. A total of 21 people including students, domain experts, and metadata professionals participated in Phase I of the project; 18 in Phase II; and 15 in Phase III of the project, which we report on in this paper. Table 1 outlines our previous and current research efforts for this project.

TABLE 1. Overview of the Previous and Current Research Efforts

Timeline and Goals	Methods and Activities	Outcomes
Phase I (2011-2012) Establish the core set of elements for a metadata schema	Domain analysis: Examined how video games are organized in current systems and collect video game metadata elements	16 CORE elements identified and defined.
	Personas: Evaluated the collected metadata elements from domain analysis based on six personas	
Phase II (2012-2013) 1) Establish the recommended set of elements for a metadata schema	Domain analysis: Continued to explore more game websites to collect additional elements and terms that can be used for controlled vocabularies	1) 46 REC (recommended) elements (including CORE16) identified and defined based on user data and facet analysis results. 2) New controlled
	User interview: Selected which metadata elements are important to users based on the in-depth interviews of 24 gamers	

2) Develop encoding schemes (controlled vocabularies) for select elements	Facet analysis: Conducted a facet analysis on select elements and define the facets and foci	vocabularies created for 17 elements.
Phase III (2013-2014) Review and evaluation of the current schema	Collaborative review: Collaboratively analyzed and revised the current metadata schema and controlled vocabularies	1) CORE and REC metadata sets thoroughly revised based on the additional user data and feedback from metadata record creators.
	Schema testing: Created metadata records of 65 sample games in order to test the applicability of the schema	
	User interview: Conducted 32 additional interviews focusing on personas other than gamers to gather feedback on the schema	2) Controlled vocabularies for eight elements also revised. New controlled vocabularies created for two additional elements.
	Survey: Conducted a large-scale survey collecting 1,257 responses to obtain more generalizable results	

In Phase I, we conducted an extensive domain analysis using empirical data currently describing video games (Lee, Tennis, & Clarke, 2012). We used existing information sources (i.e., a variety of video game-related websites and catalog records from sources like Mobygames, Giantbomb, Allgame, Gamefaqs, Gamespot, IGN, Wikipedia, WorldCat, Amazon, etc.) to understand how the domain has been shaped, how it is currently described, and where gaps appeared. We also developed six different personas—archetypes representing the needs, behaviors, and goals of a particular group of users (Cooper, 1999)—epitomizing the most common types of people interested in games. The six personas developed were game player, parent of youth game player, nostalgic collector, academic scholar, game developer/designer, and curator/librarian (Lee et al., 2013b). Based on these personas and several use scenarios, we evaluated the 61 elements and identified the 16 CORE elements deemed to be most useful for the all user personas. These elements were evaluated and further revised based on cataloging sample games. More detailed information on our research activity in Phase I can be found in Lee et al. (2013b).

In Phase II, we established the recommended set of elements for a metadata schema (Lee et al., 2013a) and began developing encoding schemes/controlled vocabularies for select elements (Donovan et al., 2013; Lee et al., 2014). In addition to the CORE 16, 30 additional elements were recommended to describe video games to the level of thoroughness useful to the various user groups. As we began to develop vocabularies for many of these elements, it became evident that metadata from extant sources significantly varies with regards to the types and granularity of the terms. For example, most websites did not provide definitions for genre labels, and those that did were not consistent with definitions from other sites. We also found information sources varied widely in reliability and availability. Dates and features varied, and descriptions of features were often colored by subjective marketing propaganda. Definitions and differences between “developer” vs. “publisher” were unclear. Differentiating among various editions of games was extremely challenging due to releases of the same game in multiple regions, for multiple consoles or systems, as well as special “collector’s editions” or other limited special releases. Finding information on previous versions of digitally distributed games such as game apps for smartphones or tablets is also exceedingly difficult (Lee, Clarke, & Perti, 2014).

In the current phase of the study, Phase III, we focused on revising and evaluating the current schema and controlled vocabularies. There were two concurrent research streams: 1) refining the schema as a group and testing its applicability using sample games, and 2) gathering

user feedback on the schema through interviews and a large-scale online survey. More detailed discussion on research activities from Phase III is provided in the following section.

Study Design and Methods

Refining and Testing the Schema

Collaborative review: The previous metadata work from Phases I and II was revised through an iterative and collaborative expert review process. Each student independently wrote up his or her feedback on the initial list of personas, the elements and structure of the schema, and the controlled vocabularies, guided by personal expertise and experience with video games. To the method of expert review, we added a level of collaborative consensus, where summary of the feedback was reviewed by the group, and every aspect was discussed until a convergence of views was achieved. As a result, new elements were included to make the schema more pertinent to digitally distributed games (Lee et al., 2014). All definitions were refined and instructions for identifying and extracting the metadata values were written for each element. A number of controlled vocabularies were revised and two new vocabularies were created.

Schema testing: The schema and encoding schemes (syntax and vocabulary) were further evaluated by applying them to examples of video games to create sample metadata records of those games. These records were created by the students participating in the course. The students had a wide range of experience in cataloging (from novices to professionals) as well as gaming experience (from casual to dedicated gamers). As a group, they cataloged a total of 65 sample games which consisted of traditional console games for a variety of platforms, digitally distributed game apps, and flash-based games. The full list of sample games cataloged is provided in the Appendix. Each game was cataloged twice according to the schema: once each by two different students. The results from comparing the two metadata records created for each game in conjunction with feedback from the students allowed us to identify successes, issues, and areas for improvement.

User Evaluation of the Schema

User interviews: In addition to the previous 24 interviews, we interviewed 32 participants, all over 18 years old, who find, play, purchase, collect, and recommend video games, in order to investigate their current practices surrounding video games. Unlike Phase II, in which the researchers mostly focused on interviewing gamers, the interviewees in Phase III represented a wider range of personas: 3 casual gamers, 5 avid gamers, 4 parents of young players, 6 game collectors, 5 game industry employees, 6 curators/librarians managing video game collections, and 3 scholars conducting game research. Interviewees were recruited by snowball sampling. Gamers were initially recruited via the SIMM's public exhibition, and librarians via the American Library Association (ALA)'s annual conference and the Young Adult Library Services Association listserv. Scholars, employees, and parents were recruited via researchers' personal connections and references. Collectors were recruited via game-related forums. The interview protocol included specific questions about gaming experiences, game-related information needs and search behaviors, feedback on the current metadata schema, etc. However, we were also interested in exploring what participants had to say about games. Interviews for some specific personas had specific questions (e.g., librarians, people in game industry, parents, scholars). All the interviews were conducted between July and October 2013, either in-person, over the phone, or on Skype. Each interview lasted approximately 45 minutes to an hour. Each interviewee was compensated with a \$20 Amazon gift card for their participation.

Interviews were fully transcribed and analyzed in order to obtain a detailed qualitative description of behavior surrounding video game use, purchases, recommendations, organizational

needs, and game-seeking. The code book contained codes for various aspects such as user behaviors, game-related resources, appeal factors, etc. The code book was created through an iterative coding process. Initially, we selected a sub-set of interview transcripts from which we developed a preliminary code book. Using this code book, two coders independently coded the transcripts once. A series of meetings discussing the code revisions followed. Using the revised code book that resulted from these meetings, the two coders reviewed and re-coded the transcripts. Afterwards, a third party reviewed all the coding work, identifying any issues and/or inconsistencies in applying the codes. The instances of questionable code applications were discussed among all three (i.e., two coders and the third party reviewer) until consensus was reached. More detailed description of the coding process as well as a full discussion of findings is provided in another article under preparation (Lee, Clarke, & Rossi, in preparation). The current article will primarily focus on user feedback with regard to the metadata.

Survey: We conducted a large-scale online survey asking 23 questions regarding users' gaming experience and game-related information needs and behaviors, and 5 demographic questions. The objectives of this survey were to determine game-related information needs and behaviors which are commonly shared by a substantial number of users, and to identify which information features are perceived as useful. Through this survey, we sought to validate our previous findings from interviews in a larger population.

The questionnaire consisted of five sections: 1) Questions about gaming experience, 2) Questions about physical games, 3) Questions about digital games, 4) Game-related information needs and search behaviors, and 5) Demographics. The survey was administered via LimeSurvey, an online questionnaire tool, hosted on the University of Washington Information School's server.

Participants were limited to people over the age of 18 who played video games. No specific game playing behavior was necessary to participate in the survey (i.e., participants did not have to play a certain number of games or hours per week to participate). Participants were recruited through a variety of physical and online communication platforms, including various game related mailing lists (e.g., ALA Connect - Games and gaming group, ATLUS forum) and Facebook groups (e.g., International Game Developers Association, Extra Credits), UW iSchool student, faculty, and staff mailing lists, UW iSchool Research Fair, researchers' social networks, and so on. Many users on Facebook and other social media sites chose to share or forward links to their friends and various communities as well. The limitation on generalizing the findings of this survey to a larger population of gamers due to the sampling method should be noted. Participants were given a choice to enter a raffle to win a total of \$200 worth of Amazon gift cards.

The survey was active for approximately 7 weeks from November 19, 2013, to December 30, 2013. A total of 2163 respondents participated in the survey; of those, 1257 completed the survey. Participants took an average of 20 minutes and 49 seconds to complete the survey.

Findings and Discussion

Overview

In this section, we provide a summary of our findings from conducting a collaborative review, testing the schema, and analyzing the user feedback provided through an online survey and interviews. We describe each of the issues identified followed by an explanation of the changes made to the schema to address those issues.

In the survey, we asked the users what information about games is most useful to them for games they are currently playing and also when they are looking for new games to play. Table 2 shows the complete list of metadata elements⁵ sorted by the proportion of positive responses. The

count in the sixth column represents the number of responses that stated the particular element is useful for at least one of the two cases (i.e., for games currently playing OR seeking new games).

TABLE 2. Metadata Elements Sorted by the Proportion of Positive Survey Responses to Their Usefulness

	Games currently playing	Percent (N=1257)	Seeking new games	Percent (N=1257)	Count	Percent (N=1257)
Price	245	19.5%	998	79.4%	1034	82.3%
Platform	594	47.3%	956	76.1%	1026	81.6%
Genre	589	46.9%	930	74.0%	996	79.2%
Series	533	42.4%	881	70.1%	953	75.8%
Style	528	42.0%	864	68.7%	922	73.3%
Gameplay videos	510	40.6%	798	63.1%	895	71.2%
Plot/Narrative	636	50.6%	758	60.3%	893	71.0%
Franchise/Universe	476	37.9%	801	63.7%	859	68.3%
Theme	505	40.2%	776	61.7%	843	67.1%
Mood/Affect	543	43.2%	738	58.7%	838	66.7%
Title	643	51.2%	599	47.7%	822	65.4%
System requirements	394	31.3%	728	57.9%	797	63.4%
Developer	307	24.4%	730	58.1%	774	61.6%
Setting	427	34.0%	641	51.0%	717	57.0%
Format	389	30.9%	619	49.2%	716	57.0%
Trailers	156	12.4%	690	54.9%	713	56.7%
Number of players	453	36.0%	586	46.6%	692	55.1%
Distributor	392	31.2%	613	48.8%	685	54.5%
Retail Release Date	114	9.1%	656	52.2%	684	54.4%
Presentation	396	31.5%	610	48.5%	674	53.6%
Visual Style	410	32.6%	604	48.1%	671	53.4%
Online capabilities	415	33.0%	564	44.9%	659	52.4%
Point of View	398	31.7%	581	46.2%	658	52.3%
Customization options	455	36.2%	478	38.0%	611	48.6%
Screenshots	167	13.3%	566	45.0%	593	47.2%
Difficulty levels	478	38.0%	351	27.9%	577	45.9%
Type of ending	432	34.4%	388	30.9%	572	45.5%
Language	320	25.5%	498	39.6%	559	44.5%
Region	241	19.2%	500	39.8%	545	43.4%
Special hardware	252	20.0%	494	39.3%	539	42.9%
Temporal aspect	319	25.4%	461	36.7%	527	41.9%
Box art/Covers	219	17.4%	378	30.1%	440	35.0%
Packaging	121	9.6%	396	31.5%	433	34.4%
Edition	179	14.2%	331	26.3%	421	33.5%
Purpose	215	17.1%	365	28.3%	415	33.0%

References to historical events	248	19.7%	343	27.3%	413	32.9%
Official website	277	22.0%	337	26.8%	411	32.7%
Publisher	143	11.4%	368	29.3%	410	32.6%
Achievements/Awards/Trophies	341	27.1%	160	12.7%	384	30.5%
Alternative title	203	16.1%	229	18.2%	340	27.0%
Game credits	161	12.8%	201	16.0%	290	23.1%
Rating	108	8.6%	235	18.7%	279	22.2%
Identifier	60	4.8%	92	7.3%	135	10.7%

The elements <price> and <platform> were the most highly rated with over 80% of respondents supporting their usefulness. This was followed by a number of subject metadata elements (e.g., <genre>, <style>, <plot/narrative>, <theme>, <mood/affect>) and metadata about how the games are related to each other (e.g., <series>, <franchise/universe>). Visual information was also highly valued, most notably, <gameplay videos>. The usefulness of elements that are more closely tied to physical aspects of games (e.g., identifier (UPC), packaging, box art/covers) tended to be rated lower. It was also possible to confirm that the usefulness of certain elements does change depending on the information seeking stages (i.e., already playing the game vs. seeking new games), as previous literature suggests (Google, 2012). The elements <title>, <difficulty level>, <type of ending>, and <achievements/awards/trophies>, were deemed more useful for the games users are currently playing. All the other elements were deemed more useful for finding new games to play.

Redefining the Personas

Upon reviewing the interview transcripts from Phase II, the researchers determined that the original six personas were not fully representative of the variety of people who interact with video games. Therefore, several additional personas were added to better represent the types of individuals who might be interested in seeking information about video games. The newly added personas are as follows: educator/teacher, casual gamer (as opposed to avid gamer), and industry professional (specifically those who are not game designers/developers). A single person may have aspects of multiple personas. This became evident as we interviewed more people. Despite our efforts to recruit people representing specific personas, they often represented multiple personas (e.g., casual gamer and parent; avid gamer and designer/developer).

Patterns of Discrepancies among Catalogers

Due to the nature of the highly descriptive data, we did not attempt to calculate specific quantitative measures for evaluating inter-indexer consistencies. Rather, we performed a qualitative review of consistencies of descriptions provided by different catalogers in order to identify problematic elements and patterns of discrepancies.

The most common discrepancies that emerged during cataloging were regarding which conceptual level a cataloger was describing—work or manifestation? This was especially problematic for games available on multiple platforms (e.g., *Candy Crush Saga* for iTunes, Android, Facebook, etc.). Some chose to describe a game at the work level, listing all the platforms for which the game is available. Others chose to describe a game at the manifestation level, such as the specific instantiation on a particular platform. This resulted in discrepancies in the amount of information provided for some of the elements. This confusion about conceptual level of description inspired development of a conceptual data model to inform catalogers about exactly

which entity they are describing (e.g., game, game edition, local release, etc.) (Jett, Sacchi, Lee, & Clarke, under review).

Subjective elements such as plot, theme, mood, etc. require familiarity with the game or playing experience. Otherwise, the cataloger needs to rely on secondary sources (e.g., watching a gameplay video, reading a Wikipedia page) to determine such information. Because catalogers drew from different sources (i.e., personal experience of playing the game vs. research on the game), this sometimes resulted in inconsistent descriptions. To address this issue, we created a list of preferred sources of information for catalogers to reference.

Some catalogers reported that the large amount of information required to describe certain elements, such as achievements/trophies/awards, customization options, credits, etc., led to inconsistent recordings, and may not result in a high return on investment of cataloger's time. We eliminated or modified these elements to narrow the scope, increase consistency, and reduce the workload.

Describing information for visual style of a game also led to discrepancy among catalogers due to lack of understanding of and familiarity with terms for describing techniques used to create certain displays (e.g., rasterized, ray traced, low-poly). We decided to revise the controlled vocabulary to focus on the visual appearance rather than creation techniques. In addition, a follow-up user study is currently underway to test if consistency can be improved with training materials. These materials contain a number of visual examples for each term, as requested by the catalogers.

Relationships among Games

Elements representing different relationships among related games (i.e., <series>, <franchise/universe>) were deemed highly useful as evidenced by both the survey results and interviews. This information seemed not only useful for gamers (e.g., *"If I was looking for a new game to play, it would probably be like the series, because it [has] definite connection to something you know"* (P27, Avid gamer)), but also users who purchase games for other people (e.g., *"My six year old likes... he loves Mario, just about any Mario game"* (P48, Parent)). Testing these elements by cataloging real games helped clarify that <franchise> and <universe> are not the same and therefore they should be split. The term <universe> is understood and used as a "super series," whereas <franchise> refers to the intellectual property associated with particular games and is related to marketing and commercial applications. The definition for the element <franchise> was revised to reflect this understanding, and <universe> was removed, as this can be represented by using a recursive relationship for the element <series>. This discussion also led to the realization that development of a data model will be necessary to better represent the different relationships among games. Furthermore, some interviewees were interested in searching and browsing games based on their influences on each other.

*"The other one is like the DNA of the game - **this game influences this game**. Those are things that even me right now I would be interested to see: that game influenced this game. That kind of circles back into "hey, these people like this game, then they also like this game." There's these connections that are sort of meta-connections between the games that are interesting to see from just a data point of view... It's like movies or song albums..." (P38, Industry professional)*

"I think that's very important in games, because games, to a certain extent, a lot of games are pretty derivative of each other, and that's ok. You know, that's how things become refined and tracing the different paths of these strains of influence, go, tracing the directions they go is really great." (P20, Avid gamer)

We expect that some of these influences may be inferred by elements such as <developer>, <theme>, <setting>, etc.; however, it would be potentially useful to specifically include some sort of “influenced by” relationship in the data model to explicitly capture this information.

Expansions/Alterations for Games

Additional content for video games such as downloadable content (DLC), modifications (“mods”), patches, etc., is an increasingly important aspect of gaming distribution and revenue generation models, particularly for digitally distributed games. The element <additional content> was added to the schema to capture this information. In addition, an increasing number of games are offering various ways for players to connect to other gamers and to online stores and servers. This being the case, information about <online capabilities> is highly relevant to many gamers (e.g., “*Capability to connect online, for phone games especially, is really important. Cause I like knowing my rank in comparison to my friends*” (P3, Casual gamer)). In order to record more specific information related to <online capabilities>, it was split into two separate elements: <networked features> and <connectivity>. The element <networked features> will be used to describe the ways in which the game can be experienced by connecting to other entities, such as online multiplayer, co-op mode, leaderboards, content download, etc., whereas <connectivity> was created to capture the information about the technical requirements (e.g., hardware) necessary for connection. By “connection,” we are not only referring to connecting to the Internet; many handheld game systems, such as Gameboy or GameGear, allow players to link their systems using a proprietary patch cable. The element <connectivity> is used to specify all these different types of connection that can be made among multiple systems. This is an increasingly important issue, especially for recent video game consoles (e.g., Xbox One initially attempting to require a user’s periodic connection to the Internet to run the console), and undoubtedly an important consideration for game players. There is also an unresolved issue of the instability of <networked features> related information; a game’s online server can be shut down due to diminishing popularity, or a company can decide to discontinue online services on a particular platform which will result in terminating online play features, leaderboards, etc. for many games⁶.

Importance of Subject Metadata

The usefulness of elements describing the content, subject, or the nature of users’ experience of games such as <genre>, <style>, <plot/narrative>, <theme>, and <mood/affect> was also highly rated by the survey participants. This is also resonated in a number of interviews:

*“So having that in there **might help to, as a search term or as an organizational term, help to separate sub-genres of, you know, a plot/narrative or theme that, you know, this is a comical way to prevent the apocalypse versus a very serious way to prevent the apocalypse, which I think might be a kind of good way to organize.**” (P28, Gamer)*

*“I think **mood and affect are the ones probably the most important to me...** because I play games for fun and entertainment and I don’t want to be disturbed or scared or depressed by a video game.” (P2, Casual gamer)*

*“**Writing, and 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.**” (P26, Gamer)*

A number of interviewees also commented that while the element <plot> can be useful for grouping games with similar plots together, they need additional information on what the game is really about. Some concern was raised as to what level of granularity should be used to describe the plot:

*“I think if you go too detailed then you are kind of giving away parts of the game, but if you go too general, I would say that these examples would be slightly too general for me to be particularly interested in the game. Particularly because **lots of games are where the protagonist explores the world**. That’s Mario, that’s most MMOs are kind of, you’re just basically exploring the world and finding quests. (P28, Gamer)”*

The new element <summary> was added to address this issue and provide more detailed information about a game’s narratives. This information would not be used for collocation purposes but rather to help users make informed decisions about individual games. <Theme> was also mentioned as useful because it can aid users’ decisions on playing a particular game (e.g., “I guess also **theme would be a big one**, because one of the options here is supernatural zombies, I’m not going to play a game with zombies... it is too real in the game” (P2, Casual gamer); “I work with a community called Queer Geek which is a GLBTQ, anything that has like a **minority presences[sic] in game [would be useful]**” (P11, Avid gamer)).

In addition, the element <links to historical events> is now subsumed under <setting> in order to remove duplication of data. <Setting> was revised to have three facets: spatial, temporal, and cultural. Under each facet, the value will be qualified as “real” or “imagined.” <Setting> can provide more information about the gameplay in addition to <genre> (e.g., “This is the kinda, it’s set in modern times in the city, you know, kinda giving them a sense of what the gameplay is like” (P43, Curator/Librarian)). Information about the connection to historical events may be useful for gamers as well as parents and educators (e.g., “[T]hey each have historical things that they’re crazy about. Like I said, [name redacted] is very WWII, but [name redacted] is very interested in Roman history and the American Revolution – which is why Assassin’s Creed is so interesting to him, so yeah, they definitely would care very much about that” (P47, Parent)). The element <temporal aspect> was also renamed to the more commonly used term <pacing> based on user warrant. An issue was raised during the review process that there can be multiple dimensions of time flow in a game. As a result, three different temporal aspects will be represented in this element: “battle system” (i.e., how the time flows in a battle), “in-game clock” (i.e., how the time passes in the game world), and “time manipulation” (i.e., how characters interact with time). For instance, *Final Fantasy: Lightning Returns* has a real-time battle system, the in-game clock runs continuously during the gameplay, and the protagonist is able to temporarily stop the time. As some interviewees suggested, this information can be critical for gameplay (e.g., “When you say a turn-based strategy game, rather than RTS [real-time strategy], those are different genres... They play differently because time is used differently” (P53, Scholar)).

Need for Visual Metadata

Of the four different kinds of visual information about games (i.e., gameplay videos, trailers, screenshots, box art/covers), survey respondents found <gameplay videos> to be most useful, and this was often mentioned as highly useful by interviewees as well:

*“ Even though it looks cute from a still image or if I want to see also [if] it is too simple, **you can only get that from seeing someone actually play** and that would also get at possibly **how difficult controls are and how difficult it is to actually play the game** a little bit.” (P2, Casual gamer)*

“I flat out don’t like playing sports games, racing games, and fighting games as much, but that you can sort of figure out with the gameplay videos. Even if it’s a genre you don’t really like, I

don't really like fighting games, but I like Super Smash Bros. I can discern that from the gameplay video. (P37, Industry professional)

"Definitely the gameplay video, that is going to sway me one way or the other. If something is like a polished piece and it looks well and it is well presented and it looks like something I've played before or something I would like to play, then that's probably [going to] get me to spend my money." (P22, Gamer)

Interviewees seem to think that trailers would be more entertaining than informing (e.g., "[T]railers are entertaining but it is usually just the CG [computer graphics] animated thing that is not even related to the game and it's not actual game play" (P22, Avid gamer)). Screenshots may be problematic because it can be difficult to determine if they are from pre-generated video clips, actual gameplay, or if they are fan-created artwork (e.g., "I feel like the problem with screenshots at least in my mind is veracity – how do you determine that a screenshot is actually from the game that you're claiming it's from?" (P5, Collector)).

Challenging Descriptive Metadata

There were several descriptive metadata elements that were particularly difficult due to issues such as the lack of available information sources, instability of information, etc. <Price>, for instance, was identified by our survey participants as the most useful element, but was particularly difficult for our catalogers to research and transcribe in a straightforward manner. Several reasons for the difficulty were 1) lack of immediately available primary sources from which to derive the information; 2) regular price fluctuations of both physical and digitally-distributed games; and 3) lack of supporting metadata to give context to the price. In order to at least partially address these issues, the date when the price information was captured and the source of the information will be recorded as qualifiers with this metadata element.

In addition, the <region> element also had to be revised from its previously broad definition of "the names used to refer to a place, region, or territory where a game is designated as playable." This definition turned out to be problematic because it was not always clear how one determines if a game is playable in certain regions. For instance, if a game app is released in multiple languages, does that mean the game was intended to be playable only in the regions where those languages are spoken? The element was renamed and redefined so that it better reflects the overall intent, which was to provide users with necessary information to play a particular game release (e.g., knowing if they will be able to play the game they bought in Japan on their PlayStation).

*"Region would be very important... my kids are always like, oh it's out in Japan but not here. It's here but my friend in Portugal doesn't have it. Especially for things that have online components, **they want to play with their friends that they've made all over the globe.** Or people that they've met in Fandoms."* (P39, Curator/Librarian)

Such information is multi-tiered, requiring in-depth knowledge of electronic circuitry, Digital Rights Management (DRM), analog/digital signal processing, and more prevalently, approximate physical location. Game companies sum up these needs into their own proprietary region codes, which often differ from system to system. The new definition specifically refers to "the classification code indicating the video encoding and regional hardware necessary to realize the game" (e.g., North America NTSC-U/C; Japan and Asia NTSC-J). An issue remains that many older games, as it turns out, do not have region codes printed on the box, manual, or cartridge, often leading to confusion as to what exactly is required to play the game, in terms of both hardware and

software. A comprehensive analysis of the technical specifics as well as the way companies describe region in their own games must be done to create a useful controlled vocabulary.

Issues in Describing Digitally Distributed Games

The elements <format>, <packaging>, and <box art/cover> turned out to be problematic, particularly for describing digitally distributed games. By “digitally distributed games,” we are referring to games that have no physical container or component, and are accessible only through download or streaming (Lee et al., 2014). The term “born-digital” in the controlled vocabulary for <format> was revised due to the lack of specificity in describing different varieties of digitally distributed games, such as downloadable, flash-based, and streaming games. Additionally, it is the nature of video games that they are all born-digital from a developmental and carrier standpoint and thus it was an inappropriate term for describing games which have no physical component. The name of the element <box art/cover> was changed to <representative art> in order to ensure its applicability to digitally distributed games in addition to physical games. The definition of <packaging> was revised so that it clearly states its applicability to physical games alone.

In addition, there were 16 survey responses to the open-ended question asking for other useful information that specifically mentioned DRM. Five of these responses simply referred to DRM without further specification. Two mentioned desires for DRM-free games. The remainder commented on specific DRM restrictions, such as *“a limited amount of installs per license or games that demand that you log into a server to play a single player game is horrible,”* *“always [sic] online is a nightmare,”* and *“does it require that the CD be in the drive at all times?”* Some respondents specifically mentioned wanting to avoid “intrusive DRM” and wanted to know what types of DRM and other limitations apply to a particular game before they purchase it. This information is not only useful to game players from a practical viewpoint, but also can be of interest to game scholars studying the legal aspect of gaming. The researchers decided to include this new metadata element with an expectation that this will be increasingly important as digitally distributed games become more prevalent (e.g., *“The game starts out with DRM, and then they take it away, [it] changes the game in sort of a very fundamental way, particularly if the game is online or needs to be online to function. So I feel like digital rights management, even if it’s just a yes or no tick box could be important, and necessarily important for the platform that it’s on.”* (P5, Collector)). Information regarding limited install activation, persistent online authentication, software tampering, and rights to copy, share, and resell the game will be recorded under this element. We expect to further develop this controlled vocabulary to fully represent the wide variety of DRM techniques and technologies in our future work.

Managing Complexity

A number of elements were removed from the recommended set and some elements were simplified in order to avoid putting undue burden on catalogers, as explained below.

Characters: For many catalogers, this element quickly became an unruly mess and unavoidable time expense. Complications were numerous, some of which cannot be easily solved. First, it is difficult to clearly indicate what constitutes a main character. Criteria such as being a playable character or named character were immediately met with counterexamples (e.g., main antagonist in the game who is not playable, unnamed (customizable) protagonist). Also, a particular character can appear across multiple games in a series, but it can be unclear if they are in fact one and the same character (e.g., all the Links in the *Legend of Zelda* series). Other issues include the ambiguity regarding types and depth of information to be transcribed, the sheer number of characters in some cases, the lack of a controlled vocabulary to handle the ascription, how to handle non-playable characters, and how to gauge the necessity of ascribing one character over another in a game.

Considering the level of complexity this element brings to the schema, it was decided that a separate database of game characters specifying the source of and relationships among the characters would need to be established in order to describe this information in a satisfying manner.

Controls: This element was challenged due to the following reasons: 1) for certain titles, the list of controls can be quite extensive and can vary from character to character (e.g., an RPG title with multiple playable characters who each have their own set of attacks, magic and healing spells), making it difficult to discern which controls should be included or not, and 2) for certain titles, the controls can actually be reprogrammed by the player. The potential usefulness of describing this information in a metadata record was also unclear because most newer games, especially digitally distributed games, come with tutorials embedded in the games themselves. This led to the decision to remove <controls> from the recommended set of elements, and place it in a more comprehensive full set, currently under development.

Achievements, Awards, Trophies: There were two main issues with this element: 1) this is a very detailed level of information that may contain spoilers, and 2) it was unclear to the catalogers what exactly constitutes as an achievement. The original intent was to record only the pre-programmed in-game achievements based on specific consoles (e.g., PlayStation Trophies, Xbox Achievements). However, this was challenged by multiple games on Steam that have growing numbers of achievements as new patches are released (e.g., *Team Fortress 2* was launched with 17 achievements that later grew to 477 achievements). The researchers determined that this information is better represented and managed in a game wiki rather than in the metadata record, and thus decided to remove this element.

Credits: This element was removed from the recommended set for the following reasons: credits turned out to be exceptionally inconsistent from one game to the next; after examining the credits information from multiple games, it became clear that a wide variety of job titles are used without clear indications of how the people actually contributed to creating the game. There is also the issue of authority control; we need to be able to correctly attribute the work to the right person. For that, we need an authority file of people involved in video game industry. Further research is necessary for <credits> to be included as a viable element in the schema.

Visual style: The visual style was deemed useful by over 50% of the survey participants and also was mentioned several times by multiple interviewees as important. However, several catalogers mentioned that it was difficult to determine the visual style of sample games, in particular, the technique facet. The technique facet was designed to represent information on the tools and techniques used to render the visuals. This may be highly relevant to game developers/designers, but have less significance to average game players. The interview data suggested that what mattered more to users was the visual style and dimension information that reflected how they experienced the game, rather than the techniques and technologies used to create that experience. (e.g., *"The visual style, I usually play realistic game or cel-shaded graphics, that's what I play too – that's the comic book style."* (P22, Avid gamer)). Therefore we decided to keep <visual style> and <dimension> as two separate elements in the recommended metadata set, and leave the visual technique to be included in the larger full set.

Customization options: Testing the schema revealed that the criteria for inclusion in this element needed further specification. There are numerous things that can be customized in many games, such as characters, difficulty levels, display, sound, etc. This quickly became excessively complicated. After much discussion, the researchers decided to describe only two particular

customization options under this element: characters (e.g., *“I’m thinking about the Mass Effect games, they’ll come in and say that they loved how customizable it is.”* (P39, Curator/Librarian), *“I really did like that in Mass Effect you can make your own person which is pretty cool. I made it [to] look like me.”* (P22, Avid gamer)) and difficulty levels (e.g., *“Difficulty levels are important, I like to feel challenged and I like to be able to advance and have a sense of completion.”* (P22, Avid gamer)) as these were most commonly mentioned by the interviewees.

Revised Version of the Schema

The schema was further refined to version 2.0 based on the collaborative review, creating metadata records of 65 sample games, feedback from the student catalogers, and user data from the interviews and the survey. Table 3 presents the elements, definitions, and instructions from Version 2.0. We also provide a crosswalk table (Table 4) comparing our schema to other general schemas/standards that are typically used in cultural heritage institutions: in particular, Dublin Core, MARC (MACHINE-Readable Cataloging), and RDA (Resource Description and Access). The lack of comparable elements in other schemas/standards illustrates the limitations of using them to fully describe video games.

TABLE 3. Metadata Elements from the Revised Schema

Element	Definition and Instruction
Title	Proper names that are used to refer to a video game, assigned by the creator. (modified from CIDOC CRM, 2014, p.16)
Edition	A word or phrase appearing in the game being described that indicates a difference in either content or form between it and a related game (e.g., second edition, greatest hits, collector’s edition, limited edition). (modified from FRBR, 2009, p. 41)
Series	Proper names of a set of related games, often indicated by consecutive numbering, continuing narrative, or similarities in game play and themes, to which the game being described belongs.
Franchise	A commonly used name referring to the intellectual property, related data, and content shared among a group of cultural objects to which the game being described belongs.
Corporate body	An individual, organization, or group of individuals or organizations responsible for creation, realization, manufacture, marketing, and/or distribution of a game. (modified from FRBR, 2009, p.25)
Platform	The hardware and operating system on which the game was designed to be played. (e.g. PlayStation 3; XBOX 360; Nintendo 3DS; Android 4.4 KitKat; Apple iOS 7, PC Windows XP, Mac OS X)
Format	The distribution medium or method that provides the executable code of a video game. (e.g. cartridge, optical, magnetic, downloadable, streaming)
System requirements	Hardware, firmware, and/or software components that are prerequisites for running the game on a particular platform. (e.g. 2GB RAM, 8GB Hard Disk Space, DirectX 9.0C, 256MB ATI HD26000 XT Video Card, DirectX 9.0c Compatible Sound Card, Quad Core 1.8GHz Processor)
Special hardware	A hardware that is required or recommended for playing the game in addition to the main platform. (e.g. motion controller; gaming headset)

Networked features	The ways in which the game can be experienced through connection(s) to other entities, such as game companies, third-party organizations, and other players. (e.g., online multiplayer, content download, leaderboards, voice, spectators)
Connectivity	The technology through which the networked features are realized. (e.g., Client-server based, Ad-hoc, Broadband, LAN)
DRM	Digital rights management technologies intended to control the use of the game.
Additional content	The type and the name of the additional content including DLC, MODs, Patches, etc.
Number of players	The number or range of the number of players the game can accommodate either separately or concurrently.
Region code	The classification code that indicates the video encoding and regional hardware necessary to realize the game. (e.g., North America NTSC-U/C; Japan and Asia NTSC-J; Europe, New Zealand and Australia PAL; China NTSC-C)
Language	The classification code for the language(s) in which the game conveys information.
Retail release date	The date of the public/commercial release of the game.
Rating	The classification of the content in the video game for informed decision making about the game, provided by organizations such as professional associations, game distributors, or creators. (e.g., MA-13 Parental Discretion Advised. Mature Audiences; Everyone. E. (ESRB)). Note that some older games do not have this rating information.
Official website	A URL of the website for the game from the companies officially associated with the game.
Price/MSRP	The manufacturer's suggested retail price (MSRP) at time of initial release in the country or region where the game was released.
Award	The names(s) of any awards the game has won and the granting body that awarded them.
Customization options	The in-game options for difficulty level and characters which can be modified by the player for personalized experience.
Packaging	All items included in the original packaging of the game. (e.g. 2 game discs, soundtrack, manual, action figures)
Genre	The overall nature of a game based on its objectives, types of rules, distinctive characteristics, modes of action, manners of gameplay, and how a player interacts with it.
Plot	The underlying events that make up the story of the game.
Summary	A brief statement or account of the main points of the game.
Theme	A common thread, motif, subject, or idea that recurs in the game. (modified from Getty Art & Architecture Thesaurus) (e.g. fantasy-princess; supernatural-zombies)
Setting	The location, time frame, and/or cultural context in which the game takes place.
Mood	The pervading atmosphere or tone of the video game which evokes or recalls a certain emotion or state of mind.
Pacing	The methods by which time passes in the game and/or manner in which events take place.

Estimated time of completion	The estimated average time to complete the game.
Type of ending	The characteristics describing how the game ends and/or post-game content.
Visual style	The predominant and recognizable visual appearance of a video game as originally intended by its creator, and/or determined in the context of creation.
Dimension	The intended perception of the depth of the represented entities inside the game. (e.g., 2D, 3D, Stereoscopic 3D, Multiple)
Point of view	Perspective from which the player experiences the gameplay. (e.g. first person; third person)
Representative art	The officially released image that is representative of the game, prominently featured in a physical or digital distribution package
Screenshots	Still images taken during the gameplay.
Trailers	Video footage released and/or endorsed by the developer/publisher of the game for promotional purposes.
Gameplay videos	Video footage of the gameplay excluding such things as introductions, cutscenes, or trailers.
Note	Any other notable characteristics of the game.

TABLE 4. Metadata Crosswalk Table⁸

Our schema	DC Terms	MARC fields	RDA no.
Title [M][R]	title	245a	5.1.3
Edition	hasVersion; isVersionOf	250	2.5.1
Series [R]	isPartOf	490	2.1.2
Franchise [R]			
Corporate body [M][R]	creator; publisher	110	11.2
Platform [M][R]		337	3.2
Format [M]	format	338	3.3
System requirements [R]	requires	538	3.2.0
Special hardware [R]	requires	538	3.2.0
Networked features [R]			
Connectivity [R]			
DRM [R]	license accessRights	506;540	4.4 4.5
Additional content [R]	isReferencedBy	525	J3.5; J4.5; J5.5 (varies depending on FRBR level)
Number of players [M][R]			
Region code [M][R]		044	3.19.6
Language [R]	language	041	6.1.1
Retail release date [M][R]	issued	260\$c	2.8
Rating [R]	audience	521	7.7

Official website [R]	relation	856u	4.6
Price/MSRP [R]		365	4.2.1.3
Award [R]		586	7.28
Customization options			
Packaging		300	3.4
Genre [M][R]	type	655	6.3
Plot [R]	abstract	520	7.10
Summary	abstract	520	7.10
Theme [R]	subject	650	13
Setting [R]	coverage	651	
Mood [R]			
Pacing			
Estimated time of completion			7.22
Type of ending			
Visual style [R]		650	
Dimension			3.17.2 (for motion picture film)
Point of view			
Representative art [R]	description		
Screenshots [R]	hasPart relation		J3.5
Trailers [R]	relation		J3.5
Gameplay videos [R]	hasPart relation		J3.5
Note		500	5.9

Conclusion and Future Work

Video games are important part of our cultural heritage. Games possess an innate ability for human expression, and reflect how we define ourselves in current history (Monnens et al., 2009). It is of utmost importance to organize, describe, and preserve this legacy for future generations. Our work from the past three years is only the start in pursuing this immense and critical research agenda.

This is an ongoing research effort as we continue to work on refining the elements and definitions to better represent new types of games. We also plan to continue to work on creating controlled vocabularies for the new metadata elements such as <connectivity>, <networked features>, <additional content>, etc. and revising the vocabularies for elements such as <format> and <theme>. In addition, there is a separate research stream to develop a metadata application profile based on CIDOC CRM (ICOM/CIDOC Documentation Standards Group, 2014), specifically targeted for game museums. This metadata profile will include a number of additional administrative elements as well as item-level descriptions (e.g., provenance, value, display record).

Even with a comprehensive list of metadata elements, there is still information that is not easy to represent. The complex relationships that exist among games (e.g., series, influences) are particularly tricky. In order to fully represent these relationships, the authors have been developing a conceptual data model (Jett et al., under review). The revised metadata elements discussed in this paper will be fully incorporated into the data model. We envision that the data model can be extended to represent ephemera and other materials related to games such as action figures,

artwork, music, etc., to create a robust and holistic representation of the video game domain that serves users' needs.

Acknowledgements

The authors would like to thank everyone who participated in the interviews and survey, and helped recruit participants by sharing our invitations. We extend special thanks to the students in the INFO498/INFX598 video game metadata courses for their valuable contributions in revising the interview and survey questions, and James Portnow at DigiPen for his help in recruiting survey participants. This research is supported by the Bridge Award from the Office of Research at University of Washington.

Reference

- Beak, J., & Olson, H. A. (2011). Analysis of metadata schemas for children's libraries. In Smiraglia, R. P. (Ed.), *Proceedings from North American Symposium on Knowledge Organization*, Vol. 3 (pp. 1-12). Toronto, Canada.
- Bloomfield, M. (2001). Indexing: neglected and poorly understood. *Cataloging & Classification Quarterly*, 33 (1), 63-75.
- Bruce, T. R., & Hillmann, D. I. (2004). The continuum of metadata quality: defining, expressing, exploiting. In D. Hillmann & E. Westbrook (Eds.), *Metadata in Practice*, Retrieved Mar 10, 2014, from: <http://ecommons.library.cornell.edu/handle/1813/7895>
- Cooper, A. (1999). *The inmates are running the asylum*. Indianapolis: Sams.
- Crookes, D. (2012). British Library starts videogame website archive project. Retrieved Sep 11, 2013, from: <http://www.independent.co.uk/life-style/gadgets-and-tech/news/british-library-starts-videogame-websitearchive-project-6858507.html>
- Currier, S., Barton, J., O'Beirne, R. & Ryan, B. (2004). Quality assurance for digital learning object repositories: issues for the metadata creation process, *ALT-J Research in Learning Technology*, 12 (1), 5-20.
- Donovan, A., Cho, H., Magnifico, C., & Lee, J. H. (2013). Pretty as a pixel: Issues and challenges in developing a controlled vocabulary for video game visual styles. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 413-414).
- Funk, M.E., Reid, C.A., & McGoogan, L.S. (1983). Indexing consistency in Medline. *Bulletin of the Medical Library Association*, 71 (2), 176-183.
- Google. (2012). Understanding the modern gamer. Retrieved Mar 7, 2014, from: <http://www.thinkwithgoogle.com/research-studies/understanding-the-modern-gamer.html>
- Guy, M., Powell, A. & Day, M. (2004). Improving the quality of metadata in e-print archives. *Ariadne*, 38. Retrieved Mar 25, 2014, from <http://www.ariadne.ac.uk/issue38/guy/>

- Hillmann, D. I. (2008). Metadata quality: from evaluation to augmentation. *Cataloging & Classification Quarterly*, 46 (1), 65-80.
- Hughes, A.V. & Rafferty, P. (2011). Inter-indexer consistency in graphic materials indexing at the National Library of Wales. *Journal of Documentation*, 67 (1), 9 – 32.
- ICOM/CIDOC Documentation Standards Group (2014). Definition of the CIDOC Conceptual Reference Model. Retrieved Mar 7, 2014, from: http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf
- IFLA Study Group on the FRBR, IFLA Section on Cataloging, Standing Committee. (2009). *Functional Requirements for Bibliographic Records: Final report*. K.G. München: Saur Verlag.
- Jett, J., Sacchi, S., Lee, J. H., & Clarke, R. I. (under review). A conceptual model for video games and interactive media. Submitted to ACM/IEEE Joint Conference on Digital Libraries (DL 2014).
- Lee, J. H., Cho, H., Fox, V., & Perti, A. (2013a). User-centered approach in creating a metadata schema for video games and interactive media. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 229-238).
- Lee, J. H., Clarke, R. I., & Perti, A. (2014). Metadata for Digitally Distributed Video Games at the Seattle Interactive Media Museum. In *Museums and the Web 2014*, N. Proctor & R. Cherry (eds). Silver Spring, MD: Museums and the Web. Retrieved January 24, 2014, from: <http://mw2014.museumsandtheweb.com/paper/metadata-for-digital-video-games-at-the-seattle-interactive-media-museum/>
- Lee, J. H., Clarke, R. I., & Rossi, S. (in preparation). A qualitative investigation of users' video game information needs and behaviors.
- Lee J. H., Tennis, J. T., Clarke, R. I., & Carpenter, M. (2013b). Developing a video game metadata schema for the Seattle Interactive Media Museum. *International Journal on Digital Libraries*, 13 (2), 105-117.
- Lee, J. H., Karlova, N. A., Clarke, R. I., Thornton, K., & Perti, A. (2014). Facet analysis of video game genres. In *Proceedings of the iConference 2014*.
- McDonough, J., Kirschenbaum, M., Reside, D., Fraistat, N., Jerz, D. (2010a). Twisty little passages almost all alike: applying the FRBR model to a classic computer game. *Digital Humanities Quarterly*, 4(2).
- McDonough, J., Olendorf, R., Kirschenbaum, M., Kraus, K., Reisd, D., Donahue, R., Phelps, A., Egert, C., Lowood, H., & Rojo, S. (2010b). *Preserving Virtual Worlds Final Report*. Retrieved Sep 11, 2013, from: <http://hdl.handle.net/2142/17097>
- Moen, W. E., Steward, E. L. & McClure, C. R. (1997). The role of content analysis in evaluating metadata for the U.S. Government Information Locator Service: results from an exploratory study. Retrieved Mar 25, 2014, from <http://www.unt.edu/wmoen/publications/GILSMDContentAnalysis.htm>

Monnens, D., Armstrong, A., Ruggill, J., McAllister, K., Vowell, Z., & Donahue, R. (2009). Before it's too late: a digital game preservation white paper. Retrieved Mar 17, 2014, from http://wiki.igda.org/images/a/ab/IGDA_Game_Preservation_SIG_-_Before_It's_Too_Late_-_A_Digital_Game_Preservation_White_Paper.odt

National Information Standards Organization (NISO). (2007). A Framework of Guidance for Building Good Digital Collections. Bethesda, MD: NISO Press. Retrieved Mar 25, 2014, from <http://www.niso.org/publications/rp/framework3.pdf>

Newman, J. (2009). Save the videogame! The National Videogame Archive Preservation, Supersession and Obsolescence. *M/C Journal*, 12 (3). Retrieved Mar 17, 2014, from <http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/167>

Owens, T. (2012). Yes, the Library of Congress has video games: an interview with David Gibson. Retrieved Jan 23, 2014, from: <http://blogs.loc.gov/digitalpreservation/2012/09/yes-the-library-of-congress-has-video-games-an-interview-with-david-gibson/>

Park, J. (2009). Metadata quality in digital repositories: a survey of the current state of the art. *Cataloging & Classification Quarterly*, 47 (3/4).

Shukair, G., Loutas, N., Peristeras, V., & Sklarß, S. (2013). Towards semantically interoperable metadata repositories: The Asset Description Metadata Schema. *Computers in Industry*, 64 (1), 10-18.

The Getty Research Institute. (n.d.) Art & Architecture Thesaurus Online. Retrieved Mar 17, 2014, from: <http://www.getty.edu/research/tools/vocabularies/aat/>

Appendix

Interview Protocol

Hi, I'm [name] Thanks for agreeing to participate in our study. Today I'm going to ask you some questions about your video game preferences and habits. We want to know what you think about video games! Just to be clear, when we talk about video games, we mean everything from PC games, console games, online games, even games that you play on your smartphone or tablet. Our definition is very wide, and we don't want to leave anything out. If you're not sure if something you're thinking about is a "video game," feel free to ask. I'm going to be audio recording this interview just so we can remember what you said. The recording will be transcribed for analysis, but your name and any other identifying information will be removed.

Any questions before we begin?

Before we dive into the questions about games, I'd like to ask you a few basic questions:

- In what year were you born?
- What is your profession?
- Do you identify with any particular ethnicity?

Now I'd like to ask you some questions about playing video games:

- What are your all-time favorite video games?
- Why are these games your favorite?

- What game-playing devices do you currently own?
- What systems/consoles/devices have you used in the past that you no longer use?
- Approximately how long have you been playing video games?
- How often do you play video games?
- How long do you usually play games in a single session?
- Of the time you spend playing video games, what percentage is spent playing alone versus playing with other people?
- Where do you usually play games?
- Why do you play video games?
- What percentage of your video game collection consists of physical games versus digital games?
- Between physical and digital games, which do you prefer and why?
- What percentage of your video game collection do you actively use?

[If applicable] Now I'd like to ask you some questions about physical video games:

- How many different physical video games do you play in an average week?
- Approximately how many physical video games do you currently own?
- Where do you usually acquire physical video games?
- What percentage of your physical video games did you acquire new vs. used?
- How often do you acquire physical video games?
- How do you learn about physical video games?
- How do you usually organize your physical video game collection?
- What do you do when you are no longer interested in a physical video game?

[If applicable] Now I'd like to ask you some questions about digital video games:

- How many different digital video games do you play in an average week?
- How many digital video games have you downloaded, streamed or accessed in the last month?
- From which sources do you usually acquire digital video games?
- How do you learn about new digital video games?
- How do you organize your digital video game collection?
- What do you do when you are no longer interested in a digital video game?
- What percentage of your digital collection is only available digitally (meaning the game was never physically published in cartridges, disc, etc.)?

Finally, I'd like to ask you some questions about how you find information about video games:

- Which websites or related resources do you visit to find more information about video games?
- What do you use these websites or related resources for?
- Which of the websites you mentioned do you find most useful when looking for information about games?
- Please tell us why you find that website most useful:
- What information about games is most useful to you for games that you are currently playing and when you are looking for new games to play?
- Is there any other game-related information that is useful to you?
- Looking over the metadata elements, which of those elements do you think would be useful to you in an information system for video games (such as a database or website)?
- Are there any that you think would be not useful or problematic?

That's all the questions I have for you. Is there anything else you'd like to add?

Relevant Questions from the Survey Questionnaire

[Q21] What information about games is most useful to you for games that you are currently playing and when you are looking for new games to play?

	Useful for games that you are currently playing	Useful for deciding which new games to play
title	<input type="checkbox"/>	<input type="checkbox"/>
alternative title	<input type="checkbox"/>	<input type="checkbox"/>
edition (e.g., collector's edition, limited edition)	<input type="checkbox"/>	<input type="checkbox"/>
format (e.g., cartridge, disc, born-digital)	<input type="checkbox"/>	<input type="checkbox"/>
series (e.g., Final Fantasy, Halo)	<input type="checkbox"/>	<input type="checkbox"/>
franchise/universe (e.g., Final Fantasy Franchise including Final Fantasy series, Crystal Chronicles, Kingdom Hearts, etc.)	<input type="checkbox"/>	<input type="checkbox"/>
platform (e.g., PlayStation 3, Xbox 360, Nintendo 3DS)	<input type="checkbox"/>	<input type="checkbox"/>
developer (e.g., Bungie for <i>Halo</i>)	<input type="checkbox"/>	<input type="checkbox"/>
publisher (e.g., Microsoft Studios for <i>Halo</i>)	<input type="checkbox"/>	<input type="checkbox"/>
distributor (e.g., Steam)	<input type="checkbox"/>	<input type="checkbox"/>
special hardware (e.g., motion controller, gaming headset)	<input type="checkbox"/>	<input type="checkbox"/>
online capabilities (e.g., playing the game online, downloading additional features)	<input type="checkbox"/>	<input type="checkbox"/>
system requirements (e.g., memory, software version)	<input type="checkbox"/>	<input type="checkbox"/>
game credits (i.e., list of people who contributed to creating the game)	<input type="checkbox"/>	<input type="checkbox"/>
official website	<input type="checkbox"/>	<input type="checkbox"/>
price/MSRP	<input type="checkbox"/>	<input type="checkbox"/>
retail release date	<input type="checkbox"/>	<input type="checkbox"/>
packaging information (i.e., what is included in the game package)	<input type="checkbox"/>	<input type="checkbox"/>
number of players	<input type="checkbox"/>	<input type="checkbox"/>
rating (i.e., ESRB)	<input type="checkbox"/>	<input type="checkbox"/>
purpose (e.g., education, entertainment)	<input type="checkbox"/>	<input type="checkbox"/>
customization options (e.g., for characters, levels)	<input type="checkbox"/>	<input type="checkbox"/>
difficulty levels	<input type="checkbox"/>	<input type="checkbox"/>
achievements/awards/trophies	<input type="checkbox"/>	<input type="checkbox"/>
region (e.g., North America (NTSC-U/C), Japan and Asia (NTSC-J))	<input type="checkbox"/>	<input type="checkbox"/>
language	<input type="checkbox"/>	<input type="checkbox"/>
identifier (i.e., unique identifiers for the games such as barcode, ASIN)	<input type="checkbox"/>	<input type="checkbox"/>
box art/cover	<input type="checkbox"/>	<input type="checkbox"/>
screenshots	<input type="checkbox"/>	<input type="checkbox"/>
trailers	<input type="checkbox"/>	<input type="checkbox"/>

gameplay videos	<input type="checkbox"/>	<input type="checkbox"/>
genre/gameplay (e.g., Action, Strategy, Puzzle)	<input type="checkbox"/>	<input type="checkbox"/>
style (i.e., more detailed categories of genre such as Platformer, Action RPG, Tower defense)	<input type="checkbox"/>	<input type="checkbox"/>
plot/narrative (e.g., Protagonist explores world, Characters prevent apocalypse)	<input type="checkbox"/>	<input type="checkbox"/>
theme (e.g, fantasy-princess, supernatural-zombies)	<input type="checkbox"/>	<input type="checkbox"/>
setting (e.g., nature, schools, modern)	<input type="checkbox"/>	<input type="checkbox"/>
mood/affect (e.g., dark, humorous, quirky)	<input type="checkbox"/>	<input type="checkbox"/>
temporal aspect (e.g., real-time, time travel, time manipulation)	<input type="checkbox"/>	<input type="checkbox"/>
presentation (e.g., 2D, 3D, side-scrolling)	<input type="checkbox"/>	<input type="checkbox"/>
point of view (e.g., first person, third person)	<input type="checkbox"/>	<input type="checkbox"/>
references to historical events (e.g., World War II, The American Civil War)	<input type="checkbox"/>	<input type="checkbox"/>
type of ending (e.g., new game+, multiple endings)	<input type="checkbox"/>	<input type="checkbox"/>
visual style (e.g., retro, anime/manga, realistic)	<input type="checkbox"/>	<input type="checkbox"/>

[Q22] Are there any other game-related information that is useful to you?
Please write your answer here:

List of Sample Games

Game Title	Platform	Developer
Agricola	iOS app (iPad)	Playdek, Inc.
Bakery Story	iOS app (iPhone)	TeamLava
BANG! [HD] the Official Video Game	iOS app (iPad)	SpinVector S.p.A.
Barry's Bad Night	Flash-based	danthemilk
Bioshock Infinite	Xbox 360	Irrational Games
Bubble Witch Saga	iOS app (iPhone)	King.com Limited
Bubsy In: Claws Encounters of the Furred Kind	Super Nintendo Entertainment System	Accolade
Bully: Scholarship Edition	Xbox 360	Rockstar Vancouver
Call of Duty: Black Ops II	Xbox 360	Treyarch
Candy Crush Saga	iOS app (iPhone)	King.com Limited
Castlevania: Order of Ecclesia	Nintendo DS	Konami
Catan HD	iOS app (iPad)	USM
Caylus	iOS app (iPad)	Big Daddy's Creations
Champions of Norrath: Return to Arms	PlayStation 2	Snowblind Studios
Commander Keen: Aliens Ate My Babysitter!	PC DOS	iD Software Inc.
Condemned: Criminal Origins	Xbox 360	Monolith Productions, Inc.
Dance Central	Xbox 360	Harmonix
Dead Space	Xbox 360	EA Redwood Shores
Jump Super Stars	Nintendo DS	Ganbarion
Donkey Kong Country	Super Nintendo	Rare LTD

Dots: A Game About Connecting	Android	Playdots, Inc.
Dragon Age II	PC Windows	BioWare
Dragon Ball Z: Legend of the Super Saiyan	Super Nintendo Entertainment System	Bandai
DragonGem	Android	ITREEGAMER
Digimon Adventure 02: D-Terminal	D-Terminal	Bandai
Dungeons & Dragons Online	PC download	Turbine, Inc.
Eternal Darkness: Sanity's Requiem	Nintendo GameCube	Silicon Knights
Fallout 3	Xbox 360	Bethesda
Final Fantasy Tactics: The War of the Lions for iPad	iOS app (iPad)	Square Enix Inc.
LINE Fluffy Diver	iOS (iPhone)	LINE Corporation
Gabriel Knight 3: Blood of the Sacred, Blood of the Damned	PC Windows	Sierra Studios
Ghost Trick: Phantom Detective	iOS app (iPad)	Capcom
Grid Game	Flash-based	Mark James
Inindo: Way of the Ninja	Super Nintendo	Koei
InSperation	Flash-based	Puzzle Lab
Juniper's Knot	PC download	Dischan Media
Kumo Lumo	iOS app (iPhone)	Chilingo
Mario Teaches Typing 2	PC Windows	Brainstorm
Mass Effect 2	Xbox 360	Bioware
Minesweeper Classic	Android	IT Benefit
Neverwinter	PC download	Cryptic Studios
No One Lives Forever 2: A Spy in H.A.R.M.'s Way	PC Windows	Monolith Productions, Inc.
Phoenix Wright: Ace Attorney Trilogy HD	iOS app (iPad)	Capcom
Physicus	PC Windows	Ruske & Pühretmaier Edutainment GmbH
Planetarian	iOS app (iPad)	VisualArts Co., Ltd.
Plants vs. Zombies HD	iOS app (iPad)	PopCap
Puyo Puyo Fever 2	Nintendo DS	Sega
Rayman Origins	PlayStation 3	Ubisoft Montpellier Studios
Reiner Knizia's Tigris & Euphrates	iOS app (iPad)	Codito Development Inc.
Resident Evil 6	Xbox 360	Capcom
Robot Unicorn Attack	Flash-based	[adult swim] games
The Sims FreePlay	Android	EA Swiss Sarl
1080° Snowboarding	Nintendo 64	Nintendo EAD
Solitaire	Android	Ken Magic
Spore	PC Windows	Maxis
System Shock 2	PC Windows	Irrational Games LLC
The Elder Scrolls V: Skyrim	Xbox 360	Bethesda
The Journey to Wild Divine: The Passage	PC Windows	The Wild Divine Project
The Legend of Zelda Skyward Sword	Wii	Nintendo EAD
Tiny Farm	iOS app (iPhone)	Com2uS USA, Inc.
Transformers: War for Cybertron	Xbox 360	High Moon Studios
Uncharted: Drake's Fortune	PlayStation 3	Naughty Dog
VidRhythm	iOS app (iPad)	Harmonix

WhizzBall!	Flash-based	Discovery Kids
Words With Friends Free	iOS app (iPhone)	Zynga Inc.

¹ <http://library.stanford.edu/news/2013/03/stanford-libraries-preserves-historical-software-collection-federal-agency>

² http://www.ims.gov/news/2013_ols_grant_announcement.aspx#CA

³ <http://meganwinget.com/videogame-preservation/>

⁴ <http://blog.ims.gov/?p=4091>

⁵ The survey asked for users' feedback on a total of 43 elements from REC set version 1.2 in which the elements <character types>, <character names>, and <controls> were removed based on feedback from the collaborative review

⁶ Nintendo Wi-Fi Connection service for Nintendo DS and Wii to end in May.

<https://www.nintendo.com/whatsnew/detail/vyWpoM6CB1e6FjW8NIY7bvzOrgBURhzw>

⁷ <http://wiki.teamfortress.com/wiki/Achievements>

⁸ [M]: Mandatory, [R]: Repeatable