A Conceptual Data Model and Schema for Curating Collections of Video Game Development Artifacts

Abstract

Jin Ha Lee at the University of Washington (UW) Information School, in collaboration with the Video Game History Foundation (VGHF), proposes a two-year research project designed to help memory institutions organize and preserve critically endangered media creation artifacts. While game preservation has gained popularity among academics and institutions dedicated to preserving our cultural heritage, few have yet to focus on development artifacts, such as artwork, game design documents, musical scores, test builds, and marketing materials. As with other media, development artifacts are vital for understanding and appreciating the history and context of produced works. To aid those attempting to preserve these endangered materials, our research team will produce and test a conceptual data model and metadata schema for representing artifacts related to the development of video games. The primary research questions in this project addresses are:

> **RQ1.** Which information entities, attributes, roles, and relationships are relevant and important to describe the artifacts created during the development of video games and interactive media?

> **RQ2.** Which information entities, attributes, roles, and relationships are perceived as valuable by various user groups/stakeholders (including librarians, museum curators, game researchers, students, and users of library and museums) interested in game development artifacts?

> **RQ3.** What issues and challenges emerge as we apply our model and schema to describe existing collections?

The project outcomes include the conceptual data model, metadata schema, related controlled vocabularies, metadata crosswalks, and best practices document with instructions on how to adopt our work in various contexts. The main contributions of this research are: 1) contributing to a theorized understanding of how to represent complex media objects, 2) enabling curators and catalogers to effectively describe and represent game related materials through standard descriptions, and 3) extending and complementing existing standards for organizing and representing popular cultural objects.

To this end, this research project consists of two major phases that build upon our previous work.

In **Phase I**, we will create a conceptual data model and metadata schema. We will combine a top-down modeling approach involving archivists and industry professionals with a bottom-up modeling approach based on extant and potential descriptive aspects of these artifacts. Our collaborators from the VGHF will assist us in both approaches. At the same time, we will be conducting interviews of stakeholders representing different user groups to better understand their needs and searching behavior related to video game development artifacts, which will inform the design of our model. The conceptual data model will be expressed in Resource Description Framework (RDF) which provides interoperability among applications using machine-understandable information on the Web.

In **Phase II**, we will test the usefulness of the data model and schema by employing a mixed-methods approach involving real users. Activities in this phase will include a large-scale survey and practical application of the model on a subset of the VGHF collection. This will allow the research team to evaluate the entities and relationships included in our data model, providing insights into how it may be improved. This will ensure that our data model will ultimately be useful and understandable to a variety of potential user groups and stakeholders, including librarians, museum curators, game researchers, students, and users of libraries and museums.

In addition to improving our theoretical understanding of the domain, this work has the potential to advance the cataloging practice of born-digital artifacts and popular cultural artifacts within libraries and museums, and serve as a catalyst for advances in the use of games in education, science, and the humanities.

A Conceptual Data Model and Schema for Curating Collections of Video Game Development Artifacts

A team of researchers from the Information School at University of Washington (UW) and collaborators from the Video Game History Foundation (VGHF) request $254k for our 2-year research project to create a conceptual data model and metadata schema for curating collections of video game development artifacts. This project corresponds to the Curating Collections category of the National Leadership Grants for Libraries and will occur from June 1, 2018 to May 31, 2020. Video game development artifacts such as artwork, game design documents, musical scores, test builds, and marketing materials derived from the creation process of video games are vital for understanding and appreciating the history and context of games and interactive media. However, they have received little attention and support from academia and memory institutions and, as a result, they are critically endangered. This project will produce and evaluate a conceptual data model and metadata schema for representing artifacts related to the development of video games to complement ongoing research on video game metadata. By employing a mix of domain analysis and document analysis through a user-centered approach, we will identify information entities, attributes, roles, and relationships relevant to these materials for targeted user groups, including librarians, museum curators, game researchers, students, and users of libraries and museums. We will also evaluate our framework by applying it to existing collections at the VGHF, identifying issues and challenges, and addressing them in the revision process. Our model, schema, metadata crosswalks, and best practice documents will be published with instructions to help adopt our work in institutions with collections of games and game related materials. Our research findings will 1) improve our theoretical understanding of how to represent complex objects and their relationships within the domain of game development, 2) help institutions with such collections to represent and organize their materials, and 3) extend and complement existing standards for representing cultural objects in libraries, archives, and museums.

I. Statement of National Need: Preventing an Irreversible Loss of Media History

Today, more than ever, digital games play a widespread role in society. The Entertainment Software Association reports that more than 150 million Americans play video games and that the industry employs more than 220,000 Americans (ESA, 2017). Over 350 US colleges and universities offer degrees or coursework to prepare game industry professionals, and games are used in education, science, and engineering as learning and literacy tools. From friends bonding while playing *Rock Band* at a party to preschool children playing shape and color games on iPads; from local libraries hosting *Mario Kart* competitions to fans replaying *Dragon Age* to explore each intriguing branch of the narrative tree—video games and other digital interactive media are an embedded aspect of our social, cultural, and economic activities.

Like print publishing, film, and music, digital games are products of the cultural industries, that assortment of organizations which produce and distribute social symbolism. To properly understand the history of each of these media, we need to know the context and details of their creation. Draft manuscripts, research notebooks, and other related artifacts are often key to deeper understanding regarding the processes of symbol creators and the intentions designed into their works.

Unfortunately, unlike many of its counterparts in the cultural industries, digital games have received relatively little attention and support from academia, museums, libraries, and other institutions concerned with the study and preservation of culture and cultural objects. Though the first commercial digital home game console was released in the 1970s, only in recent years have we seen academia and memory institutions accelerate their acceptance of digital games as cultural objects. Radio, television, and film were also subject to this form of neglect in their infancy, and media historians and archivists have lost access to significant resources due to society's slow realization of the cultural impact of these media.

Fortunately, an increasing number of institutions are now collecting and providing access to digital games as part of our cultural heritage. For instance, the Strong National Museum of Play in New York collects and preserves electronic games of all types—arcade, console, computer, handheld, and mobile—that have enjoyed popularity over a sustained period and have influenced the video game industry or popular culture and

society in general (The Strong National Museum of Play, n.d.). Living Computers: Museum + Labs (LC:M+L) curates a large and growing collection of functional computer and game systems, as well as a collection of games that visitors can play on original machines. Internationally, Ritsumeikan Center for Game Studies at Ritsumeikan University, the National Media Museum in the UK, and the Computerspielemuseum in Berlin also work to catalog, classify, archive, and preserve digital games. The American Library Association (ALA) and its Games and Gaming Roundtable support and promote gaming in libraries via International Games Week (ALA, 2017) and a growing number of university libraries circulate video games for the purposes of academic inquiry.

Preserving digital games is no easy task. Early disregard by academics and memory institutions means that much the industry's early history has already been lost, lies in the hands of private collectors, or is boxed up in the closets and attics of players. More recently, the age of the Internet has shifted much of gaming to multiplayer experiences and online delivery, where game producers often do not provide "the game" in a physical package that is relatively easy to archive. The UK-based Digital Preservation Coalition (DPC) considers many forms of digital games to be *critically endangered* (Digital Preservation Coalition, n.d.), meaning "they face material technical challenges to preservation, there are no agencies responsible for them, or those agencies are unwilling or unable to meet preservation needs." Some of the *aggravating conditions* the DCP has identified for the preservation of digital games include:

- Dependency on gaming console;
- Loss or lack of documentation including uncertainty over intellectual property rights;
- Publisher no longer exists;
- Game withdrawn or replaced;
- Absence of collecting mandate from memory institution.

To date, game archival efforts at these institutions have focused largely on cataloging and preserving *final representations* of these cultural works. The majority of digital games that have been created during the industry's relatively short history are already no longer easily accessible for study and play. Even when we can access games, that playable code is only a part of the story. Researchers and memory institutions have given far less attention to **artifacts associated with game creation**, such as game design documents (GDDs), technical design documents, (TDDs), art bibles, style guides, musical scores, test builds, voice-over auditions, marketing materials, and company-specific archival items. One exception is Winget's work (Winget & Murray, 2009; Winget & Sampson, 2011), confirming the game development process does "produce significant and important documentation as traditionally conceived by collecting institutions (p.29)" but also finding that standard practices fail to adequately collect and preserve the full range of artifacts created. Building on this research, our proposed work aims to create a model and schema that can be used for comprehensively capturing information about those artifacts.

As with other products of the cultural industries, loss of historic development artifacts has a deleterious effect on our long-term ability to study and understand cultural objects. To understand how and why games were made, how they were advertised and sold, how they were received and consumed by players of their time, and what kind of social impact they made, historians and researchers rely on such materials to tell a complete story. In the same way that music development artifacts (e.g., alternate recordings, sheet music, studio photos) from the Abbey Road recording studio help us understand the Beatles' evolution as they created *Sgt. Pepper's Lonely Hearts Club Band* more clearly than just listening to the album, game creation artifacts help us understand interactive media far better than just playing games. It is also critical to preserve these materials to better educate and inspire the next generation of creators. Without an organized effort to collect, document, and preserve these materials, we risk losing them forever.

Some institutions have recognized this danger and have begun to correct course. For instance, in addition to collecting and preserving original hardware and software, the National Videogame Archive in Great Britain and the Strong National Museum of Play also collect artifacts from the creation process. However, to date, this effort is limited in scope and largely uncoordinated with other organizations involved in cultural preservation or digital game production. Furthermore, there is no standard to represent these materials in a way that is interoperable between these organizations.

In addition, many of the artifacts from the game creation process are born-digital themselves. Each DPC aggravating condition identified for digital games also applies to artifacts associated with their development, but additional conditions apply as well, such as:

- Conflation of bit preservation with logical preservation;
- Dependence on exotic or obsolete or proprietary formats, processes, or devices;
- Inappropriate dependence on service provider;
- Lack of awareness, concern, or planning from creators and corporate owners;
- Poor storage conditions (temperature, humidity, dust, electromagnetic effects)

While few are currently aware of these challenges, many will eventually recognize the loss, from curious fans to students and historians of the industry, media, and culture, as well as the librarians, museum professionals, and archivists at the memory organizations who serve those individuals. Right now, before the loss is felt more acutely, we can begin to address this underappreciated public need by building a shared conceptual data model and metadata schema for representing and organizing these endangered artifacts to help preserve historical information about one of America's greatest contributions to popular culture.

II. Project Design

A. Research Objectives
To overcome these challenges and aid those attempting to preserve endangered artifacts so vital to our understanding of interactive media, our research team will produce and test a conceptual data model and metadata schema for representing artifacts related to the development of video games to complement ongoing research on video game metadata. In order to best create our model and schema, we will answer three primary research questions:

> **RQ1.** Which information entities, attributes, roles, and relationships are relevant and important to describe the artifacts created during the development of video games and interactive media?
> **RQ2.** Which information entities, attributes, roles, and relationships are perceived as valuable by various user groups/stakeholders (including librarians, museum curators, game researchers, students, and users of library and museums) interested in game development artifacts?
> **RQ3.** What issues and challenges emerge as we apply our model and schema to describe existing collections?

The research findings will improve our theoretical understanding of the entities and their relationships within the domain of game development. We will ground the research and field test our findings by developing metadata records for the collection owned by the VGHF and museums/libraries/archives with similar materials in their collections. The project outcomes include the conceptual data model, metadata schema, related controlled vocabularies, metadata crosswalks, and best practices document. These will benefit curators and archivists who will be able to adopt some or all of our work for the benefit of the players, collectors, academics, game industry professionals, and game development organizations.

B. Contributions
(1) Develop theory regarding the representation of complex media objects. The main scholarly contribution of our work is a theorized understanding of how to best represent the entities and relationships in the domain of video game and interactive media development. However, many other forms of electronic, born-digital, and interactive media (e.g., e-books, computer software, digital images, and smartphone applications) can benefit from research on non-book metadata. Inquiry into the differences in metadata needs for physical and digital artifacts speaks to larger questions about the transition from physical to digital materials and the implications of that transition for libraries, archives, and museums. Our work may also be applicable to other segments of the cultural industries, such as film and animation, which are also created from a complex development process where many intermediate development artifacts are produced.

(2) Better organize and represent video game collections in museums and libraries. Beyond academic research, our conceptual data model and metadata schema will enable catalogers to describe video games and related materials more accurately and thoroughly, improving the quality of metadata shared among organizations and used to locate materials. Standardized and thorough descriptions of these materials affords improved access to museum curators, archivists, and librarians who acquire, catalog, and provide reference services; users of these collections; and commercial organizations trying to organize and preserve their own development assets. As an increasing number of organizations adopt this new open standard, it will support interoperability among institutions, facilitating new consortia and resource sharing opportunities. As research partners, the VGHF is committed to adopt the work.

(3) Extend and complement existing standards for organization and representation. This research complements and extends our prior work with the Video Game Metadata Schema (VGMS) (Game Research Group, 2016; Lee et al., 2017), the conceptual model for video games and interactive media (Jett et al., 2016), and research from GAMECIP research team (De Groat et al., 2015;) as well as extending and complementing other standards. Our team will examine Encoded Archival Description (EAD; "a non-proprietary de facto standard for the encoding of finding aids for use in a networked (online) environment" (The Library of Congress, n.d.)), Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images (CCO; "a published manual for describing, documenting, and cataloging cultural works and their visual surrogates" (Baca et al., 2006)), Resource Description and Access (RDA; "a package of data elements, guidelines, and instructions for creating library and cultural heritage resource metadata that are well-formed according to international models for user-focused linked data applications" (CLA, 2010)), and the Conceptual Reference Model from the Comité International pour la Documentation at the International Council of Museums (CDOC CRM; "provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation" (ICOM/CIDOC Documentation Standards Group, 2011)). In addition to the conceptual data model and the metadata schema, we will provide metadata crosswalks and best practices document explaining how our work can be translated and adopted in existing standards. This will allow maximum flexibility in applying our work in a variety of contexts, including full adoption (the VGHF) and partial adoption to complement an existing cataloging framework (e.g., libraries with games cataloged in RDA and MARC (MAchine-Readable Catalog) format).

C. Relevant Efforts at Other Institutions

The Preserving Virtual Worlds project was a joint effort by Rochester Institute of Technology, Stanford University, the University of Maryland, and the University of Illinois at Urbana-Champaign, and was supported by the Library of Congress. This project focused on preserving older video games and software, and establishing best practices and strategies for game preservation. The second phase of the project, funded by IMLS, focused on determining significant properties for educational games. 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., 2010). Responding to this call, the conceptual model for video games and interactive media, VGMS, and seven related controlled vocabularies were created by the UW Game Research (GAMER) Group and Seattle Interactive Media Museum in 2012 (Lee et al., 2013a; Lee et al., 2013b). GAMECIP, led by the University of California, Santa Cruz Library, UCSC Computer Science, and Stanford University Library, also investigated metadata needs and citation practices surrounding computer games in institutional collections, producing a schema and best practices to improve library and institutional practice for computer games (De Groat et al., 2015; The Online Audiovisual Catalogers, 2015). However, both of these projects mainly focused on the end product: final representations rather than materials derived from the creation process. Our work will expand these schemas by addressing excluded but relevant and important materials.

Specifically related to game development materials, Megan Winget led an IMLS-supported project studying the creation process and artifacts of the video game industry by observing and interviewing game developers (Winget & Murray, 2008; Winget & Sampson, 2011). Our project outcomes will extend Winget's work, describing and representing the types of information objects that Winget identified, and emphasizing their relationships to user experiences with games.

D. Preliminary Research

This year, we have built a foundation for this project through preliminary explorations of the domain. The UW team, in collaboration with Kelsey Lewin from the VGHF, surveyed video game professionals, researchers, students, and fans on their perceptions and personal challenges in exploring and discovering these kinds of artifacts. The survey explored the general attitudes and relationships people have with these artifacts. Questions were designed to discover users' goals when searching the game development artifact collection as well as the metadata and access points that help them achieve their goals. We collected 526 responses which are currently under analysis. Most respondents expressed significant interests in these artifacts—including concept artwork, unfinished builds, and press releases—but agreed that accessibility is difficult because materials are spread across a large number of locations rather than collected in any one portal or database. The majority of participants wanted to examine these materials to learn their historical context (82.7%), to better understand the game development and marketing process (80.2%), and to find inspiration for their own work (45.9%). Many also expressed frustration that files from most resources are generally not well-labeled or given any attribution, making discovery nearly impossible.

Preliminary findings provided insight into those people interested in game assets and ephemera collections; their expectations for searching, browsing, and interacting with such collections; the kinds of materials they expect to find; and their intentions. This early work facilitates deeper exploration with those likely to use such an archive. The proposed project continues our investigations with relevant stakeholders via semi-structured interviews, which will provide much needed depth regarding their information needs, wants, requests, and other behaviors related to these materials.

E. Schedule and Outcomes

We propose a 2-year research project with the first year focusing on the creation of the model and schema and the second focusing on the evaluation and refinement of the work.

Year 1: We will focus on cooperative development of the model and schema, combining top-down (consulting domain experts) and bottom-up (analyzing existing data structures and records of game artifacts) modeling. The principal investigator (PI) has an established history of user-centered knowledge organization research, and the research team will use that expertise to work with target users to identify their needs, explore their practices regarding game development artifacts, and map the way in which they conceptualize the domain. The concurrent bottom-up approach will inform the data model by examining working archives, data structures, metadata records, and organizational systems to understand how the domain is currently being represented and explore the ways in which it could be represented in the future. This approach will identify attributes of artifacts not currently recorded in extant systems yet critical to understanding the complex relationships in the domain. The entities and relationships identified in the bottom-up approach will inform the model established by the top-down approach. Our collaborators from the VGHF will fully participate in both approaches. Concurrently, we will conduct in-depth, semi-structured interviews with museum curators, librarians, archivists, game researchers, students, and library and museum patrons. Extending our preliminary work with more specific inquiries, we will explore how these different user groups perceive and express as their needs for accessing game development artifacts, and how they expect to find the materials they need in each particular context. This user-centered approach will ensure that our produced outcome will be meaningful and relevant to these different stakeholders. The conceptual model will be expressed in Resource Description Framework (RDF) which provides interoperability among applications using machine-understandable information on the Web. The advisory board will kick off the project with an in-person meeting in summer 2018. The board will also review our progress with the model/schema, considering the user data we obtained via interviews, and provide feedback in an online meeting in winter 2019.

Year 2: We will test and refine the model through targeted user surveys and practical cataloging of sample materials—varied by type, time period, platform, and genre—from the VGHF collection. These tests will allow us to evaluate the entities and relationships in our conceptual model, as well as provide insights on possible improvements, based on the following criteria:

1. Accuracy (are the entities and relationships accurately represented?);
2. Understandability (is it intuitive and does it explain the terms in a straightforward way?);
3. Extensibility (is the model easily extensible without requiring major restructuring?);
4. Conciseness (is the model free of redundant information features and relationships?);
5. Learning capabilities (does the model help users understand the domain better?);
6. Usability for users (is the model convenient and useful for end users?); and
7. Usability for catalogers (is the model convenient and useful for catalogers?).

Based on the user data and test cataloging experience, we will revise the model and schema to maximize the applicability of our work in real-life cataloging contexts. The research team has successfully employed a similar approach while creating and testing the conceptual model for video games and VGMS, and are very familiar with the methodologies involved in each activity. In addition to budgeted resources, the effort will include volunteer graduate student researchers using the methods, processes, and lessons that made VGMS development successful (e.g., graduate courses and Capstone projects). We will hold an online meeting for the advisory board in autumn 2020 to review our proposed revisions to our model/schema based on Phase II data and provide feedback. The board will conduct a final in-person project wrap-up meeting in spring 2020.

The following table summarizes our research project plan. The grant period will run from June 1, 2018 to May 31, 2020. More detailed information about the project timeline is provided in the attached *Schedule of Completion*.

Timeline	Goals	Methods	Activities	Outcomes
Phase I (2018-2019)	Identify relevant information entities, attributes, and relationships for game development artifacts via domain analysis and document analysis [Answering RQ1]	Top-down modeling	Establish a logical data model by consulting a diverse group of game experts and enthusiasts with deep understanding of the domain	A conceptual data model and the metadata schema will be established, reflecting 1) domain experts' view, 2) structure and content of the materials themselves, and 3) video game development artifacts related information needs, uses, and search behaviors of potential users.
		Bottom-up modeling	Further refine the model by examining video game development artifacts as information objects and analyzing existing data structures and records of these artifacts	
	Understand the perceived value of different information entities, attributes, and relationships for game development artifacts by various user groups [Answering RQ2]	User interview	Conduct in-depth interviews of 20-30 users representing different user groups (i.e., developers/designers, museum curators, librarians, game researchers/students, fans) to help us understand their information needs and searching behavior as well as identify potential metadata elements	
Phase II (2019-2020)	Verify the accuracy of the representation of users' perception in our model and schema [Answering RQ2]	Survey	Conduct a large-scale survey (300-400 responses) for evaluation of the proposed conceptual data model and metadata schema	The conceptual data model and metadata schema will be revised based on 1) the feedback received from various stakeholders, and 2) the cataloging experience. Best practices documents and metadata crosswalks will also be created based on the final version of the model/schema.
	Catalog a test collection of video game development artifacts [Answering RQ3]	Usability testing	Evaluate the usability of our data model and metadata schema by cataloging a sample video game collection	

F. Research Team and Advisory Board

The core research team consists of five members: two from the UW iSchool and three from the VGHF. We have assembled a project team with exemplary skills, experience, and accomplishments related to the proposed work. As a group, the research team are constituents of the researcher, archivist, industry professional, and player populations that the work targets. The commitments that core team members have made to the project will ensure that our project outcomes are of the highest quality and will be adopted by libraries, archives, and museums with collections of video games, game assets, and ephemera.

Principal Investigator: Dr. Jin Ha Lee will serve as the PI for this proposed research. Dr. Lee is an Associate Processor, MLIS Program Chair, and Director of the GAMER Group at the UW Information School (iSchool). Her research focuses on providing better organization of and access to video games and interactive media, understanding user behavior related to games, and using games for informal learning. The PI and her research team at the UW have extensive experience working with video game metadata and establishing standards related to describing games. They created the conceptual model for video games, the VGMS, and seven relevant controlled vocabularies in collaboration with the Seattle Interactive Media Museum. Parts of the VGMS and its controlled vocabularies have been adopted by Michigan State University (MSU) and LC:M+L for cataloging their video game collections. The VGMS and its controlled vocabularies are also published in the Open Metadata Registry (OMR) (Open Metadata Registry, n.d.), and Gameplay Genre and Narrative Genre terms were assigned Library of Congress Course Codes (vgmsgg, vgmsng) so they can be used in library catalogs. Based on the VGMS genre vocabulary, her team is currently working with the Center for Game Studies at Ritsumeikan University in Japan and Computerspielemuseum in Germany to build a multilingual taxonomy for video game genres (in English, Japanese, and German) to be promoted as an international standard. The conceptual model for video games established by the GAMER Group (Jett et al., 2016) informed the design of the MediaDB created in Ritsumeikan University's initiative to organize and provide access to popular cultural objects from Japan, including video games, anime, and manga. The PI will function as the overall project manager, coordinating all activities with our research partners, advisory board members, and study participants.

Student: Marc Schmalz, a Doctoral Student at the UW iSchool, member of the GAMER Group, and game industry professional, will support the PI with all her responsibilities.

Research Partners: Three staff members from the VGHF will work with the PI and student researcher. These are Frank Cifaldi (Founder and Director, VGHF), Kelsey Lewin (Communications Director, VGHF; Co-Owner, Pink Gorilla Games), and Travis Brown (Technical Director, VGHF). Together, they have led a number of projects related to video game preservation, categorization, and education. They have extensive experience handling, analyzing, and evaluating video game artifacts for a number of institutions, including the Smithsonian, the Library of Congress, the National Videogame Museum, and the Strong National Museum of Play. The VGHF and its leading members often act as agents between potential donors and the academic, government, or charitable institutions that collect video game artifacts, providing expertise on both the artifacts and the best practices for preserving them. They provide research aid to educational projects such as documentaries, academic papers, and pop-up exhibits. Currently, they are working with Game Informer to organize, preserve, and digitize a collection of video game ephemera, including over 15,000 optical discs and hundreds of prototype game builds dating back to the early 1990s. For the proposed project, the VGHF staff will offer 1) their domain expertise, 2) their extensive network of contacts which can be used for identifying participants from diverse user groups, 3) their current collection with concrete samples for establishing the model/schema as well as testing and evaluation, and 4) their technical expertise which will allow us to implement our designs for their collection.

Advisory Board: To best pursue our goals, we have formed an advisory board consisting of members of our target user populations who will oversee all project activities and guide our work. Members include Andrew Borman, Digital Games Curator at the Strong National Museum of Play; Henry Lowood, Curator for History of Science & Technology Collections and Film & Media Collections at Stanford University Libraries; Jerome McDonough: Associate Professor at the School of Information Sciences, University of Illinois at Urbana-Champaign; Cynde Moya: Collections Manager at Living Computers: Museum + Labs; Dr. Laine Nooney: Assistant Professor of Media and Information Industries at New York University, and Shana Bryant: Senior

Producer at Private Division/Take Two Interactive. These advisors are national leaders in creating, organizing, preserving, and providing access to video games and interactive media. The IMLS has supported both Dr. Lowood and Dr. McDonough's work on organizing and preserving video games and interactive media (Preserving Virtual Worlds II project (LG-06-10-0160-10) and GAMECIP project (LG-06-13-0205-13)).

G. Budget

We are requesting funds for: 1) summer salary and benefits for the PI; 2) support for the graduate assistant, including tuition, stipend, benefits, and fees; 3) travel support for the PI, graduate assistant, and research partners; 4) honoraria and travel/meeting support for advisory board members, and 5) human subjects payment for study participants. While much of our work with participants can be performed via teleconferencing software, we have budgeted for some travel in order to directly work with collections of physical artifacts archived in remote locations, and to disseminate our findings at conferences and conventions. Other resources are available at no additional charge through the UW, such as Internet access, teleconferencing software, and web hosting services. Any required web content will be shared through the existing GAMER Group WordPress site, which is created and managed by the GAMER Group and hosted by the UW. The schema and controlled vocabularies will be published in the OMR to maximize the shareability. Full details of expected expenses are available in the attached *Budget Justification*.

In addition to compensated time, the project also relies on the voluntary unpaid participation of students performing coursework for testing the model and schema in Year 2 in order to maximize productivity without significantly increasing the cost. At the UW iSchool, courses like *Metadata for Interactive Media* and *Special Topics in Information Studies* are well-established ways of involving undergraduates and masters students in the GAMER Group's research, and students have also pursued independent studies and Capstone projects involving game research in the recent past. This benefits not only the research team by increasing productivity, but also the students as they gain practical experience working with metadata to balance the theoretical knowledge from core courses on information organization, and pursue opportunities to work in local organizations, as happened when UW iSchool students worked with LC:M+L to organize their game and software collections. The UW iSchool also involves many of its most research-minded and driven students through volunteer research. Students who worked with the PI and the GAMER Group have successfully conducted multiple research projects on various aspects of organizing video games and published their works (GAMER Group, n.d.).

Our third research question addresses a type of risk regarding the project: "What issues and challenges emerge as we apply our model and schema to describe existing collections?" While the user-centered design strategy will mitigate some of these concerns, the research team knows from experience that no metadata schema is perfect for all audiences. We *will* encounter issues when curators and archivists use our research for practical purposes for a wide variety of collections, but our team has a commitment to maintain and improve this work beyond the term of the grant, as we do with the VGMS which is reviewed and revised regularly.

III. Diversity Plan

Minority representation in gaming is important, both in terms of in-game characters and in-studio professionals. The industry has a deserved reputation for being comprised of mostly young Caucasian males, and a tendency through much of its history for authoring games that appeal to the same. The importance is highlighted by the ongoing Gamergate harassment campaign carried out by a regressive faction of gamers seeking to suppress diverse voices and representation in the industry. The GAMER Group is acutely aware of these diversity concerns in the domain, and is committed to gaming for everyone.

The GAMER Group has a track record of supporting inclusive gaming and exploring representation within games. For example, students have recently developed a controlled vocabulary for game protagonists which, when implemented, affords the opportunity to study representation in games and the ability to find characters like oneself represented in interactive media. Additionally, the group's ongoing game accessibility metadata project is working in consultation with the International Game Developers Association (IGDA) to create a standard which can be used to provide accessibility information about a game prior to purchase. In the way that Entertainment Software Ratings Board (ESRB) ratings provide guidance on content, this new standard

would provide guidance on accessibility, helping an underserved portion of the gaming community find appropriate content.

The core research team and the advisory board for this proposal are carefully put together to include people with diverse gender, race, and perspectives. We will continue the GAMER Group's history of inclusive research by seeking out diverse participants for research and evaluation, as well, at both the individual and organizational levels, relying on sources such as Gaymer X, AbleGamers, I Need Diverse Games, GeekGirlCon, and IGDA Special Interest Groups (e.g., Women in Games, Blacks in Games).

IV. National Impact

The team's ultimate goal is to transform the way memory institutions handle game-related artifacts, stopping their loss and making them available to future students, patrons, and game workers. The grant period sets the foundation for this important long-term work.

The practical benefits of the project will be immediately felt at the VGHF, where the conceptual model and schema will first be implemented, allowing them to process and preserve their collection of game-related materials. Members of the advisory board will also be well suited for early adoption of the conceptual model and schema, and their participation is certainly an indication of interest for their organizations. We hope our industry outreach via organizations like IGDA and game conventions like Penny Arcade Expo (PAX) will help professionals better organize development artifacts, making them more effective and efficient in their archiving practices, and affording them easier handoff to memory institutions. Ultimately, adoption by memory institutions will bring improved access for all of our target user populations.

We will ensure our project outcomes are readily adaptable by conducting user-focused research at each step and seeking the council of participants and the advisory board in each phase. As explained above, the conceptual data model and schema will be expressed in RDF to ensure the interoperability among applications using machine-understandable information on the Web and will be published in the OMR that is freely accessible. In addition, our deliverables include crosswalks and best practices document—tools which maps our schema onto existing institutional standards such as CIDOC CRM and FRBR (IFLA Study Group, 2009)— assisting others with implementing and adapting our work to their institutional and personal needs.

The GAMER Group and the VGHF are committed to crafting our research for maximum shareability, accessibility, and usability, and have a record of having done so with the VGMS. All official releases and published academic papers related to the development of the VGMS and its controlled vocabularies are posted on the GAMER Group website, hosted by the UW. The conceptual model, schema, and CVs are also freely available through the OMR. Project outcomes from this proposed research will also be made available through the GAMER Group website. This is a sustainable solution because the UW and iSchool are committed to supporting our lab. Project materials will be hosted indefinitely by the UW in the same way it has supported our distribution of work related to the VGMS. Materials will also be distributed through the VGHF website.

A. Communication Plan

(1) Dissemination of Research Findings. Research findings will be disseminated in traditional academic formats, such as journal publications (*Library and Information Science Research*, *Cataloging & Classification Quarterly*, and *International Journal on Digital Libraries*) and research conference proceedings (ASIS&T, JCDL, and the iConference). Practical findings and project deliverables will be shared with the library community through roundtables, forums, and pre-conference training sessions, such as the ALA Annual and Midwinter conferences. Other archival meetings, like the Society for American Archivists' Annual Meeting or museum conferences like Museums and the Web, offer opportunities for less formal discussion, such as roundtables, where we can convey lessons learned and best practices while sharing our findings. This information may be useful to other groups dealing with non-traditional materials such as emergent collections of digital ephemera or e-books.

In addition to academic venues, our research findings will also be offered to professional gatherings. Our research team has a history of participating in the PAX series of game conventions, and developers' conferences such as PAX Dev and the Game Developer's Conference (GDC) also provide opportunities to engage with industry professionals.

By making our research products freely accessible online and by actively promoting our work in relevant venues, we expect that organizations with video game collections will utilize the schema—saving staff time, effort, and financial resources—rather than developing home-grown systems. With all components of the research outcomes online, an organization is free to select components most appropriate for its own use. Under the Creative Commons license, many institutions who might not otherwise consider collecting and cataloging materials related to the production of video games will be more inclined to do so, thus helping archives and museums increase their collections in areas of patron need.

(2) **Transparency and Accessibility of the Research Products.** A full report on our conceptual data model and metadata schema, including the definitions of entities and relationships, controlled vocabularies, crosswalk and best practice guides with instructions for use, will be openly published on the GAMER Group and VGHF websites, as well as any metadata generated for the sample collection and any crosswalks for interoperability with other standards. All the information will be freely available under a Creative Commons Attribution-ShareAlike 3.0 Unported License. We plan to implement the work on the OMR as well. Any library, museum, archives, studio, or game professional with an Internet connection will have access to the schema for review and use. The PI and the GAMER Group, as well as VGHF, will also be available for online consultation for providing guidance on any institutions who are interested in adopting our research outcomes, as we have supported MSU, LC:M+L., Ritsumeikan Center for Game Studies, and Computerspielemuseum.

(3) **Targeting Our Findings/Products.** As participants within the game preservation community and with connections to industry, we have multiple means of informing our intended audience about our work. Listservs are still a popular communication channel in library communities, and we expect to publicize many of our publications and presentations on relevant Listservs. Participants from our previous research studies often request to be informed of subsequent findings, so we have established an email distribution list of interested parties. We can also directly contact libraries, archives and museums known to have video game and related collections. In addition to email, social media outlets like Facebook, Twitter, and Discord have the potential to reach a wider variety and different demographics of library, museum and archives personnel. Social media and blogging activities by the GAMER Group, the VGHF, and the UW iSchool can highlight the project at various stages, drawing attention to our work.

B. Sustainability Plan

(1) **Ongoing Institutional Support of Project Activities and Products.** The VGHF, GAMER Group, and UW iSchool were committed to this project before the grant opportunity and will continue to support it after the grant period ends. Several UW iSchool classes (*Metadata for Interactive Media*, *Special Topics in Information Studies*) are well-established and offered every year, and both undergraduates and masters students have pursued game-related independent study and Capstone projects in the recent past.

All published work will be available on both the VGHF and GAMER Group websites indefinitely. As an established organization, the VGHF is dedicated to continuing its mission of cataloging, preserving, exhibiting video games, interactive media and related physical, digital and abstract artifacts for educational and research purposes. Cataloging of these materials and the requisite metadata work that is necessary for rich and accurate description will continue after the grant ends.

The UW has been committed to the iSchool for more than 100 years. University resources regularly support websites and other online resources in conjunction with research projects, such as Voices From the Rwanda Tribunal, a collection designed to make primary source materials surrounding the International Criminal Tribunal for Rwanda freely and openly accessible to people of Rwanda and the world. Multi-lifespan accessibility is a key component of this project, and this example demonstrates the commitment of personnel and resources at the UW to online longevity.

(2) **Continuing Benefits Resulting from the Project.** Our conceptual data model and metadata schema are but two steps along a long research path at the UW iSchool. Insights gleaned from our work will not only help us better understand the domain of video games, but such ontological understanding underlies all future attempts to increase discoverability and access for games and interactive media, and for development artifacts related to those media. Following the creation of standards during this period of this grant, the GAMER Group hopes to implement those standards and establish online repositories for game-related metadata.

Schedule of Completion

PHASE	ACTIVITY	2018-2019				2019-2020			
		Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring
Phase I: Creation	Kick-off in-person meeting with the advisory board	■							
	Identify and recruit domain experts	■	■						
	Identify and consult with domain experts (Top-down modeling)	■	■						
	Identify and examine existing documents, data structure, and relevant standards (Bottom-up modeling)	■	■						
	Develop interview instrument, pretest	■							
	Get an IRB approval for the interview	■							
	Conduct user interviews	■							
	Analysis of interview data		■	■					
	Develop model/schema (including controlled vocabularies)	■	■	■	■				
	First online review meeting with the advisory board			■					
	Finalize the initial version of model/schema to be tested				■				
Phase II: Evaluation	Develop survey instrument to evaluate model/schema, pretest					■			
	Get an IRB approval for the survey					■			
	Launch survey and collect data					■	■		
	Analysis of survey data						■		
	Second online review meeting with the advisory board						■		
	Test catalog sample VGHF collection with the model/schema						■	■	
	Revise the model/schema based on user data and test cataloging experience						■	■	■
	Final in-person wrap-up meeting with the advisory board								■
Multiple Phases	UW Course: Metadata for Interactive Media		■				■		
	UW Course: Capstone			■				■	■
	UW Course: Directed Fieldwork	■	■			■	■		
	Dissemination via conferences, journals, and other events		■	■	■	■	■	■	■

Note: Shaded bars represent an activity's most intensive period; in practice, activities may begin before and extend beyond

DIGITAL PRODUCT FORM

Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

☐ Please check here if you have reviewed Parts I, II, III, and IV below and you have determined that your proposal does NOT involve the creation of digital products (i.e., digital content, resources, assets, software, or datasets). You must still submit this Digital Product Form with your proposal even if you check this box, because this Digital Product Form is a Required Document.

If you ARE creating digital products, you must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

Part I: Intellectual Property Rights and Permissions

A.1 What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

All digital products we intend to create will be licensed under a Creative Commons Attribution 3.0 - Noncommercial U.S. License. We will explain to potential users they are free to use, redistribute, and modify any of these resources as long as they attribute the source.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

The ownership rights our organization will assert over new digital products will focus on attribution and appropriate credit. Our digital products will be for open access; we only need appropriate credit. We will notify potential users about relevant terms or conditions by placing a human-readable summary of the Creative Commons Attribution 3.0 U.S. license in our materials.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

N/A

Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

A. Creating or Collecting New Digital Content, Resources, or Assets

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

We will be producing the following digital content: 1) Conceptual data model for representing video game development artifacts, 2) Metadata schema for describing the artifacts, 3) Relevant controlled vocabularies (CVs) to be used with the schema, 4) Metadata crosswalks mapping our schema and other existing standards for describing cultural objects, and 5) Best practices document with instructions on how to adopt our work. We will produce Microsoft Word documents for all digital products. In addition, the model, metadata schema and CVs will be expressed in RDF form and be published in Open Metadata Registry (http://metadataregistry.org/).

A.2 List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

All the hardware and software necessary for this research are provided by the University of Washington Information School. The Project Director (PD) and the graduate student researcher will be using workstations provided by the UW Information School. We will use NVivo for coding our data including audio recordings of the interviews, transcripts, and notes. The Information School license Oxygen XML (XML editor) that can be used for encoding the metadata records. We will use Google Team Drive to save and share any project data among the research team members.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

XLSX (Excel spreadsheet for metadata crosswalks), DOCX (Word documents for all digital products and interview transcripts), MPEG4 (audio recordings of interview, 256kbps), XML documents (metadata schema and CVs in the Open Metadata Registry). Sample data may be shared as TXT or CSV formats.

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

Our digital products will go through an iterative design process involving the core research team as well as the advisory board in order to ensure the quality of the outcomes. In addition, via the user survey, we will be able to collect feedback on our model, schema, and CVs which will be used for revising our outcomes. We will also test catalog video game development artifacts from the Video Game History Foundation (VGHF)'s collection and based on that experience, refine our work. The types of video game development artifacts will change over time, just as video games do. Therefore in order to maintain the relevance of our project outcomes, after we initially publish them, we will regularly review and revise them, just as we have been doing for our model, schema, and CVs for video games (http://gamer.ischool.uw.edu/official_release/).

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

Our digital assets will be shared in perpetuity through public-facing websites provided by the UW iSchool, just as we have been doing for our model, schema, and CVs for video games (http://gamer.ischool.uw.edu/official_release/). The VGHF will share them via their website as well. Copies will be provided through the Open Metadata Registry.

C. Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore,

PREMIS) and metadata content (e.g., thesauri).

One of the main goals of the project is to create a metadata schema for video game development artifacts and therefore, the sample game development artifacts will be described using our own project outcome. We will be creating metadata crosswalks by mapping our schema to other existing standards for representing and describing popular cultural objects in multiple contexts including: MARC Machine-Readable Cataloging (MARC) and Resource Description and Access (RDA), Encoded Archival Description (EAD), Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images (CCO), and the Conceptual Reference Model from the Comité International pour la Documentation at the International Council of Museums (CIDOC CRM).

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

Metadata generated through test cataloging of the VGHF collection will be created in Microsoft Excel and shared as XLSX, TXT, or CSV files. The metadata records will be maintained in the GAMER group's server space provided by the University of Washington as well as in VGHF.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

We will be using the websites of Game Research Group and VGHF, as well as major social media platforms such as Twitter and Facebook and various listservs and slack channels relevant to libraries (e.g., ALA Games and Gaming Round Table) to promote our work to facilitate widespread discovery and use of our work.

D. Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

Our digital assets will be shared in perpetuity through public-facing websites provided by the UW iSchool, just as we have been doing for our model, schema, and CVs for video games (http://gamer.ischool.uw.edu/official_release/). The VGHF will share copies via their website as well. Copies will be provided through the OMR as long as that site allows. All of these sources are expected to remain public for their lifetimes. All of the digital products should be accessible via standard web browsers.

D.2 Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

- Video Game Metadata Schema, Controlled Vocabularies (CVs), and related documentations: http://gamer.ischool.uw.edu/official_release/
- Video Game Metadata Schema: http://metadataregistry.org/schema/show/id/132.html
- Gameplay Genre CV: http://metadataregistry.org/vocabulary/show/id/381.html;
- Narrative Genre CV: http://metadataregistry.org/vocabulary/show/id/432.html;
- Mood CV: http://metadataregistry.org/vocabulary/show/id/379.html;
- Theme CV: http://metadataregistry.org/vocabulary/show/id/426.html;
- Setting – Place CV: http://metadataregistry.org/vocabulary/show/id/424.html;
- Setting – Time period CV: http://metadataregistry.org/vocabulary/show/id/423.html;
- Setting – World CV: http://metadataregistry.org/vocabulary/show/id/422.html;

- Visual Style – Style CV: http://metadataregistry.org/vocabulary/show/id/430.html;
- Visual Style – Color CV: http://metadataregistry.org/vocabulary/show/id/428.html;
- Visual Style – Light CV: http://metadataregistry.org/vocabulary/show/id/431.html;
- Narrative Tropes CV: http://metadataregistry.org/vocabulary/show/id/429.html;
- Character Tropes CV: http://metadataregistry.org/vocabulary/show/id/427.html;
- Library of Congress Genre/Form Source Codes for VGMS Gameplay Genre CV and Narrative Genre CV: https://www.loc.gov/standards/sourcelist/genre-form.html;
- Links to Publications and Presentations: http://gamer.ischool.uw.edu/publications/

Part III. Projects Developing Software

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

N/A

A.2 List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

N/A

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

N/A

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

N/A

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

N/A

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

N/A

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.

N/A

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify

any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

N/A

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

N/A

C.3 Identify where you will deposit the source code for the software you intend to develop:

N/A

Name of publicly accessible source code repository:

URL:

Part IV: Projects Creating Datasets

A.1 Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

In Phase I of the project, we will be generating research notes from our consulting sessions with game domain experts during our top-down modeling process, and also during our bottom-up modeling process, as we identify and record what kinds of video game development artifacts exist and how they are currently being organized. These activities will happen in Autumn and Winter quarters of 2018-2019 academic year. In addition, we will be collecting interview data from users representing different user groups of video game development artifacts. The interview data will be collected in Autumn 2018 and Winter 2019, and the data analysis will happen in Winter-Spring 2019. The notes and interview data will be coded using NVivo.

In Phase II of the project, we will be collecting user feedback on our project outcomes via online survey. The survey data collection will occur in Autumn 2019 and Winter 2020, and the data analysis will happen in Winter 2020. The survey data will be collected via Google Forms and analyzed using Microsoft Excel. Phase 2 also involves testing our conceptual model and schema on a practical collection at the VGHF. This will be done by researchers or volunteer students recording information in Microsoft Excel or Google Sheets. The test cataloging activities will happen in Winter and Spring 2020, generating metadata records of sample video game development artifacts. These sample metadata records will be shared via the GAMER Group's website.

Our research notes, raw survey data and interview transcripts will not be shared publicly, but any publications reporting the findings will be shared via ResearchWorks, a digital repository at University of Washington.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

We will be seeking the approval from the University of Washington Institutional Review Board (IRB) for the interview and survey components. We will aim to get the IRB approval by Autumn 2018 for the interview component, and by Autumn 2019 for the survey component. We do not expect this process to be difficult, given that we have successfully obtained the approval for our previous video game metadata work projects.

A.3 Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

Names and contact information (email address) will be collected from interviews, and survey participants who wish to enter for a small prize drawing. This information will be stored separately from the data set. All participants will be assigned unique identifier codes. Links between personal identifiers and subject identifier cords for interviewees and design sessions participants will be kept separately from the data set and after retention schedules are met, will be destroyed.

A.4 If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

Documentation of informed consent, as required by the UW IRB, will be maintained by the PI in locked files and on password-protected computers at the University of Washington per retention schedules.

A.5 What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

In Phase I, researchers will take notes as they consult domain experts and advisory board members. Researchers will also examine game development artifacts and existing databases of such materials, and take notes on how they are currently being organized. The interview data will be collected via in-person or online interviews, conducted by the PI and/or student researcher.

In Phase II, we will set up an online survey using Google forms and collected users' feedback to our project outcomes. In addition, researchers will record metadata of sample video game development artifacts using Microsoft Excel or Google Sheets.

A.6 What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

We will generate a codebook during our qualitative coding process of the interview data (in Phase I) and survey data (in Phase II) and apply it to code the data. The codebook will be stored in DOCX and/or CSV format in the Google Team Drive and implemented within NVivo. Researchers will be using our conceptual model, schema, CVs, and best practices guide to generate the sample metadata records, which will be stored as CSV files. Throughout the whole design process, researchers will be producing notes for improving the conceptual model, schema, CVs, and guide, which will be stored as DOCX files or google documents.

A.7 What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

Our digital assets will be shared in perpetuity through public-facing websites provided by the UW iSchool, just as we have been doing for our model, schema, and CVs for video games (http://gamer.ischool.uw.edu/official_release/). The VGHF will share copies via their website as well. Copies will be provided through the OMR as long as that site allows. All of these sources are expected to remain public for their lifetimes.

A.8 Identify where you will deposit the dataset(s):

Name of repository: UW iSchool GAMER Group Website
URL: http://gamer.ischool.uw.edu/official_release

Name of repository: Video Game History Foundation Website
URL: https://gamehistory.org/

Name of repository: Open Metadata Registry
URL: http://metadataregistry.org/

Name of repository: ResearchWorks at University of Washington
URL: https://digital.lib.washington.edu/researchworks/

A.9 When and how frequently will you review this data management plan? How will the implementation be monitored?

Review and maintenance of generated assets will occur even after the grant period has expired. Our digital assets will be revised as required even after the period of the grant, just as we have been doing for our model, schema, and CVs for video games which are also available at the site above. Typically, we make minor revisions to the schema once every one to two years. Controlled vocabularies are reviewed and revised every 3-4 years. The model we have created for video games has been reviewed as necessary, when there was a new development related to the objects being modeled (e.g., development of augmented reality games), and we will practice a similar kind of review process for the model for game development artifacts.