

How Will We Preserve Virtual Worlds? About the Preserving Virtual Worlds Project

Lead institution: University of Illinois at Urbana-Champaign
Principal Investigator: Jerome McDonough (jmcdonou@uiuc.edu)
Project coordinator and main contact: Janet Eke (jeke@uiuc.edu; 217-333-4701)

Partners: University of Maryland, Stanford University,
Rochester Institute of Technology, Linden Lab

This document provides additional detail and background information about the Preserving Virtual Worlds project recently funded by the Preserving Creative America (PCA) initiative under the National Digital Information Infrastructure Preservation Program (NDIIPP) administered by the Library of Congress. Project work will begin in **January 2008** and continue through December 2009.

For more information about the Preserving Virtual Worlds project, please contact project coordinator Janet Eke (jeke@uiuc.edu). For more information about NDIIPP and its associated initiatives, please visit <http://www.digitalpreservation.gov>, or contact Guy Lamolinara at the Library of Congress (glam@loc.gov).

13 August 2007

Contents

- A. Summary
 - B. Background: The Problem to Be Addressed
 - C. Project Goal
 - D. Project Activities
 - E. Partners and key personnel
 - F. Appendix
 - F.1 Case Studies Under Consideration (NOT finalized)
 - F.2 Additional background on schema development and archiving activities
-

A. Summary

Interactive media are highly complex and at high risk for loss as technologies rapidly become obsolete. The **Preserving Virtual Worlds** project will explore methods for preserving digital games and interactive fiction. Major activities will include developing basic standards for metadata and content representation and conducting a series of archiving case studies for early video games, electronic literature and *Second Life*, an interactive multiplayer game. *Second Life* content participants include *Life to the Second Power*, *Democracy Island* and the International Spaceflight Museum. *Partners: University of Illinois at Urbana-Champaign (lead), University of Maryland, Stanford University, Rochester Institute of Technology and Linden Lab.*

The Preserving Virtual Worlds project is funded by the **Preserving Creative America** initiative under the National Digital Information Infrastructure Preservation Program (**NDIIPP**) administered by the **Library of Congress**.

[Above includes excerpt from Library of Congress Press release; 3 August 2007.]

B. Background: The Problem to Be Addressed

This project addresses a neglected topic in digital media preservation: methods, infrastructure, standards, and technology for preserving the complex software, content, and interactivity in computer games and electronic literature, as well as the transactions and interactions that constitute the user's experience of them.

The Preserving Virtual Worlds project recognizes that hypertext, on the one hand, and computer and video games on the other, have assumed a prominent place among media for entertainment, communication, and social interaction. The cultural profile of these virtual worlds has extended to education (training, "serious games"), artistic expression, social networking, and political commentary ("newsgames"). Such interactive media have become an important part of contemporary cultural expression and creativity in the United States, and that importance is not only about the economic growth of computer, video, networked and mobile games as an entertainment industry, which has equaled or surpassed movies, television and other media: at least as important is their impact on society and culture.

Anyone familiar with interactive fiction, the social worlds of MUDs and massively multiplayer games, the technical mastery of first-person shooters, the visual storytelling of adventure and role-playing games, the intellectual depth of historical and strategy games, or the emergent behaviors and cultures in open-ended simulations, knows that "virtual worlds" is a label that covers an astonishing breadth of art, entertainment, political expression, simulation, competition, and narrative experience. Despite all these reasons for taking them seriously, relatively little work has been done on the preservation of virtual worlds, which present particular challenges because of their interactivity, their frequent software modification and revision, networked collaboration, use of 3-D graphics and sound, and the like. Each of these adds new complexity to the problems of preserving digital content.

Electronic literature, video games and computer games must be understood as creative, born-digital works with distinctive aesthetic qualities that not only take advantage of digital technologies but also push the limits of digital media. These works are typically more experimental and diverse than other kinds of born-digital artifacts more familiar to libraries at this point—for example, digital documents. Electronic literature and digital games provide new kinds of test-beds for digital preservation. Addressing the problem of their preservation means preparing for a future in which an increasing proportion of what we create will be born-digital and will take fuller advantage of networked, new-media environments. These virtual worlds are actualized in user experiences that are sometimes unique, often social, and always necessary for understanding these worlds. Just as an archived book is of limited use if researchers cannot open its cover and

read it, an archived world will be of limited use if researchers cannot visit it. Unless we also develop solutions for preserving user experiences, future generations will have no way to understand how these experiences became such an important part of our culture.

C. Project Goal

Our goal is to help develop mechanisms and methods for preserving digital games and interactive fiction by

- 1) developing basic standards for metadata and content representation; and
- 2) investigating preservation issues through a series of archiving case studies representing a) early games and literature and b) later interactive multi-player game environments.

Key deliverables include development of metadata schema and wrapper recommendations, the archiving of key representative content and the development of generalizable archiving approaches for preserving this content. Our approach is intended to address both the pressing need to preserve the bits and available representation information of early and significant works now, and the need to begin to address more difficult issues surrounding long-term preservation of more recent multi-player interactive virtual worlds.

D. Project Activities

The Preserving Virtual Worlds project is broken roughly into three phases: 1) Scoping and background research; 2) Schema development; 3) Implementation (archiving of test cases) and Wrap-up. Some additional detail is provided below. For further information please contact Janet Eke (jeke@uiuc.edu).

Phase	Top-level tasks include...
Phase 1: Scoping and Background Research	<ul style="list-style-type: none"> • Identify representative types of preservation issues/considerations posed by the case set (including Second Life) • Develop a beginning framework for characterizing game and interactive fiction by type of preservation problem (based on case set) • Identify and attempt to resolve specific IP issues associated with individual works in case set proposed for archiving • Conduct background research for preserving complex interactive user-behavior, including surveying existing relevant taxonomies for documenting game behaviour and interactive fiction behaviour and analyzing examples of selected game-play data in order to better understand the role such content could play in preservation • Conduct background research for schema development, including researching emulation and migration strategies currently in use, and reviewing preservation policies of collecting institutions
Phase 2: Schema Development	<ul style="list-style-type: none"> • Develop new schema to capture technical metadata and other representation information for the data formats included in our case studies • Develop new schema for description of Context Information for digital objects • Make recommendations for use of selected existing wrapper formats
Phase 3: Implementation (ingestion of test cases) and Wrap-up	<ul style="list-style-type: none"> • Ingest a selection of early interactive fiction and early games (dependent on resolution of technical and IP issues) using existing institutional repositories at Stanford, UIUC and (TBD) Maryland • Ingest selected participating Second Life projects (dependent on resolution of technical and IP issues) using existing institutional repositories at Stanford, UIUC and (TBD) Maryland • Analyze institutional repositories' capabilities to support full metadata set necessary for curation of games and interactive fiction • Complete final project reports and recommendations
Key deliverables include...	
<ul style="list-style-type: none"> • Identification of representative types of preservation problems posed by the case set • Development of beginning framework for characterizing game and interactive fiction by preservation problem • Surveys of existing taxonomies for documenting game behavior and interactive fiction behaviour • New schema to capture technical metadata and other representation information for the data formats included in our case studies • New schema for description of Context Information for digital objects • Recommendations on use of selected existing wrapper formats • Archived content (selected test cases) and repository analysis • Feedback/recommendations to LC for their national collecting plan 	

E. Project Partners

Institution	Expertise/role
<p>UIUC (Graduate School of Library & Information Science, and UI Library) – Lead institution http://www.lis.uiuc.edu/ http://www.library.uiuc.edu/</p> <p>Leads</p> <ul style="list-style-type: none"> • Jerry McDonough, PI (jmcdonou@uiuc.edu) • Janet Eke, project coordinator (jeke@uiuc.edu) 	<ul style="list-style-type: none"> • digital preservation theory and research • repository development and management • XML schema development, metadata schema and standards development • metadata and repository workflow implementation • project management
<p>University of Maryland (Maryland Institute for Technology in the Humanities (MITH)) http://www.mith2.umd.edu</p> <p>Leads</p> <ul style="list-style-type: none"> • Neil Fraistat (fraistat@umd.edu) • Matt Kirschenbaum (mkirschenbaum@gmail.com) • Kari Kraus (karimkraus@gmail.com) 	<ul style="list-style-type: none"> • textual scholarship • notation systems • schema development and format-testing • electronic literature and Second Life subject expertise
<p>Stanford University (Stanford Humanities Laboratory) http://www-sul.stanford.edu/depts/hasrg/histsci/</p> <p>Leads</p> <ul style="list-style-type: none"> • Henry Lowood (lowood@stanford.edu) • Henrik Bennetsen • Michael Shanks 	<ul style="list-style-type: none"> • computer game subject expertise, including history of digital games • game technology • software archives • Second Life development and subject expertise
<p>Rochester Institute of Technology (Game Design & Development) http://games.rit.edu</p> <p>Lead</p> <ul style="list-style-type: none"> • Andrew Phelps (amp@it.rit.edu) 	<ul style="list-style-type: none"> • computer game subject expertise * technical expertise, including understanding game engines and emulation
<p>Linden Lab (Second Life) http://lindenlab.com/</p>	<ul style="list-style-type: none"> • subject expertise • consultant • Participating Second Life projects include: <ul style="list-style-type: none"> • Life to the Second Power (the Stanford and Lynn Hershman project) • Democracy Island (New York Law School's Democracy Design Project, Beth Noveck) • The International Spaceflight Museum (Spaceflight Museum Planning Group)

F. APPENDIX

F.1 Case Studies Under Discussion (Interactive Literature and Games)

Please NOTE that final determination of test cases for archiving will be made in Phase 1 of the project, and is dependent on resolution of IP and technical issues, and on project resources. **NONE** of the following has been finalized.

1. Adventure, aka ADVENT and Colossal Cave
2. Mystery House
3. Mindwheel
4. Works of Deena Larsen
5. Second Life electronic books ("prim books")
6. Spacewar! (MIT, 1962)
7. Star Raiders (Atari, 1979)
8. Sim City (Maxis, 1989)
9. Civilization I/II (MicroProse, 1991-1996)
10. DOOM (id, 1993)
11. Warcraft I/II/III (Blizzard, 1994-2003)
12. PLATO
13. Life to the Second Power (Second Life)
14. The International Space Museum (Second Life)
15. Democracy Island (Second Life)

F.2 Additional Background

1. Metadata Schema Development

Problem to be addressed: *In order to function as part of the digital library, digital content needs to be wrapped in what the OAIS Reference model calls an "information package." Current wrapper formats do not provide suitable ways of documenting interactive fiction and games at the bit-level: specifically, they fail to provide the "representation information" needed to map the raw bits into higher-level data constructs. In the case of highly complex, interactive objects such as hypertext fiction and games, inadequate representation information will severely hamper preservation of these works.*

There have been a variety of recent efforts to develop XML-based 'wrapper' formats to bundle all of the content and metadata for a digital object in a single package. Notable examples include FOXML, METS, MPEG-21 DIDL, MXF and XFDU. Within the digital library community, a great deal of emphasis has been placed on the need to have these wrapper formats play the role of an "information package" as described within the Open Archival Information System Reference Model (see Figure 1, in Appendix B)¹. However, none of the packaging formats developed to date can fully support the OAIS Reference Model notion of an information package for interactive fiction or games. A critical component of the OAIS information package is "representation information," the set of information necessary to interpret binary data. Without representation information, a digital object is essentially an undecipherable string of ones and zeros. All of the various packaging standards mentioned provide facilities for referencing representation information, but this makes them highly dependent on external schema. Unfortunately, the schema developed for recording representation information during the past several years do not yet provide the level of documentation necessary to understand digital file formats fully. To date, because a great deal of

¹ From "Preservation Metadata and the OAIS Reference Model: A Metadata Framework to Support the Preservation of Digital Objects," a report by the OCLC/RLG Working Group on Preservation Metadata, June 2002. http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

the digital library community's work has focused on documents with fairly simple digital representations, and because the community has a strong bias towards open, standard formats, this limitation has not emerged as a critical flaw. In the long term, however, the simplifying assumption that digital content will be simple rather than complex and will be produced in open formats is dangerous, especially for highly interactive virtual worlds such as interactive fiction and games, which are very likely to come in proprietary rather than open formats, with competitive disincentives to documentation.

Proposed activities and deliverables

In order to help assure the longevity of these complex works, the project will embark on an effort to build upon existing work in metadata wrappers and develop structures which will allow us to record the complete set of Representation Information and Preservation Description Information needed to support long-term preservation.

Since it is unclear whether migration, emulation or a combination of both may best serve to enable these objects to survive, we will seek to create or elaborate metadata standards that will support both approaches. We anticipate that successful completion of this task will require:

- development of new schema to capture technical metadata and other representation information for the data formats included in our case studies;
- new schema for description of Context Information for digital objects;
- new schema for preserving complex interactive user-behavior;
- new schema for structural metadata to encode interactive fiction;
- a set of suggested elaborations of existing wrapper formats (along with recommendations on use practices) to allow for complete support of representation information.

The project will demonstrate the use of the new schema and the inclusion of Representation Information through a revised version of the METS format and/or development of the Electronic Literature Organization's (ELO) proposed X-Lit format.

2. Archiving of Test Cases

Problem to be explored: *It is important that we support multiple approaches to the preservation of complex works, taking into account both emulation and migration strategies, and considering different objects of preservation, from the software bit-stream to the player experience. How will the software, digital assets, and content that form the basis of virtual worlds be collected and preserved? Given the nature of virtual worlds as interactive spaces, how will the activities of players/readers (including their activities as content creators) be captured? To examine these issues in a real-world setting, the project will attempt to ingest both early examples and later examples of interactive fiction and computer games into existing institutional repositories using the schema developed in Phase 2.*

There are two basic strategies for preserving complex works: emulation and migration. With respect to virtual worlds, emulation requires the simulation of the original operating system and environment for which the game was originally designed, using contemporary hardware and software, while migration requires one to translate digital objects, relationships, and behaviors into an alternative hardware- and software-dependent format, in order to free it from the constraint of having to exist within the environment for which it was originally designed. With respect to the activities of players/readers within these worlds, both emulation and migration might allow "replaying" a game or a reading. Emulation may provide a more exact replication of the original experience and also preserves the original format of the work, preserving our knowledge of design techniques of the period; migration may have a greater likelihood of insuring long-term access. Each approach involves compromises and costs, and part of the purpose of this set of experiments would be to understand what those are, and which might be acceptable

under what circumstances.

Proposed activities and deliverables

Based on the METS metadata wrapper development and X-Lit work done in connection with the ELO, the project will archive a selection of early interactive fiction and early games from the Stanford University, University of Maryland and University of Illinois at Urbana-Champaign collections using existing institutional repositories (dependent on resolution of IP and technical issues). Archiving activities will consist of bitstream-level preservation in a metadata wrapper, and a commitment by the owning institution to support long-term curation of the resources. We expect that the approaches developed here for preservation of interactive fiction and games will have wider applications to other forms of interactive content.

In addition to preserving bits and representation information, the project will investigate more complex approaches to preservation. This will include examining issues surrounding the preservation of later interactive multi-player game environments using the Linden Lab Second Life virtual world as a case study.

Specific tasks and activities are expected to include:

- identification of the types of preservation problems posed by the representative case set;
- surveys of existing taxonomies for documenting games, game behavior and interactive fiction behavior;
- analysis of examples of game-play data (such as server-side transaction logs) in order to better understand the role such content could play in preserving the virtual world cultural record;
- research into emulation and migration strategies currently in use;
- consideration of where metadata comes from, how it might be auto-generated, and implications for developer workflow to support preservation.