

The International Journal of Digital Curation

Issue 2, Volume 6 | 2011

Saving Second Life: Issues in Archiving a Complex, Multi-User Virtual World

Jerome McDonough & Robert Olendorf,
Graduate School of Library & Information Science,
University of Illinois at Urbana-Champaign

Abstract

Virtual environments, such as Second Life, have assumed an increasingly important role in popular culture, education and research. Unfortunately, we have almost no practical experience in how to preserve these highly dynamic, interactive information resources. This article reports on research by the National Digital Information Infrastructure for Preservation Program (NDIIPP)-funded Preserving Virtual Worlds project, which examines the issues that arise when attempting to archive regions from Second Life. Intellectual property and contractual issues can raise significant impediments to the creation of an archival information package for these environments, as can the technical design of the worlds themselves. We discuss the implication of these impediments for distributed models of preservation, such as NDIIPP.




Introduction

The Preserving Virtual Worlds Project is a research effort, conducted as part of the Library of Congress' National Digital Information Infrastructure for Preservation Program, to investigate issues around the preservation of video games and interactive fiction. Researchers from the Rochester Institute of Technology, Stanford University, the University of Illinois and the University of Maryland have been examining a variety of computer games, from early games such as Spacewar! on the PDP-1 to more modern multi-user virtual environments. As part of this research, our group has been collaborating with Linden Lab to investigate issues around the preservation of Second Life.

While a virtual environment, such as Second Life, may seem an odd choice for the focus of a preservation project, such environments are becoming increasingly important to the scholarly community and to the larger public. Economically, computer games have become a significant part of the global economy; global sales of video games were estimated at \$46.5 billion in 2009 (Wu, [2010](#)), and within the United States, direct and indirect employment by the gaming software industry accounts for over 80,000 jobs (Siwek, [2007](#)). Socially, video games have become one of the most popular forms of entertainment within the United States, with 67% of American households playing computer or video games (Entertainment Software Association, [2010](#)), and with online gaming sites in the United States reporting over 190 million visitors in a single month in 2008 (comScore, [2009](#)). Games are also starting to have a significant cultural impact, influencing other forms of media (e.g., the rise of movies derived from video games, including *Tomb Raider*, *Resident Evil*, *Final Fantasy* and *Doom*) and in turn being influenced by the wider culture, such as the placement of advertisements in XBox games by the Obama presidential campaign (GamePolitics.com, [2008](#)). Bainbridge ([2007](#)) has also noted the potential impact of virtual environments on the research community, both as virtual laboratory spaces and as settings for economic and social research. Increasingly, any complete understanding of modern society and culture requires an understanding of the world of gaming, and if cultural heritage institutions are to adequately serve researchers, we must develop means of preserving these new information resources.

Unfortunately, the importance of these resources often appears to be matched by their transience. Our group has been tracking the status of various virtual worlds similar to Second Life during the course of our project. Thirteen different multi-user virtual environments, some of them gaming oriented, some social, ceased to exist during the course of our research, including some relatively well-known systems such as there.com, the Matrix Online, the Lively system from Google, and EA-Land (perhaps better known by its previous name, The Sims Online). The Sims Online lasted less than six years before Electronic Arts shut it down.

The ephemeral nature of some of these worlds has influenced the experimental approach we have taken in trying to archive regions from Second Life. When a virtual environment ceases to exist, the software employed by a gaming company to manage and maintain that world is unlikely to be made available. We took as an operating assumption for our project that we could not depend on the on-going availability of the server platform for the worlds we wished to archive, and that we should therefore attempt to archive these worlds in a manner which would, to the extent possible,



separate the content of the virtual environment from the underlying platform and allow the first to survive, even if the second did not. Publishers have long relied on markup languages such as SGML to separate the structure of content from its presentation in order to insure the content's on-going viability for reuse (Clark, [2008](#)). Our hope is that by employing an approach to archiving virtual worlds that separates their data structure from the software enabling their presentation, we will help contribute to the ongoing accessibility of these worlds as materials for research and scholarship.

While attempting to separate the content of the virtual environment from its underlying platform, our project has also committed to trying to preserve the interactive nature of the different games and interactive environments we are examining, as far as possible. A fundamental difference between the virtual environments our project is examining and many previous digital preservation projects is that these environments are highly interactive and highly dynamic – the game you played yesterday will not be the game you play today. Preserving this dynamic component of games and virtual worlds has meant that a purely documentary approach to preservation through the use of screen shots, video capture of game play, textual walk through of game play, etc., was not sufficient. Such documentation might provide a useful supplement to our preservation efforts in many cases, but perhaps the most significant property of virtual environments is their interactivity, and so our efforts have tried to preserve this as much as possible.

We have also been trying to determine the feasibility of archiving regions of Second Life while operating as ordinary users without employing special access mechanisms or privileges that might be available to Linden Lab employees. While Linden Lab is a partner in our research, cultural heritage institutions cannot necessarily rely on the cooperation of gaming companies when attempting to obtain copies of virtual environments for preservation purposes. By approaching the preservation of Second Life regions from the perspective of a third party without any special access to the server software or data sets, we hope to gain a greater understanding of the problems which may confront cultural heritage institutions when trying to archive these settings as part of their normal, day-to-day work.

A First Look at Second Life

Second Life is a multiuser online virtual world developed by Linden Lab that first became available to the public in 2002. Second Life is an example of what are known as “social worlds” (Book, [2004](#)), virtual environments in which there are no predefined rules or objectives and in which the primary use is social interaction with other users (known in Second Life as “residents”) within the virtual environment. Second Life generally conforms to a client/server architecture, with different servers operated by Linden Lab handling different components of the service:

- Login Server – manages user authentication and login processes;
- User Server – manages instant messaging sessions;
- Space Server – manages the routing of messages between residents based upon their location in the virtual space;
- Data Server – manages connections to the various data bases containing Second Life's data and log information;
- Simulators – each simulator manages the state of a single region in Second Life, including the state of both objects and the terrain, and managing the simulation of physics for the region.


Linden Lab has made the source code for the client software for Second Life publicly available under the terms of the GNU General Public License, version 2.0, with the artwork for the viewer licensed under the Creative Commons Attribution/Share-Alike 3.0 license. The software for the server components is closed source and is not publicly available.

Second Life's virtual environment is organized into different 256 x 256 meter regions, typically referred to as islands or sims (short for "Simulators"). A region can be owned either by Linden Lab or by one of Second Life's residents. Each region is essentially an independent environment within Second Life, hosted on its own dedicated server, although residents can move freely between the different regions. The contents of a given region consists primarily of a combination of 3D objects; graphical texture files, which can be overlaid on the 3D objects; audio files providing background noises; and scripts which enable interactivity with the various objects within a region. All of this content is hosted on Linden Labs servers, with the different servers managing the interactions between residents' client software applications and the various regions.

Objects in Second Life are created by linking together one or more shape primitives, or "prims". The primitives come in eight basic shapes (box, prism, cylinder, sphere, torus, ring, tube and sculpted) that can be further modified from their default shapes. A prim is more complex than its name might imply, as each prim has an associated inventory that can contain scripts, notecards, textures and other items. By combining these primitives, residents can build an astonishing variety of complex objects. Textures (image files applied to the faces of the primitives) give objects the illusion of possessing even more detail. Each object has a set of metadata elements associated with it, including the identity of the individual who originally created the object (the intellectual property rights holder), known as the Creator, and the individual who has possession of the object, known as the Owner. See Figure 1 for an example of the editing pane displaying metadata for a complex object, a house.



Figure 1. Manipulating Objects in Second Life.




Objects can be made to interact with avatars, other objects or even communicate with external resources by the inclusion of scripts that are bound to objects. Scripts, written in the Linden Scripting Language (LSL)¹ are integral to making Second Life an interactive environment, and by including sounds, textures or a variety of other assets, Second Life objects can be made to accomplish a wide variety of functions. In any given region there can be several thousand objects. These objects can have a variety of different creators and owners. Additionally, the textures, scripts and other assets associated with the objects may have different creators that are difficult, if not impossible, to identify.

This unique environment presents some special problems compared to other objects of digital preservation. First, there are technological impediments both to obtaining certain portions of Second Life's content and to insuring ongoing access to that content. Due to the permissions system employed by Second Life's servers, an object's scripts and some other prim content are, by default, unavailable to anyone but the object's creator, and so cannot be accessed (or copied) by another resident. Even if the scripts should be obtained, they are written in Linden Scripting Language, not a standard scripting or programming language, so they are not particularly useful outside the Second Life environment. It would be necessary to recreate the content in an environment such as OpenSimulator (Fishwick, 2009), reverse engineering the scripts or translating them to another language to achieve the desired behavior. Even recreating content that is available for copying, such as objects' geometries and textures, can be problematic. While it is possible to extract shape data recording objects' geometry and store it in an XML format that can be transformed to other standard schema representations of 3D objects, subtle variations in virtual environments (for example, how the physics of such phenomenon as light and gravity are implemented within the simulated environment) can lead to an imperfect reproduction of the original experience. Extracting an object from Second Life and using the information for that object to instantiate a new instance in an open source virtual environment such as OpenSimulator or Sirikata (Horn et al., 2009) may produce a subtly different user experience of the object.

Another potential difficulty in archiving Second Life content relates to issues of intellectual property and contract law. In any given region, there are likely to be objects from a large number of different content creators. The regions we investigated contained objects from three to fifty content creators. The Terms of Service agreement covering residents' access and use of the Second Life service specifically states: "You retain any and all intellectual property rights in content you submit to the service" (Linden Lab, 2010). The Terms of Service also forbid any infringement of intellectual property through unauthorized copying of content available through the service, stating that: "You must obtain from the applicable Content Providers any necessary license rights in Content that you desire to use or access" and "You agree that you will not copy, transfer, or distribute outside the Service any Content that contains any Linden In-World Content, in whole or in part or in modified or unmodified form, except as allowed by the Snapshot and Machinima Policy, or that infringes or violates any intellectual property rights of Linden Lab, other Content Providers, or any third parties." The contractual and legal frameworks for Second Life therefore forbid any effort to make a preservation copy of the contents of an island without the explicit

¹ Full documentation of the Linden Scripting Language is available through the Second Life LSL Portal at http://wiki.secondlife.com/wiki/LSL_Portal.



permission of all of the content creators who have objects present within that island. Making a preservation copy of an island thus requires that all content creators for objects in that island be contacted with a request to allow us to archive their creations. Potentially a large fraction of any region's creators could refuse to allow us to archive their creations, and in that case, the best we can manage is to archive those portions of the island for which we have received permission to make an archival copy.

A final difficulty with archiving Second Life content is the issue of determining the full scope of what is worth archiving. Even if we were to manage to capture all of the data from a region, including animations and scripting, we may still be left with little knowledge of how the region was actually used. A complete copy of a Second Life island, instantiated separately from the Second Life service, is a world preserved in amber. It may endure through time, but it will not stay alive. Second Life is a highly dynamic, interactive, multi-user resource, and attempting to archive it is at some level akin to attempting to archive the state of Illinois. We cannot preserve the island exactly as it is, because the nature of the island is that it is an evolving, dynamic region. All we can really do is document the island, its people and its culture, and attempt to preserve that documentation. A snapshot of the physical state of an island in Second Life is certainly one important form of documentation, but it is reasonable to ask whether it alone is sufficient, or whether we must also collect other documentation regarding the behavior of users visiting the region, or relating to the social function of the region.


An Approach to Archiving Second Life

The approach we have taken to archiving content from Second Life involves a multistage process. We first create a manifest of a region's contents and gather basic metadata about the objects within the region. We then contact all of the content creators for objects in the region and ask for permission to archive their creations. We then download all of the structural data for objects in a region for which we have permission to archive. We determine what additional information, beyond the raw data for the region, we might wish to archive along with the data drawn from Second Life. Finally, we create a submission information package that includes the various data objects from the world along with related metadata and other ancillary data. Details on our process follow.

Creating a Manifest

We first create a complete manifest of a region's contents and collect basic descriptive metadata regarding the objects within that region using scripted Second Life objects (probes). The probes are controlled from within Second Life using additional scripted objects worn by a resident's avatar. The probes exhaustively search a region for objects. They then collect basic information about each object (object name, description, position, owner, creator and permissions). This descriptive metadata is sent via HTTP to a web server where the data is processed and inserted into a MySQL database. While the process appears straightforward, there are a number of technical issues that actually make this difficult.

The probe uses the LSL function `llSensor()` to detect objects. The `llSensor()` function has a maximum search range of a 96-meter radius sphere. This presents a problem in that spheres do not stack well. A search conducted on adjacent spheres within a world would, of necessity, omit a fair amount of space from its search. Therefore, the probes limit their data collection to objects within a cube of the



maximum volume that can still be bounded by the sphere; overlapping search spheres allow us to insure that the entire volume of the region is searched. A second and more problematic limitation of the `l1Sensor()` function is that it only returns the sixteen closest objects in the search space and it is impossible to tell if there are more objects within the sensor's search radius than those sixteen. If more than sixteen objects are in the search space, the more distant objects from the probe are silently dropped from the `l1Sensor()` function's output. To overcome this problem, a three dimensional search tree algorithm was devised. A probe starts at its maximum search space. If it detects 16 objects, the search volume is divided into eight equal subspaces. Each of those subspaces is searched. If any of those subspaces contain 16 objects, they are in turn divided into eight equal subspaces and the process continues recursively until all of the subspaces have less than 16 objects, or until a subspace radius less than 0.001 meters is encountered.

Another problem with our probe mechanism for gathering object information is that Second Life regions can be fairly complex, creating a hazardous environment for the probes. Some regions in Second Life have been set by their owners to be "no script" territory; if our probe enters such a region, its scripts cease to run and the probe essentially dies. In those instances, our options are to contact the landowner to negotiate access to allow our scripted probes to run, or to exclude the territory from our archive. We have also encountered areas that our probe could not enter due to obstacles blocking their path. In some of those cases, while our probe continued to operate, it would become trapped in one location and be unable to continue navigating its way out of that space to complete its search. Second Life scripts are limited to 64 kilobytes of memory, making the construction or use of a more sophisticated algorithm for navigating Second Life (and avoiding the problems of having our probes become trapped) problematic. Using the simpler expedient of launching probes from different locations to insure full coverage of a region enabled us to successfully scan those cases.

Obtaining Permissions

After having obtained an inventory of the objects within a region, we then proceed to contact the creators of those objects to obtain permission to archive their materials. Given the potentially large number of content creators in any given region, we use a semi-automated system to obtain permission to archive objects. The avatar names and UUIDs (a unique identifier given to every asset in Second Life) of all the content creators from a region are pulled from the data base, and sent to a Second Life object that serves both as a dedicated server for handling requests from our web server, and also acts as a portal for interaction with avatars. The portal receives the avatar information and sends a notecard within Second Life to the creators, describing our project, what we want to do with the objects they created, and an in-world teleport link they can use to travel to our portal, where they can access our web interface for granting or denying permission to archive their materials.

At the portal, creators use the object's interface to access the web interface for permissions management. The creators are first asked to set up an account, and then they are presented with a list of all their objects we wish to archive. The creator may choose to allow us to archive all of their objects, some of their objects or none. Creators are also allowed to specify an embargo period, during which their content will not be made public.



Harvesting Content Objects

The geometrical data for all of the objects for which we have obtained permission to archive, along with their associated textures, are gathered using the Copy Bot software. Copy Bot is a text-based Second Life client that allows the user to, among other things, download all of the data for any object, including information on ownership, shape, position, size, and other details, to their local computer. The output arrives as an XML file, and the associated textures for the object are downloaded concurrently.

Copy Bot is written to download a single object at a time. To download all of the objects in a region, we first pull all of the object UUIDs and locations from our database. The web server then sends HTTP requests to an attachment worn by the Copy Bot avatar, with the UUID and the location of the object. The attachment then sends a message to the bot to move close to the object of interest and download it. Once all of the data is downloaded, it is moved to the web server, in preparation to create the submission information package.

Obtaining Additional Preservation Metadata

The OAIS reference model defines a number of different types of information that should be recorded and maintained for a digital resource that you wish to preserve, including:

- Descriptive information – information that supports an archive’s users in finding, identifying and retrieving archived material;
- Context information – information that documents the relationship between an archived object and the larger environment in which it is embedded, including why it was created;
- Fixity information – information (including authentication mechanisms and authentication keys) that allow a user to confirm that archived material has not be changed in an unauthorized manner;
- Representation information – information that maps digital information into more meaningful concepts, e.g., mapping from bit sequences into character data;
- Provenance information – information documenting the history of the archived information.

In addition to providing this information for the individual objects within a region that are archiving, we also wished to be able to record metadata for the region as a whole. To accomplish this we developed an XML schema defining an XML file format that allows us to record some minimal information about the region we have scanned, including its name, location within the Second Life system and a reference to any covenant² the landowner for the region may have established, the date and time the scan of the region was conducted, and information on all of the individual objects

² Covenants are rules over and above the Second Life Terms of Service and Community Standards which the owner of an island has established for that region. Any resident wishing to purchase a parcel of land on an island agrees to be bound by the terms of the covenant for the island. A common use of covenants is to establish “zoning laws” for an island, dictating what types of structures can be built.

contained within that region. We also developed web-based and Second Life-based tools to aid in the generation of this metadata. Minimal descriptive information for the regions and objects within them are obtained from Second Life itself, and basic fixity information for objects is generated automatically. Context, representation and provenance information are somewhat more difficult to generate automatically and at this point are being created manually.

Because one of the most important aspects of the Second Life is the interaction of the users' avatars with both the environment and each other, the information packages that we have created for Second Life regions contain additional context information to document the region's history and use. As with all metadata, contextual information could be applied at all levels of the content object, from the region down to individual objects. In addition, we included the ability to add contextual narratives into the information package for a region. The archivist could visit various places in a region and describe in free text any important properties, events or other aspects associated with that location. In addition, snapshots of the location, events or objects associated with the narrative can be uploaded as well. Figure 2 shows one of forty-seven screen captures which we took of the International Space Flight Museum (ISFM) in Second Life (one of the regions we are archiving) as part of the context information we are collecting regarding that region. Additional context information for a region, such as the ISFM, might include a harvested copy of the ISFM's separate website. In addition to these region-specific forms of context information, we are also including in our information packages for the various regions additional documentation on Second Life itself, including the Second Life Linden Scripting Language wiki, the Second Life website and a number of tutorials on using Second Life.



Figure 2. International Space Flight Museum.

As the majority of the information about objects in Second Life that we download is stored in XML format, and the texture files downloaded for those objects are all in the JPEG 2000 format, the necessary representation information for Second Life content is actually relatively small. Additional representation information may be required for context information objects (such as harvested websites). Provenance information, at this point, is extremely minimal; we record the date and time a region

was scanned, and the individual responsible for conducting the scan. We also obtain information from Second Life regarding object ownership.

In order to record the relationships between a particular data object and its representation information, we have developed an OWL ontology, which includes concepts from both the OAIS reference model and from the Functional Requirements for Bibliographic Records Final Report (IFLA Study Group, 2009). OAI-ORE records can then be used to aggregate an object's data files and the files containing representation information for those data files, as well as recording which representation information is necessary to decode which data file. The combination of our XML schema for recording the basic information defining a region and its contents, the Copy Bot XML schema for recording details on each individual object, our OWL ontology and a packaging format, such as OAI-ORE, gives us the ability to record the full set of information we need to document a Second Life region. Figure 3 gives a simplified overview of an instance of a region containing two objects with an external website providing additional context information for the region.

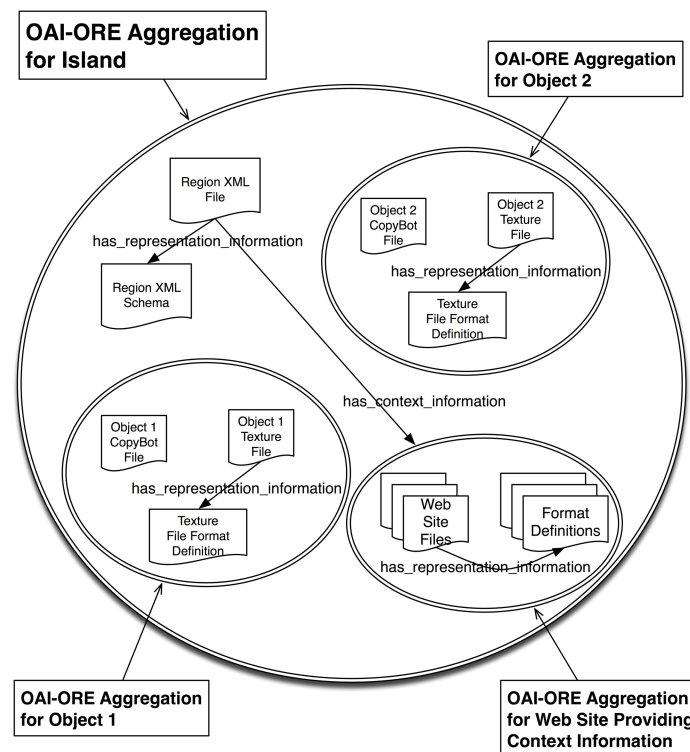


Figure 3. Submission Information Package Framework for a Second Life Region.

Results from Archiving Experiments


We chose eight regions as test cases for our archival process (Table 1). Of those, we were not able to archive Democracy Island due to problems introducing our probes into that region. While the archiving process was generally successful, we identified a number of problems and difficulties.

Island	Description	Object Count	Owners	Creators	Responding Creators (%)	Creators Refusing Permission
Life Squared	A reincarnation of the archive of artist Lynn Hershman Leeson, housed in the Special Collections Library at Stanford University.	742	6	20	2 (10%)	0
Stanford University Libraries	A virtual library space established by Stanford University Libraries in Second Life to support online collaboration, virtual classroom space, and virtual exhibits of library materials.	1800	24	63	5 (8%)	2
Democracy Island	A project of the New York Law School, Democracy Island “overcomes some of the difficulties associated with civic participation and engagement in real space...by offering an on-line space that can be conveniently accessed from home or work” (Institute for Information Law & Policy, 2010).	Could Not Scan.				
International Spaceflight Museum (Spaceport Alpha)	A virtual museum hosting exhibits and events about spacecraft and space travel, managed by the ISM Corporation, a Kansas non-profit corporation.	1951	30	70	7(10%)	0
Rumsey Historical Maps in Second Life (Consisting of four regions: Rumsey Maps 1, Rumsey Maps 2, Rumsey Maps 3 & Rumsey Maps 4)	An instantiation of selections from the David Rumsey map collection, utilizing the features of Second Life achieve effects such as overlaying a map of a particular terrain on a properly scaled model of the terrain.	Rumsey Maps 1: 319 Rumsey Maps 2: 861 Rumsey Maps 3: 595 Rumsey Maps 4: 386	Rumsey Maps 1: 5 Rumsey Maps 2: 3 Rumsey Maps 3: 4 Rumsey Maps 4: 4	Rumsey Maps 1: 6 Rumsey Maps 2: 3 Rumsey Maps 3: 6 Rumsey Maps 4: 4	Rumsey Maps 1: 0 (0%) Rumsey Maps 2: 0 (0%) Rumsey Maps 3: 0 (0%) Rumsey Maps 4: 0 (0%)	Rumsey Maps 1: 0 Rumsey Maps 2: 0 Rumsey Maps 3: 0 Rumsey Maps 4: 0

Table 1. Archival Test Regions in Second Life.

Creating the Manifest

The probes were largely successful in creating the manifest for each region. The time to complete the scan ranged from approximately ten to thirty minutes. In all but one case, the probes successfully scanned the regions, and uploaded the basic object data and metadata to the database. The primary difficulty lay mostly with the building and region management practices exhibited by many of the residents. Each prim in Second Life counts as a single object. Best practice for constructing complex objects in Second Life dictates that an object constituted of multiple prims should have all of the



prims linked into a single link set. This allows the whole link set to be manipulated as a single object. However, it is possible to position prims together in such a way that they appear to form a single coherent object without actually linking them. In regions where residents did not carefully link their creations, there were a much larger number of objects to scan, increasing overall scan time. This also has ramifications for the permission process; content creators are less likely to be aggravated by the process of giving permission to archive a single complex object than having to grant permission to archive each of the individual prims constituting the object.

We were unable to scan one region, Democracy Island, due to an overbuilding problem within the region. The managers of the region allowed the general public to build there and, as a result, the allotted maximum number of prim objects permitted within a single region was already used up before we attempted to archive it. Our probes constitute new objects that must be added to a region before a scan can commence. With all available allotment of objects for the region already consumed, we could not instantiate our probes to conduct the scan. Additionally, the island had been subdivided into a number of different parcels with different permissions settings, each with a different owner. The owners of the region and the parcels were not easily contacted to address the issue.

Obtaining Permissions

While technically simple to implement, the permissions process was perhaps the most difficult part of the project. There were two primary issues. The first was the low response. Of the three regions where we had more than ten creators, we ranged from a five to ten percent response rate from content creators when we asked for permission to archive their creations. Fortunately, of all the respondents, only two refused to have their objects archived. Even among those respondents who agreed to have their objects archived, however, many had significant concerns about the possibility of having their creations pirated and sold. Often the initial reaction to our request for permission to archive was anger. We were able to alleviate these fears, but doing so often required a great deal of one-on-one discussion with individual content creators.

One of the other greatest concerns of many of the respondents was the web interface to the permissions system. In order to verify that the individuals interacting with our web portal for permissions management were the same individuals controlling the avatars listed as content creators for the objects we wished to archive, we required that they initially access the system via our Second Life portal. For many, this potentially represented a significant loss of privacy since we could theoretically record their IP addresses as they connected to our web portal and correlate those addresses with their avatars. Users of virtual environments, such as Second Life, have often demonstrated a strong disinclination to allow the gap between their online and offline identities to be breached (McDonough, 2000). Given that we were unlikely to be able to overcome user resistance to any action which might link their online and offline identities, in cases where content creators expressed this concern, we spoke with their avatars in-world, obtained their permission to archive through chat or voice communication, and submitted their response to our request to archive to our web portal ourselves.

Harvesting Content Objects

Our greatest difficulty with regards to obtaining a copy of the content in Second Life was in obtaining scripts and the other contents of prim's inventories. Scripts especially determine how an object interacts with avatars, other objects and its environment. Scripts can also be used to act as email clients, web servers and web clients. There is currently no easy way to collect a prim's inventory contents automatically. The archivist can ask the object creators for copies of the object's scripts and other contents. However, this can only be done on an object-by-object basis, and there is no way to determine if an object is scripted without manually examining the object. We currently do not gather scripts and other prim contents. Even if we did, however, it would be of limited use since Second Life scripts are in a proprietary language with limited use outside of the Second Life environment.

Obtaining Additional Preservation Metadata

Gathering and creating the metadata was a far more time consuming process than gathering the content. While some of the descriptive metadata and fixity information could be harvested from Second Life or generated automatically (i.e., object creators, MD5 hash values for content, etc.), most metadata required research and had to be created *de novo*. Given the large number of objects represented in the regions, manual generation of such metadata obviously does not scale well as a solution for archiving these worlds.

Given the limitations on our ability to obtain a complete version of the content of the regions we scanned, recording additional context information documenting the regions' creation, meaning and use is critical to insuring scholars' understanding of these regions in the future. But any archivist or curator tasked with collecting such information immediately runs into a difficult problem, familiar to many institutions which have embarked on web archiving projects (Hswe, Kaczmarek, Houser & Eke, [2009](#); Masanès, [2002](#); Reilly et al., [2003](#)): the problem of selection, or to put it another way, deciding when to stop collecting. Second Life is not an island (or more accurately, a set of islands) unto itself. It is an Internet service that is integrated with the larger universe of Internet services, including the World Wide Web. Second Life objects have links out to various external websites, and various websites feature SLURLs (Second Life URLs) allowing individuals with a Second Life client to teleport into particular locations within Second Life. As Lee ([2011](#)) has observed, "If one wanted to know the full context of an entity, one would need an omniscient awareness of all things." The set of information that we might potentially record as context information, even for a particular region in Second Life, is far larger than it is practical to try to save. How, then do we approach the problem of selection with regards to context information for Second Life regions?

The amount and kind of context information we might collect varied region by region in our examples. Some regions, such as the International Spaceflight Museum, had significant external resources, such as websites that supplemented the content of the region. Some regions also had various blog entries about them. Websites created by the regions' creators were always included in the context metadata. Blog entries, and other third party references were included if their content appeared to be informative enough to warrant inclusion. We have also included still images of the worlds we've archived within our information package, and video when available, to document use of the virtual worlds. While this provides information to help users of

the archived versions of these worlds understand them better, we have also tried to stay cognizant of a fact noted by Lee (2011), that “users can also contribute contextual information about points in the life of a digital object after it has been transferred from its original use environment and into the archive.” The features in our XML file format for documenting a region that allow for ongoing curatorial comments to be inserted in the information package are one way in which we’ve sought to support the ongoing addition of context information to these objects. We have also partnered with the Internet Archive to establish a subcollection within their Moving Image Archive on Archiving Virtual Worlds that will allow for ongoing collecting of video context information about virtual worlds.³


A final, vital consideration with respect to context information is the Second Life software itself. As noted in the introduction, our approach to archiving these worlds has been to try to do so in a way that would allow access to the content of the world to survive, even if the underlying platform is no longer available. However, the Second Life platform is a significant intellectual artifact in its own right, and clearly is of significance for those wishing to understand these worlds’ creation and use. The server platforms on which Second Life runs are proprietary, and there is no public documentation for them. The Second Life viewer, however, is open source. We therefore opted to include wiki pages about the Second Life system architecture in our information packages for these regions. We also included source code and binaries for the Second Life viewer, plus web pages documenting the code. There are a number of other viewers derived from the original Second Life viewer that we could have included, but chose not to, primarily to reduce the size of the final package. We also included the Second Life scripting portal, to provide documentation on the Linden Scripting Language.

Fixity metadata was generated as the package was created. We used the SHA1 hash algorithm to generate a checksum for the individual data objects in the package. In addition to information about who initiated the archiving of a region and the date and time our scans were conducted, provenance metadata also includes the software required to gather the content and create the metadata. This includes the software we created to scan regions, Copy Bot and Wget, which we used to acquire copies of websites used for representation, provenance and context metadata.

Conclusions

Any honest analysis of our efforts to archive regions in Second Life using the approach described above would have to conclude that our success has been partial at best. A variety of social and technical factors have imposed severe limitations on our ability to archive Second Life regions. On the social side, the intellectual property law and contractual frameworks governing Second Life content make it impossible for us to legally extract content from Second Life without the permission of its creators. However, as our archiving experiments demonstrate, such permission is extremely difficult to come by, and it is reasonable to ask whether archiving a Second Life region is worth the time and energy if only 10% of the region’s content can actually be preserved. This is particularly true since the technical limitations on extracting content from Second Life mean that some of the content we extract is going to be partial.

³ See http://www.archive.org/details/virtual_worlds for access to the Archiving Virtual Worlds moving image collection.



Without access to objects' inventories (and hence their scripts), some of the most interesting, dynamic features of Second Life are unavailable for preservation copying.


A simple solution to these problems would be to abandon efforts to archive these worlds as third parties. Linden Lab possesses both the legal authority and the technical access necessary to make a complete copy of the various regions in Second Life. The original NDIIPP plan (Library of Congress, [2002](#)) set forth a vision in which a national network of partners would collaborate on insuring ongoing access to digital materials. Linden Lab would, at first glance, seem to be well positioned, legally and technically, to assume the role of curator-in-chief for Second Life content. However, there are a number of reasons why this is not necessarily a practical solution. The first is that Linden Lab is a commercial entity of limited resources, and while they appreciate the interest that librarians, archivists and scholars have taken in their virtual world, arranging for ongoing archival access to their virtual environment is not part of their core mission, and would require a significant allocation of resources for managing the archiving process and creating the infrastructure for stored content that Linden Lab could certainly use more productively elsewhere.

Additionally, while Linden Lab's staff is clearly eminently qualified to create a virtual world, they are not necessarily the best choice to curate a virtual world. The role of curator of a set of content is a very different one from that of creator, with a different set of demands. Smith's ([2006](#)) comments on the challenges that NDIIPP faced as it began to turn its attention to the archiving of commercial content are particularly relevant in this context:

“The past five years have shown that the “real challenges” in digital preservation are not primarily technical or procedural: they are the policies, the politics, and the economic drivers of digital preservation that serve to divide stakeholders as often as they unite them in a common cause. It is no longer true, as it may have been in 2001, that content producers, distributors, and consumers do not understand the risk of data loss. Many of the key stakeholders, from archivists to publishers, film studios to software engineers, scientists to city water engineers, record company executives to real estate developers, are very worried about how loss of data could adversely affect them. But their interests in preservation at best overlap. Just as often they are in conflict, or appear to be in conflict, because they do not share common understandings of the value of that information – for whom, for how long, for what purpose.”

To ask a commercial enterprise, such as Linden Lab, to simultaneously assume the role of a commercial content creator, working to respond to the demands of its customer base, and the role of curator, working to insure the longevity and value of its content to the larger and less well-defined community of current and future scholars who might be interested in examining materials documenting Second Life's development and history, is to place Linden in a fundamentally untenable situation. It is neither fair nor wise to ask a company which must answer to the demands of its investors to also answer to the demands of scholars who have not yet been born.

A possible solution to this would involve a very careful allocation of roles between Linden Lab and cultural memory organizations (research libraries, archives and museums) that would allow Linden Lab to employ their position as the only



organization with legal access to the full range of Second Life content to commit a copy or copies of Second Life regions to a preservation infrastructure created and curated by others. Given the intellectual property restrictions in place on Second Life content, this would of necessity involve Linden Lab using their authority to sublicense access to the materials in Second Life in order to place a copy within a dark archive, with the content made available only after it has lapsed into the public domain. This would allow Linden Lab to concentrate on their core mission, “to connect us all to an online world that advances the human condition” (Linden, [2006](#)), while placing the responsibility for curation and preservation of this content into the hands of institutions already committed to the preservation of knowledge.

Unfortunately, the dark archiving requirements for this content do mean that any institution committing itself to maintaining this content is committing itself for a long haul. Even if we were to assume that placing content in Second Life counts as pseudonymous publication (as it is published with an avatar name associated with it, rather than the user’s real name), copyright would adhere to Second Life content for 95 years from its date of publication (Copyright Law of the United States, [2009](#)). Asking an archive to hold digital content for nearly one hundred years before it is made publicly available is asking for a reasonably significant commitment. While the approach we have taken in attempting to archive Second Life regions produces results which are extremely partial and imperfect, they do have the advantage that, having secured the permissions of the copyright owners, much of the material can be made available today. An optimal solution might involve some combination of a commitment of a complete copy of a region in Second Life at a given point in time by Linden Lab with an attempt by the archiving institution to employ mechanisms similar to the ones we developed to try to secure permission to make some or all of creators’ content available through the archive on some terms. While our experiences in negotiating with rights holders in Second Life indicate that content creators are unlikely to approve any actions which result in their content being made available in a way such that others might be able to commercially exploit it, they might be amenable to having an archived region be viewable on the premises of an archive, with the understanding that the archive’s users would be unable to make copies of the archived objects.

There are also a number of technical issues which need to be resolved if archiving of virtual environments such as Second Life is going to occur at any scale. While our project has managed to exploit existing standards for archival information packaging and metadata, such as OAI-ORE and OWL, to create submission information packages for Second Life regions, the XML files for individual objects and for the regions are not standardized. There are emerging standards for exchange of 3D assets between virtual environments, such as the COLLADA schema (Barnes & Levy Finch, [2008](#)), which the participants in the Preserving Virtual Worlds project intend to test to determine how successfully they allow virtual objects to retain their significant properties when moved between virtual environments.

The curation of materials in these worlds also relies on the generation of significant amounts of metadata, and for our experiments, most of this metadata were created manually using an XML editor. This is an awkward tool for curators attempting to record additional information about both the worlds they are curating and the objects within them. There is a great deal of research that needs to be done to

improve user interfaces for data curation in general, and virtual environments are no exception. Tools that would allow curators to easily add annotations to 3D environments while in-world, and simplify the task of adding and reviewing context and representation information for an archived world would be of immense value.

Given all of the impediments we identified to the preservation of virtual environments such as Second Life, and the likelihood that archiving of these materials will require substantial investment on the part of cultural memory organization, it is reasonable to ask whether they are worth the trouble. We believe they are. As noted previously, despite their relative novelty, such environments are already well established within popular culture and are beginning to assume greater importance within people's lives. Given the impact that virtual environments, such as Second Life, have already had in fields such as public health education (Boulos, Maged & Toth-Cohen, 2009), therapeutic applications (Kim et al., 2008; Reger & Gahm, 2008) and national security (Subrahmanian & Dickerson, 2009; Hale, Stanney & Malone, 2009), we can safely assume that they will become increasingly significant to the education and research communities as well. Given this, a commitment by libraries, archives and museums to preserve these worlds seems well justified, and further research on how to preserve these resources is vital.

Acknowledgements

The research described in this article was made possible by a grant of the Library of Congress' National Digital Information Infrastructure and Preservation Program.

References

- Bainbridge, W.S. (2007). The scientific research potential of virtual worlds. *Science*, 317, 472-476. [doi:10.1126/science.1146930](https://doi.org/10.1126/science.1146930)
- Barnes, M. & Levy Finch, E. (Eds.). (2008). *COLLADA: Digital Asset Schema Release 1.5.0 Specification*. Clearlake Park, CA: Khronos Group; Tokyo, Japan: Sony Computer Entertainment. Retrieved June 25, 2011, from http://www.khronos.org/files/collada_spec_1_5.pdf.
- Book, B. (2004). Moving Beyond the Game: Social Virtual Worlds. In *Proceedings of the State of Play 2 Conference, New York, NY*. Retrieved June 25, 2011, from <http://deby.net/FILES/3d/ARTICLES/moving%20beyond%20the%20game%20-%20social%20virtual%20worlds.pdf>.
- Boulos, K., Maged, N., & Toth-Cohen, S. (2009). The University of Plymouth Sexual Health SIM experience in Second Life: Evaluation and reflections after one year. *Health Information and Libraries Journal*, 26, 279-288. [doi:10.1111/j.1471-1842.2008.00831.x](https://doi.org/10.1111/j.1471-1842.2008.00831.x)
- Clark, D. (2008). Content management and the separation of presentation and content. *Technical Communication Quarterly*, 17, 35-60. [doi:10.1080/10572250701588624](https://doi.org/10.1080/10572250701588624)



comScore (2009). *Game on! Online gaming surges as gamers seek out free alternatives in tight economy* [Press release]. Retrieved June 25, 2011, from http://www.comscore.com/layout/set/popup/Press_Events/Press_Releases/2009/1/Online_Gaming_Grows.

Copyright law of the United States and related laws contained in Title 17 of the United States Code. (2009). Retrieved June 25, 2011, from <http://www.copyright.gov/title17/circ92.pdf>.

Entertainment Software Association. (2010). *Essential facts about the computer and video game industry: 2010 sales, demographic and usage data*. Retrieved June 25, 2011 from http://www.theesa.com/facts/pdfs/ESA_Essential_Facts_2010.PDF.

Fishwick, P.A. (2009). An introduction to OpenSimulator and virtual environment agent-based M&S applications. In *Proceedings of the 2009 Winter Simulation Conference, Austin, TX*. Retrieved June 25, 2011, from <http://www.informs-sim.org/wsc09papers/015.pdf>.

GamePolitics.com (Oct. 9, 2008). *Report: Obama ads in burnout paradise* [Web log message]. Retrieved June 25, 2011, from <http://www.gamepolitics.com/2008/10/09/report-obama-ads-burnout-paradise>.

Hale, K.S., Stanney, K.M. & Malone, L. (2009). Enhancing virtual environment spatial awareness training and transfer through tactile and vestibular cues. *Ergonomics*, 52, 187-203. doi:10.1080/00140130802376000

Horn, D., Cheslack-Postava, E., Azin, T., Freeman, M.J. & Levis, P. (2009). Scaling virtual worlds with a physical metaphor. *IEEE Pervasive Computing*, 8(3), 50-54. doi:10.1109/MPRV.2009.54

§

Hswe, P., Kaczmarek, J., Houser, L. & Eke, J. (2009). The Web Archives Workbench (WAW) tool suite: Taking an archival approach to the preservation of web content. *Library Trends*, 57, 442-460. doi:10.1353/lib.0.0046

IFLA Study Group on the Functional Requirements for Bibliographic Records (2009). *Functional requirements for bibliographic records: Final report*. The Hague, Netherlands: International Federation of Library Associations and Institutions. Retrieved June 25, 2011, from http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf.

Institute for Information Law & Policy (2010). *Do Tank and the Democracy Design Workshop: Ideas for democratic action*. New York Law School. Retrieved June 25, 2011, from <http://dotank.nyls.edu/DemocracyIsland.html>.

- Kim, S.I., Ku, J., Han, K., Hyeongrae, L., Park, J., Kim, J.J. & Kim, I.Y. (2008). Virtual reality applications for patients with schizophrenia. *Journal of CyberTherapy and Rehabilitation*, 1, 101-112.
- Lee, C.A. (2011). A framework for contextual information in digital collections. *Journal of Documentation*, 67, 95-143. doi:10.1108/002204111111105470
- Library of Congress. (2002). *Preserving our digital heritage: Plan for the National Digital Information Infrastructure and Preservation Program*. Retrieved June 25, 2011, from http://digitalpreservation.gov/library/resources/pubs/docs/ndiipp_plan.pdf.
- Linden, P. (2006). The mission of Linden Lab [Web log message]. Retrieved June 25, 2011, from Second Life Blogs: <http://community.secondlife.com/t5/Features/The-Mission-of-Linden-Lab/ba-p/533170>.
- Linden Lab (2010). Terms of service: Second Life. Retrieved June 25, 2011, from <http://secondlife.com/corporate/tos.php>.
- Masanès, J. (2002). Towards continuous web archiving: First results and an agenda for the future. *D-Lib Magazine*, 8(12). doi:10.1045/december2002-masanés
- McDonough, J. (2000). *Under construction: The application of a feminist sociology to information systems design*. (Doctoral dissertation, University of California at Berkeley).
- Reger, G.M. & Gahm, G.A. (2008). Virtual reality exposure therapy for active duty soldiers. *Journal of Clinical Psychology*, 64, 940-946. doi:10.1002/jclp.20512
- Reilly, B., Palaima, C., Norsworthy, K., Myrick, L., Tuchel, G., & Simon, J. (2003). Political communications web archiving: Addressing typology and timing for selection, preservation and access. In J. Masanès, A. Rauber, & G. Cobena (Eds.), *Proceedings of the 3rd Workshop on Web Archives, Trondheim, Norway, in conjunction with the 7th European Conference on Research and Advanced Technologies for Digital Libraries, ECDL 2003*. Retrieved June 25, 2011, from <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=reilly>
- Siwek, S.E. (2007). *Video games in the 21st century: Economic contributions of the US entertainment software industry*. Washington, DC: Entertainment Software Association. Retrieved June 25, 2011, from <http://www.theesa.com/facts/pdfs/VideoGames21stCentury.pdf>.
- Smith, A. (2006). Distributed preservation in a national context: NDIIPP at mid-point. *D-Lib Magazine*, 12(6). doi:10.1045/june2006-smith



Subrahmanian, V.S. & Dickerson, J. (2009). What can virtual worlds do for national security? *Science*, 326, 1201-1202. [doi:10.1126/science.1182660](https://doi.org/10.1126/science.1182660)

Wu, J. (2010). *Global video game market forecast*. Newton, MA: Strategy Analytics.