

A Blueprint for Preserving Virtual World Cultural Heritage using Preservica & Custom Metadata Schema

Marie Vans¹ and Patricia Franks²; ¹HP Labs, ²San José State University, iSchool; ¹Fort Collins, CO, ²San José, CA

Abstract

Faced with the possibility of running out of storage space and losing important born-digital cultural heritage artifacts, we are embarking on a project to design a process for preserving objects generated by our virtual world-based group so that these issues will not become critical in the future. Because of the high-turnover nature of student-based groups, preservation of objects associated with annual and learning events is essential for continuity in the community. We present our blueprint for preserving these artifacts using a combination of an established archiving system together with a customized metadata schema that reflects the specific needs of VCARA and an easy to use interface for browsing collections.

Introduction

The Virtual Center for Archives and Records Administration (VCARA) is a group of current and former students, along with faculty from the San José State University iSchool. This group regularly meets and organizes events such as conferences, colloquiums, and virtual exhibits. This year marks the 10th anniversary of VCARA in the virtual world of Second Life. ^[1] Currently there is a collection of ten years' worth of VCARA objects including documents (meeting notes, policies, images, event documentation, contacts, social media posts, etc.) and podcasts videos and audio recordings of events and exhibits. Most of these artifacts can be viewed as representations of VCARA's born-digital cultural heritage, ^[2] and as such we are concerned with making sure we are able to persevere them. Until recently, most of these important objects were stored in a Google drive and associated with the VCARA group or distributed amongst several different social media sites (Facebook, Twitter, YouTube, Pinterest). Additionally, there is a iSchool server on which older, multi-media materials from 2010-2013 are stored.

As we pointed out in 2015 ^[3], the use of free accounts such as Google Drive and social media can be risky because sites can close without notice and/or free space limits imposed can be so low that they are easily reached, after which a cost will be associated with continued use of the service. Another issue we did not anticipate was the loss of data due to the fact that student participants often own files of interest, which they share with the group while actively involved but may delete when they withdraw from the group.

These issues led us to investigate and design a more robust approach to persevering these cultural heritage objects. In this paper, we will describe a solution based on the digital preservation system, Preservica^[4] along with metadata schema specifically designed for virtual world artifacts. While we acknowledge the fact that this company can also go out of business in the future, the experience of designing and implementing a system for preserving our virtual world cultural heritage will be applicable to other born-digital collections and digital preservation systems.

Motivation

Previously, we presented on our approach to preserving virtual world educational activities using a combination of Google Drive storage, social media networks, and video ^[3]. Over the years, we have used up the available free space on Google Drive as content has grown. We were faced with a choice: continue using Google for storage and start paying for space used beyond the free 15 gigabytes or find a longer-term solution to preserve documents, events, and other artifacts. This is crucial, not only for helping new students understand the VCARA culture but to preserve the work done by students who have left and are no longer participating in VCARA activities.

Problem

In addition to paying for storage, we faced a more pressing problem: the possibility of losing important artifacts because they were owned by former students. Google drive allows different files to be owned by different users under the same directory structure. Students take on different roles during the semester such as meeting scribe, advertising director and cameraperson. Each artifact they create under these roles end up in the shared directory but are actually owned by the student. These files are counted towards the 15-gigabyte limit for the student's account and in the event they are no longer participating in VCARA activities and bumping up against the space limit, they may delete files they no longer need, resulting in a loss of the object for VCARA. What we needed was a way to preserve artifacts in a central location, with strict controls on permissions, and in a manner that makes it easier for future students to utilize.

Background

There have been several attempts at preservation of objects associated with virtual worlds ^[5,6] and related concepts such as virtual museum exhibits ^[7] and game environments ^[8]. Most are concerned with preserving the objects used in the virtual world, exhibit, or game. McDonough et. al specifically documented inworld activities by preserving machinima (video recordings of events in virtual worlds). ^[9] We are concerned with a higher-level view, which includes typical Word-type or text documents, images in .jpeg or .png formats, podcasts, and video and audio of events that occur in-world. Because the environment in which most activities occur is a proprietary system, preserving the objects that exist in the world is beyond the scope of our project. It is enough to document how we incorporate customs and practices formed inside the virtual world.

Preservica

Preservica is a proprietary archiving system that supports OAIS-compliant workflows. It automatically inherits the security and redundancy provided by Amazon Glacier storage services as Preservica is built on top of it. Four years ago, when we first looked at it as a possible solution for preserving VCARA artifacts, it was more expensive than paying for an expansion of Google Drive space. At that time, the most cost-effective product was

\$3,950.00 per year for 100 gigabyte of data storage and for 250 gigabytes, the cost was \$6,950 [3]. Fortunately, a collaboration between Dr. Franks and Preservica, resulted in access to Preservica [4] for course projects taught at the iSchool. This access opened up the possibility for developing a prototype to demonstrate our approach to archiving VCARA artifacts and making them available for future use. This relationship allowed us to try the system while supplying feedback for improvements. It will also allow us to develop an example of a potential solution to the problem of preserving virtual world cultural heritage that can be used to obtain funds for a dedicated solution for VCARA.

Preservica addresses many challenges that archiving of digital assets present, including support for migration of obsolete media and file formats, user friendly data management, and automated data and metadata ingestion as well as a customizable interface for use by internal and external users. Because an archival process is already built-in, the time needed for setting up a process from scratch is greatly reduced. Preservica also has a metadata template creation function that allows us to create a customized schema based on, or on top of, a well-known schema such as Dublin Core. Since this is the first effort of which we are aware that defines a metadata schema for virtual world cultural heritage objects, this work should be useful for other virtual world groups.

Metadata Schema

When ingesting material into Preservica, basic and technical metadata can be automatically attached to each component. We are interested in capturing different types of information for each component, depending on the function of the artifact. However, the default metadata generated is not sufficient for our needs. Not only do we need additional common metadata, such as that provided by the Dublin Core Schema [10] (Title, Creator, Subject, Description, etc.) but also additional information such as location of object within the virtual world, images of objects, as well as links to objects, such as PowerPoint presentations, photoshop files, images, and team documents. We developed a special set of metadata schemas based on a list of object types, both inworld and associated artifacts that represent activities outside of the VCARA Island. These types are shown in table 1.

Table 1. Object Types for Metadata Schemas – Inworld & Supporting materials. Examples shown in Appendix A.

Object Types – Inworld	Inworld Example
Media on Prim	Webpage Access to blog
Book	Interactive/page turning
Signage	Current events, talks
Slide Set	Past Conference Presentations
Furniture	Desks, podiums, chairs, tables
Plants	Trees, flowers, etc.
Other Décor	Pictures, paintings
Container	Holds other objects
Information Dispenser	URL, Notecard, Landmark givers
Building/Structure Components	Entire buildings and building components
Object Types-Supporting	Example
Directory	Team Drive Directory
Microsoft Word Document	Meeting notes, forms, working documents

Google Document	Team Collaboration Docs
Adobe Photoshop	Inworld signage, marketing
JPG Image	Information, marketing
PNG Image	
TGA Image	
Image Other	
Microsoft PowerPoint File	Inworld presentations
Google Slides	Team Collaboration Docs
Plain Text File	Notecards, Speakeasy
.MPEG4 Video File	Machinima
.MP3 Audio File	Inworld presentations
PDF File	Information/Team Documents
Microsoft Excel Spreadsheet	Track Inventory

We have created a custom metadata scheme for each of these Object types. For ingestion, Preservica includes a SIP (Submission Information Package) that will automatically create the metadata for each file that has an associated .metadata text file present in the folder containing the object. Figure 1 shows an example of a single JPEG image and the customized metadata ingested using this method. For each object type, we have created an XML file for use as a template. Figure 2 shows one template for the Media on Prim object type and figure 3 shows the XML file for a Microsoft Word Document type. Most customization can be seen using the <dc:description> tags.

Universal Access

Finally, an important goal for this project is the need to make public content stored in Preservica visible to external audiences. Although Preservica allows clients to connect to their own website, we elected to use their out-of-the-box solution for a WordPress [11] front-end that allows users to browse collections. Administrators can select which components in a collection are viewable. Layouts are customizable to create browsing and searching experiences that are based on user need. Functionality such as streaming of multi-media files and full screen image and document vitalization allows detailed image viewing. Figure 4 shows the front page of the VCARA Team Collection. Using the search bar at the top of the page allows users to access to content based on collections, metadata, content type. For example, figure 5 shows an example collection gallery from one of the courses taught on the VCARA island. Figure 6 demonstrates what a user would see in WordPress for the example Preservica file shown in figure 1. Note that the metadata ingested by Preservica shows up at the left of the page. This functionality allows, for example, students to look for specific types of documents using search rather than digging through the files in the TEAM Drive on Google.

Results

We are currently finishing up the Preservation Plan for VCARA materials that date back 10 years. The examples shown here were used to familiarize ourselves with the environment and to determine whether the system would fit our needs. As part of the planning process we have defined Ingestion workflows that use our customized metadata schema specifically addressing virtual world learning community artifacts. We are currently working on finishing up Access, Preservation, and Data Management workflows. We expect to have all workflows operational by the end of the 2019-2020 academic year (May 2020). Besides finishing the workflow definitions, we are finishing up the process

of ingesting the rest of the materials from all 10 years and adding metadata.

Conclusions and Further Work

VWs are not games, although they have been compared to them for the purposes of documenting events and preserving artifacts. The turnover of students in virtual world communities means that either the wheel must be reinvented for future students in terms of processes and documents or we can preserve the things that work and let future students learn from them. We are working on a blueprint for preserving cultural heritage artifacts that other virtual world learning communities can leverage. We have shown how we've created custom metadata templates specifically for use in this environment. We have also demonstrated how Preservica ingests this metadata along with the artifacts. We plan to have the rest of the processes defined and operational by May of 2020. Completion of this process will provide us with a proof-of-concept we can use to seek funding to support a dedicated digital repository for virtual world cultural artifacts and to identify virtual world partners interested in preserving and sharing their virtual world cultural heritage with others.

References

- [1] Second Life. Homepage. <https://secondlife.com/>. Accessed 10/16/18.
- [2] Ruan, J., & McDonough, J. P. Preserving born-digital cultural heritage in virtual worlds. In *IT in Medicine & Education*, 2009. ITIME'09. IEEE International Symposium on (Vol. 1, pp. 745-748). IEEE. (2009).
- [3] Franks, P. C., & Vans, A. M. (2015, May). Preserving Virtual Worlds Educational Events using Social Media Networks and Cloud Storage Services. In *Archiving Conference* (Vol. 2015, No. 1, pp. 56-59). Society for Imaging Science and Technology.
- [4] Preservica. Homepage. <https://preservica.com/>. Accessed 10/16/18
- [5] Antonescu, M. D., Guttenbrunner, M., & Rauber, A. (2009). Documenting a virtual world-A case study in preserving scenes from Second Life. In *The 9th International Web Archiving Workshop (IWAW 2009)*.
- [6] Pittarello, F., & De Faveri, A. (2006, April). Semantic description of 3D environments: a proposal based on web standards. In *Proceedings of the eleventh international conference on 3D web technology* (pp. 85-95). ACM.
- [7] Foo, S. (2010). Online virtual exhibitions: Concepts and design considerations. *DESIDOC Journal of Library & Information Technology*, 28(4), 22-34.
- [8] McDonough, J. (2013). A tangled web: Metadata and problems in game preservation. *The Preservation of Complex Objects*, 3, 49-62.
- [9] McDonough, Jerome P., Robert Olendorf, Matthew Kirschenbaum, Kari Kraus, Doug Reside, Rachel Donahue, Andrew Phelps, Christopher Egert, Henry Lowood, and Susan Rojo. *Preserving virtual worlds final report*. 2010.
- [10] Dublin Core Metadata Initiative. Homepage. <http://dublincore.org/>. Accessed October 16, 2018.
- [11] Wordpress.org. Meet WordPress. Homepage. <https://wordpress.org/>. Accessed October 16, 2018.

Author Biography

Marie Vans is currently a Research Scientist with Hewlett-Packard Labs in Fort Collins, Colorado. Her main interests are security printing and imaging for document workflows, statistical language processing, and other approaches to document understanding. She holds a Ph.D. in Computer Science from Colorado State University. She also recently completed a second master's degree in Library and Information Science at San José State University.

Dr. Patricia C. Franks, Professor in the School of Information, is a Certified Archivist, Certified Records Manager, and Information

Governance Professional who coordinates the Master of Archives and Records Administration (MARA) degree program. She was the creative force behind the design and development of VCARA in Fall 2009. As administrator, Pat supervises student volunteers, special studies students, and interns eager for experience in all aspects of managing the Virtual Center. She was the principal investigator on a San Jose State/California State University grant, "Public records-public trust: Reclaiming history," which resulted in the first VCARA annual conference. She coordinates the Virtual Worlds Guest Lecture series as part of the iSchool's Webinar offerings. Dr. Franks is co-author with Brande Gex of a chapter included in *Virtual Worlds for Online Learning: Cases and Applications* (2015), "SLIS Island 360⁰—Is There Value to Virtual Worlds?" She is co-editor, along with Lori A. Bell and Rhonda B. Trueman, of the 2016 publication, *Teaching and Learning in Virtual Environments: Archives, Museums, and Libraries*.



Figure 1. Preservica ingested file: Dublic Core

```

<?xml version="1.0" encoding="UTF-8"?>
- <oai_dc:dc xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ oai_dc.xsd">
  <dc:title>Name of Media On Prim Object</dc:title>
  <dc:creator>Mnemosyne Seminario</dc:creator>
  <dc:subject>Object Category</dc:subject>
  <dc:description>Building</dc:description>
  <dc:description>Object Name</dc:description>
  <dc:description>Group</dc:description>
  <dc:description>Land Impact</dc:description>
  <dc:description>Images in Inventory</dc:description>
  <dc:description>Scripts</dc:description>
  <dc:description>URLs, LMs, or Notecards</dc:description>
  <dc:description>Prim Type (Box, Sculpt, etc)</dc:description>
  <dc:description>Second Life Location (sim name, x, y, z)</dc:description>
  <dc:description>Notes Column</dc:description>
  <dc:description>Who has Access: Access=VCARA Team Drive + SJSU </dc:description>
  <dc:publisher>Mnemosyne Seminario</dc:publisher>
  <dc:contributor>Names of other contributors</dc:contributor>
  <dc:date>2019-02-24</dc:date>
  <dc:type>Media On Prim</dc:type>
  <dc:identifier>UUID/ID</dc:identifier>
  <dc:source>Team Drive Image Location</dc:source>
  <dc:language>English.</dc:language>
  <dc:rights>San Jose State University, iSchool</dc:rights>
</oai_dc:dc>

```

Figure 2. XML Template for Media-on-Prim Object Type

```

<?xml version="1.0" encoding="UTF-8"?>
- <oai_dc:dc xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/elements/1.1/"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ oai_dc.xsd">
  <dc:title>Word Doc Template</dc:title>
  <dc:creator>Mnemosyne Seminario</dc:creator>
  <dc:subject>Conferences</dc:subject>
  <dc:description>Use Cases</dc:description>
  <dc:description>Notes Column</dc:description>
  <dc:description>Directory Contained in</dc:description>
  <dc:description>Who has Access: Access=VCARA Team Drive + SJSU </dc:description>
  <dc:publisher>Owner</dc:publisher>
  <dc:contributor>Names of other contributors</dc:contributor>
  <dc:date>2019-02-24</dc:date>
  <dc:type>Microsoft Word Document</dc:type>
  <dc:format>.doc</dc:format>
  <dc:identifier>string</dc:identifier>
  <dc:source>https://docs.google.com/document/d/1p0941z9FXfZFG09nbxI4ADtHyZ0A3QKu23UmgmxMAU/edit</dc:source>
  <dc:language>English.</dc:language>
  <dc:rights>San Jose State University, iSchool</dc:rights>
</oai_dc:dc>

```

Figure 3. XML Template for Microsoft Word Document



Figure 4. WordPress front page for Collections browsing & access

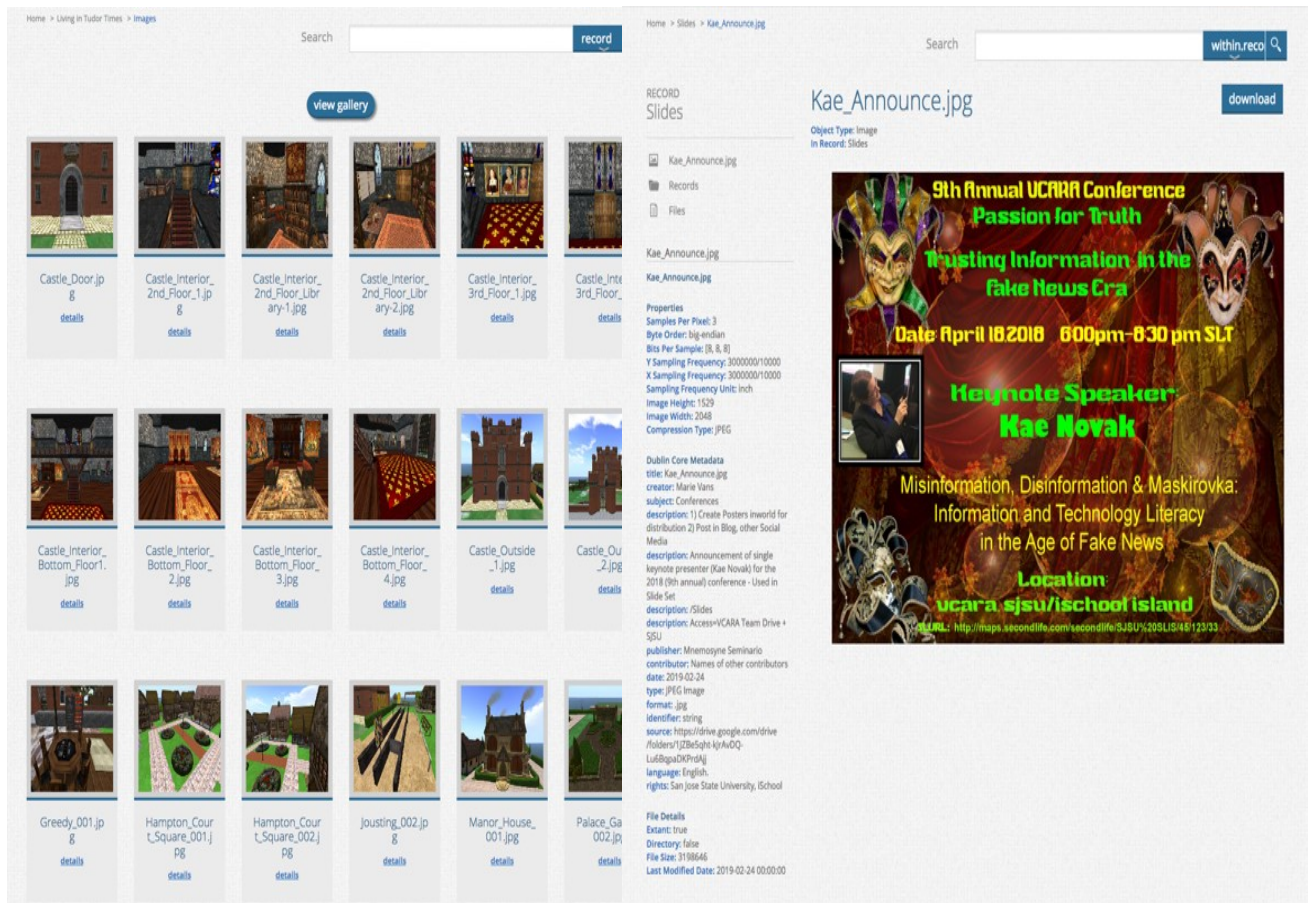


Figure 5. WordPress example collection gallery

Figure 6. WordPress version revealing metadata ingested through Preservica