# Museums and The Cloud

Lauren Marx

laurenanibas@gmail.com

Johns Hopkins University, Museum Studies Digital Curation Certificate Program

December 10, 2017

Fall 2017

## Abstract

Should museums be utilizing cloud storage for their digital assets? There are several deployment models (public, private, community, hybrid) and service models (software, platform, infrastructure, and backup) all of which have their own advantages and disadvantages for museums. This paper examines a few cloud service providers (Amazon, Shared Shelf, Preservica, and ResourceSpace) along with other forms of storage on-site (NAS, SAN, and DAS) that are available. By understanding what a museum needs for digital preservation of their digital assets it will allow the museum to determine if the cloud would be appropriate for their digital collection. This paper describes whether and how several museums have decided for or against cloud services (Wisconsin Veterans Museum, Philadelphia Museum of Art, Minnesota Historical Society, Harvard University, and the Walters Museum of Art in Baltimore, Maryland). Analysis of how the museums use the cloud, alternative solutions, and what museums should be asking before using the cloud, such as legal restrictions and types of digital assets being stored, will provide some insight and help museums decide if the cloud is a plausible option.

Table of Contents:

Introduction 4

Literature Review 4

What is the Cloud 4

Deployment Models 6

Service Models 9

Advantages and Disadvantages of the Cloud 10

Cloud Options 14

Amazon 14

Shared Shelf 16

Preservica 17

ResourceSpace 18

Besides the Cloud, What is There 18

Museum Needs for Digital Preservation 20

Research Method 22

Interview: Wisconsin Veterans Museum 22

Interview: Philadelphia Museum of Art 23

Interview: Minnesota Historical Society 24

Interview: National Gallery of Art in Washington 24

Case Study: Harvard University 25

Case Study: Walters Art Museum in Baltimore Maryland 25

Analysis 26

Alternative Solutions 26

Is Cloud the Way to Go? 27

What Museums Should Ask Themselves First 27

Conclusion 28

References 30

Appendix A_____35

Appendix B_____36

Appendix C_____37

"The cloud": we hear this technology buzz word but what does it mean? The cloud is essentially the internet. Cloud computing is storing and accessing data and programs over the internet instead of the computer's hard drive. How does this relate to the museum field? As digital assets grow within museums there is a need to find solutions on how to store the digital assets they are gathering. In this paper I examine if museums should be utilizing cloud services compared to the other options available for the digital assets.

I address what cloud services are currently out there, their advantages and disadvantages, and look at some cloud providers. I also talk about what makes cloud services a good option for a museum, or why they do not. There are many options out there for the cloud and other services that it can get confusing and difficult for a museum to decide on the right path for it. It is important that museum staff understand and use the best storage for their digital assets. My aim is to provide a better understanding of the assorted options available, their advantages and disadvantages, what museums should be looking for, examples from current museums, and how to make the right choice.

## Literature Review

We hear the term "cloud" in reference to computers or technology, but what exactly does that mean, what does it do, and how does it work? Cloud computing refers to storing, accessing, and sharing digital assets through the internet. The digital assets are stored on physical servers, which are then maintained and controlled by a cloud computing provider (Fastmetrics, 2017, para. 3-4). The servers managing the cloud computing system will also manage the traffic and what the customer needs, such as types of data, security, and so on, so that everything runs smoothly for the institution. The servers use a kind of software called

middle ware, which allows for network computers to communicate with each other (Fastmetrics, 2017, para, 27). This gives the institution the possibility to access their digital assets anywhere with internet and for more collaborative work.

The National Institute of Standards and Technology (NIST) defines cloud computing as, "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Mell & Grance, 2011, p. 2)." Within the definition there are five essential characteristics of cloud computing: on demand self-service, broad network access, resource pooling, rapid elasticity, and measured service (Ferriero, 2010, para. 5).

On demand self-service is when the cloud services can be left alone without the institution having to interact with the service provider. This allows the institutions to not worry about checking in frequently to ensure the digital assets are accessible. Broad network access means that the cloud services are available from various locations using standard protocols. This gives the institution the ability to have access to the digital assets and resources from any machine (Simmon, 2011, pp. 6-7). Resource pooling is the computing infrastructure that is shared with more than one user, which is helpful to lower costs. The fourth is rapid elasticity, which means that the cloud services can be "rapidly" edited and released to scale (Simmon, 2011, pp. 8-9). This gives the institutions options to expand or shrink the amount of services they need. The last is the measured service. Cloud services are measured with enough detail to support the requirements of the user (Simmon, 2011, p. 10). All of this is what makes up the cloud, but how does the cloud work?

The cloud is not just one type of computing system but has several deployment models and service models available. The most common deployment model for the average person tends to be the public cloud. The public cloud is accessible to the public via public networks. It is owned by a commercial organization that have large data centers around the world and who sell their cloud services (Beagrie, Charlesworth, & Miller, 2015, p. 5). The public cloud tends to be more cost effective when it is used for short term activities, as storage size and additional security measures add additional costs. With the appropriate management it can be effective for long-term activities as well, but that will require more planning, time, and IT staff. This deployment option provides more opportunities for collaboration between and within institutions as they can use the same service, making sharing of resources easier (Ludden, 2014, para. 13).

There are also negative aspects to the public cloud, such as cloud providers often offer one set systems without a lot of customization. The service level agreements and set of terms and conditions are also standardized, all of which can be a problem for museums and their diverse collections (Beagrie, Charlesworth, & Miller, 2015, p. 8). Another negative is that long-term activities are not as well supported within the public cloud. For museums this may be okay for a smaller project, but overall, they will want a deployment model that can handle long-term activities.

The second type of deployment model for the cloud is the private cloud. The private cloud allows institutions to take ideas from the public cloud and then apply those to their own data centers (Beagrie, Charlesworth, & Miller, 2015, p. 5). Larger institutions with large IT staff will benefit from this model because they can take on more workload and adjust for demand in

a timely manner. It is run solely for the institution and can either be managed by the institution itself or by a third party (Radack, 2012, p. 3). This allows for the institution to customize the cloud service more to their needs and collections. The private cloud can also better handle long-term activities, unlike the public cloud. Since the private cloud is accessed over private networks it feels more safe and secure for the institutions (Beagrie, Charlesworth, & Miller, 2015, p. 5).

The disadvantages to using the private cloud deployment model is that it is most effective with institutions that already have significant investment in their data, equipment, and IT staff (Beagrie, Charlesworth, & Miller, 2015, p. 5). It is then more difficult for small institutions or institutions with a smaller budget to use the private cloud. The costs can be higher due to the private nature of customizable services that the institution would be receiving.

The third type is a community cloud, which is shared by several institutions. This type of cloud will support a specific community with shared concerns for their digital collections (Radack, 2012, p. 3). It is typically managed by one of the institutions or a third party. It is like the public cloud service, but access is limited to certain set of users. An example of the community cloud is the version that Amazon provides to the Federal and State Governments of the United States of America, which is shared across departments (Beth, 2017, para. 1). A benefit for this type of cloud deployment is the opportunity for sharing resources and cost of services with other institutions. The institutions can create an information network and share their ongoing work with each other (Goldner, 2010, pp. 273).

The community cloud also has some disadvantages for their users. Since this cloud deployment is shared by several organizations there is the risk of disagreement between organizations on how digital assets should be stored and accessed. There may come a time that an organization wants to leave the community cloud service, which could also be problematic depending on the number of organizations involved. It is important to have contracts and policies in place to prevent any future problems (Chen, 2010, p. 5).

The last deployment model is the hybrid cloud. This is a composition of two or more of the previously mentioned cloud types (Hope, Thornhill, & Carr, 2017). Most of the literature on the cloud deployment models classifies the hybrid as a combination of the public and private cloud, but could potentially include community cloud as well. This allows for institutions to customize services to better fit their needs. The institution would be able to use the large storage and batch abilities for parts of a collection with the public cloud, store more sensitive material in the private cloud, and collaborate with other institutions with the community cloud (Radack, 2012, p. 3). It is important to have a plan set before implementing the hybrid cloud deployment so that it runs smoothly.

The hybrid option removes some of the negative aspects of the other deployment models to create a system that works for the institution. One negative aspect of the hybrid option is that there will be more planning and coordinating with the different deployment models. There will need to be policies that make it clear to the staff what model will be used for what digital assets. The cost may also increase with the use of two different models and potentially two different providers of the services (Hope, Thornhill, & Carr, 2017).

In addition to the deployment models, there are several service models that the cloud makes available for institutions. These are Cloud Software as a Service (SaaS), Cloud Platform as a Service (Paas), Cloud Infrastructure as a Service (Iaas), and Cloud Backup as a Service (BaaS). The first, Software as a Service, is an application the institution accesses over the internet. It is a software application that will allow an institution to perform a group of coordinated functions, tasks, or activities (Radack, 2012, p. 4). The institution will not control the cloud infrastructure (network, servers, operating systems, storage, application capabilities). The institution may have limited control over some application configuration settings. The provider will be responsible for deploying, configuring, maintaining, and updating the operation of the applications (Radack, 2012, p. 4). Some examples of Cloud SaaS would be PastPerfect Online and MuseumPlus (Clarke, 2013, p. 11).

The next service model is Platform as a Service, which allows the institutions to create their own custom applications. The institution can deploy its applications that it created onto the cloud infrastructure. This gives them the ability to create without having to worry about the complexities of managing the cloud infrastructure (Griffith, 2016, para. 7). They will have control over the application configurations and possibly the hosting environment configurations (Chen, 2010, p. 4). This will allow for more customization and give the institution an opportunity to create what they need for their digital assets. It will however require more IT staff and work to get the service up and running to the institutions standards.

The third service model is the Cloud Infrastructure as a Service. This is when companies provide the backbone to be rented out. It allows the institution to control the operating systems, storage, deployed applications, and limited control of select networking components.

It enables the institution to use processing, networks, and other fundamental computing resources (Radack, 2012, p. 4). Essentially the users are taking on a more IT operations role by creating, installing, monitoring, and managing the services and applications that are within the cloud infrastructure. Some examples of an IaaS are Archivist Toolkit, PastPerfect, Argus, and TMS/EmbARK (Clarke, 2013, p. 11).

The last service model is the Backup as a Service. This is exactly as it sounds, a backing up of the system for an institution (Clarke, 2012, p. 11). Instead of conducting a backup on-premises the institution would purchase backup and recovery services from a cloud data backup provider (Rouse, 2016, para. 1). This allows for the institution to not worry about managing hard disks, or other forms of storage, since the service provider will take care of the maintenance. Some examples of an BaaS are Carbonite, SOS Online Backup, and Vembu (Rouse, 2016, para. 15). While this is an option as a cloud service, much of the literature on cloud computing did not include this within their discussions.

After getting a better understanding of what cloud computing is we can analyze the advantages and disadvantages to using the cloud. One of the major benefits is the potential to save money with the pay-as-you-go system. Most, if not all, cloud services have the institution pay for services that they use instead of a flat rate (Moad, Bactha, & Stein, 2009, para. 3). While having large collections can make it costlier, it is beneficial to smaller institutions or institutions that want to use the cloud part of their digital assets. This also allows for easy scalability. The institution can add more space or reduce their space whenever they need to (Clarke, 2013, p. 12).

Access of the digital assets anywhere is another advantage to using cloud computing. If the user has internet they will be able to access the digital assets. It also allows for the user to access the information via different devices, such as computers, phones, and tablets. The cloud is also considered much more user friendly than the alternative storage options (Betts, 2015, para. 21). Since it is more user-friendly staff with less IT knowledge will be able to work with the system much easier. The cloud providers also make it easier for staff by sharing the responsibility of the digital assets and take some of the responsibility if there is any data-loss. When the storage is on-site the IT staff will be held responsible for all digital assets and potential data-loss (Betts, 2015, para. 26).

Cloud computing gives the users the ability for fast deployment. The institution can sign up and use the services almost instantly. Cloud service providers typically do not require yearly contracts making it easier for institutions to try the cloud service. If it is not working for them they can change or cancel the service (Beagrie, Charlesworth, & Miller, 2015, p. 6). This allows institutions to be more adventurous and creative with some of their digital assets if they know it does not have to be long-term. Cloud services can be adapted and changed much easier, such as storage size and the features or applications used. In addition to easy adapting, the cloud conducts its updates and ensures the services are all up to date, whereas with on-premises services it is up to the staff to make sure that everything is up to date (Platz, 2017, para. 4).

Backup of data is another advantage that the cloud offers to institutions, although it should not be the only form of backup. Since the cloud is over a network and not on an actual hard drive it can serve as a form of back up. In addition to all the other advantages there is the

ability to share and collaborate with other institutions. Depending on the service provider there can be many ways to collaborate with other users by sharing data, resources, or creating a project together (Goldner, 2010, pp. 272-3). One example of this is the University of Virginia using Shared Shelf to develop a metadata core, ArcheoCore, for archeology artifacts. They wanted a cloud service so that they could update the database while they are on-site. This project was successful that is gained the attention of Dumbarton Oaks, both have begun discussions on collaborative projects for sharing archeology data (Berenz, Burns, & Stylianopoulos, 2013, p. 7). In addition to these advantages of collaboration, depending on the deployment model chosen, it can help reduce costs and time by sharing those burdens with another institution.

Cloud computing can also have its disadvantages for institutions as well, one of them being internet access and bandwidth. While using the cloud may cut back on costs with the pay-as-you-go system there are additional costs elsewhere that need to be considered (Clarke, 2013, p. 12). Since all the work will be done online, the internet access and bandwidth need to be high enough to handle the amount of people working on it and the amount of data being saved to it. There is also a chance, while rare, that the cloud service could crash meaning the institution would not have access to the data until they were back up and running (Lee, 2017, para. 6).

Institutions need to prepare budgets for the year and cloud services make that difficult. It is important to keep in mind the cost of operation, compliance, security, and migration to/from the cloud when trying to determine the cost of using the cloud (Radack, 2012, p. 4). Some institutions have an issue with the idea of renting the storage space versus buying, but

the long run it may not be the more affordable option. Cloud services overall do not seem to handle long term projects as well as well as they have not been around very long to know how well they can hold digital assets for the long-term. This is something that continues to improve but should be considered by institutions when looking at what is best for their digital assets (Smith, 2014, p. 7).

While most cloud providers ensure standard user rights of ownership there are some that include language in their contracts that give them more control over the user's data. There is also a transfer of control over the digital environment that will now be handled by the cloud provider instead of within the organization (Instrumental, Inc., 2013, p. 6). The institution and the cloud provider will have to work together and coordinate more to ensure things are handled properly. There is also a debate on who owns the digital assets stored and whether there is a difference between the digital assets uploaded, and the digital assets created within the cloud (Griffith, 2016, para. 25). Institutions need to include this within the contracts so that they are given ownership of all digital assets (Beagrie, Charlesworth, & Miller, 2015, p. 8).

Another disadvantage, which is most commonly brought up, is the concerns over security. Many institutions worry that the cloud server could be hacked and could then corrupt their data, or the hackers gain access to sensitive data (Chen, 2010, p. 4). Now that the data is no longer on- premises the institution will have to think about how to protect the digital assets technologically versus physically. Some of the threats that institutions should be aware of are data breeches, hijacking of accounts, malware injection, and data loss (Mattoo, 2017, pp. 46-7). Cloud providers know security better than museums and provide security measures against these attacks. While providers, such as Amazon, may be attacked more they also have the

knowledgeable staff that know how to block the attacks (Chen, 2010, p.7). Institutions will still need to consider additional backup options in case any of these threats occur. Along with the other disadvantages previously mentioned, there will be a new skill set that will have to be learned and kept up to date on by the institution (Clarke, 2013, p. 12). While it may be very helpful these skills will take time and resources to acquire them. These traits apply to most of the cloud providers, but each cloud provider is different and emphasize different areas in their services.

Amazon is one of the bigger and more popular cloud service providers. Amazon Web Services are also often used by other cloud providers. It provides several services, but the main two options for storage are, Amazon S3 and Glacier. S3 is meant for institutions with large collections as it offers options with over 5000 TiB of data. One TiB is approximately 1.1 terabyte (TB) of digital information. S3 also provides an import and export tool that will make uploading digital assets easy (Instrumental, 2013, 10-1). It offers redundant storage in either multiple facilities or on multiple devices. It calculates checksums on all network traffic to find any possible corruption of data, in addition they also perform regular checks on the data. A checksum is the outcome of running an algorithm on a piece of data. Comparing the checksum that is generated from the museums version with the one provided by the source of the file. It helps to make sure that the copy of the file is authentic. S3 offers a variety of authentication mechanisms, rights-granting options, and encryption options (Instrumental, 2013, 10-1). It allows the institution to also apply their own encryptions if needed. S3 restricts the access of the digital assets to the owners, unless specified. They will provide the owner with audit

reports and access logs so that they can see what digital assets are being accessed, how often, and more (Instrumental, 2013, 10-1).

Amazon S3 offers their users a discount if the message, “Internal Error” or “Service Unavailable” appears and the uptime drops below 99.9%. While this additional support is beneficial and keeps Amazon working hard to ensure access, there are still some disadvantages to using S3 (Instrumental, 2013, 10-1). One major disadvantage is that Amazon will not be liable for any data loss. An advantage to many of the other cloud providers is that they will take some of the responsibility if anything were to happen to the institutions data. Larger cloud providers can store multiple copies in different location, whereas some smaller providers cannot, so they are more likely to take some of the responsibility if anything were to happen to the data. With Amazon S3 the institution is responsible for the maintenance and security of their digital assets (Instrumental, 2013, 10-1).

Amazon Glacier is better for long-term storage and retrieval for large data sets. While it can store as much or more than S3, retrieval can be an issue. It can take up to 3 to 4 hours to get access to the digital assets and then they are available for 24 hours. Within the Glacier archives the digital assets can be added, deleted, and read but cannot be edited (Instrumental, 2013, 10-1). This means that anytime an institution would need to make an edit on a digital asset they would have to do so on their device, then upload it to Glacier, and delete the older version. Each time the institution uploads a new file they will receive a unique archive ID and will most likely be charged as Amazon typically charges per upload (Instrumental, 2013, 10-1). While time consuming and frustrating, Amazon claims they have this system to prevent

accidental tampering of the digital assets. Overall Amazon provides reliable services with some tools in place to ensure accessibility and authenticity of the digital assets.

A second cloud service is Shared Shelf, which was developed by ArtStor. It is a cloud based media file management System as a Service. It is a cloud service that for all art collections focuses on the cataloging and file management. With Shared Shelf, the institution will be able to go from start to finish with their digital assets by creating metadata schemas, catalog the data, upload and store additional files, along with publish and share the content (Berenz, Burns, & Stylianopoulos, 2013, p. 1). Shared Shelf allows for more customization with the records while also providing various controlled vocabularies to ensure best practices. One benefit to Shared Shelf is that it gives the institution the option to set up access points for various users. Permission for editing and level of access can be assigned on a user by user basis (Berenz, Burns, & Stylianopoulos, 2013, p. 3). This allows for head of departments or administrative positions to have more control over content being added. Shared Shelf incorporates cataloging standards like VRA Core and Dublin Core along with using Cataloging Cultural Object (CCO) guidelines (ArtStor, n.d., Manage, Catalog, and Share).

Shared Shelf provides institutions with the chance for collaboration with others. It can be difficult for institutions to collaborate if they have different databases due to the variety of options out there. By using a cloud based service they will be able to share project templates, which can save resources since it is already set up. An institution would be able to use all or part of another institutions template to start their project and then customize further if they need to (Berenz, Burns, & Stylianopoulos, 2013, p. 1). The information within Shared Shelf can also be easily sent to Omeka, via an add-on that they provide, where the institution can then

create their own virtual exhibition. Shared Shelf currently stores multiple copies of user created data in geographically different server locations. They will soon change their configuration to store multiple copies of user created data on both Amazon Web Services and their geographically-dispersed physical server centers. By adding the Amazon Web Services it will strengthen their services by providing to advantages of Amazon while still giving the museums the personal services that Shared Shelf has. While Shared Shelf works with the institutions and allows for a certain level of customization, it does only work with images and the data related to the images. It would not be a service that would provide storage and access to other digital assets that an institution may have.

A third provider is Tessella Preservica, which focuses on data archiving and long-term preservation. Preservica uses Amazon Web Services (AWS) so it can offer the storage capacity of over 5000TiB, like S3 and Glacier. Data preservation is claimed to be guaranteed by Tessella due to their regular checks on the digital assets (Instrumental, 2013, 12-3). They have high data integrity by creating multiple copies and store them in multiple data centers. The data is then checked against each other and uses a combination of checksums, which can occur at custom intervals or when the data is retrieved (Instrumental, 2013, 12-3). This method of multiple copies is based on the LOCKSS (Lots of Copies Keep Stuff Safe) system. LOCKSS wants to decentralize and distribute the digital assets incase of technological, economic, and social failures (LOCKSS, Preservation Principles).

Preservica also has cyclical redundancy to check for any corruption in addition to the other checks for integrity. Since Preservica uses AWS, they provide the same encryption options at the other Amazon services provided. With Preservica the user controls the access to

the data and the metadata (Instrumental, 2013, 12-3). They can manage and store the digital assets with metadata and data tags, allowing for better and easier searching. One last beneficial feature of Preservica is that they provide tools to migrate files away from obsolete formats to ones that will remain accessible (Instrumental, 2013, 12-3).

The final cloud company examined is ResourceSpace which is a web-based Digital Asset Management Software. It is open source so there are no license fees or vendor lock ins. While ResourceSpace is offered via the cloud there are options to install the software on the computers as well. With this provider you can install the Amazon EC2 package, which allows user to obtain and configure capacity with minimal friction, but they do not use Amazon S3. Some of the features that they provide is asset sharing and collaboration (Capterra, 2016, ResourceSpace). The institution will be able to share their digital assets and collaborate with other institutions, researchers, or the public a lot easier. ResourceSpace provides cataloging and categorization of the digital assets (Capterra, 2016, ResourceSpace). The categorization along with the variety of searching methods makes it user-friendly and easy to access the digital assets. It also provides metadata management, image editing, and video management. Having these services all within one system will be more convenient and save the institution time (Capterra, 2016, ResourceSpace).

Another benefit of ResourceSpace is that they provide usage tracking and analytics so that the institution can see what digital assets are being accessed most often, which can help them decide what types of digital assets they should be preserving or keeping long-term (Capterra, 2016, ResourceSpace). There are a few negative aspects of ResourceSpace, most of which seem to be the interface. Many reviews of the DAMS mentioned that it was easy to click

the wrong thing and not as straightforward or easy to understand (Capterra, 2016, ResourceSpace). Information on additional cloud providers and comparisons is provided in an appendix.

Besides the Cloud, what else is there for institutions to use so they can have access and storage of their digital assets? One option is Network Attached Storage (NAS) which mounts itself on a local network (Lee, 2017, para. 9-10). It combines the local hard drive and cloud storage as it includes a processor, memory, and space for hard drive storage that is all connected to a local network (Betts, 2015, para. 4). The NAS can be accessible over the internet or the institution can restrict access to just its network for extra security. NAS comes with several slots for more than one hard drive, which allows for backup. It also gives the institutions room for quick expansion of their storage (Betts, 2015, para. 22).

Some of the benefits to NAS is that the institution will physically own the drive that the data is stored on, unlike the Cloud which is owned by a third party. It comes in a variety of prices and complexities making it beneficial in the museum field where the needs of the institution vary (Betts, 2015, para. 8). The NAS is better suited for museums that have technologically astute staff while the Cloud is more user friendly. Typically Cloud providers offer a two-factor authentication to access the digital assets while only some NAS provide it (Betts, 2015, para. 19). Since NAS can be accessed over the internet security can still be an issue that an institution would need to address.

A second option is Storage Area Network (SAN). It offloads data storage from the desktops and server machines then reorganizes them into an independent, high-performance network. It is a network of interconnected storage devices, accessed through a local area

network (Lee, 2017, para. 14). When the institution wants to access a storage device it will send out a block-based access request for the storage device. It provides institutions with the option to easily add storage when they need it (Lee, 2017, para. 15). The main disadvantages to SAN is that the hardware is costly, and it is very complex to build and manage this storage option.

What we most commonly think of when it comes to storage is Direct Attached Storage (DAS). It is the storage that needs to be physically connected to the device. Some examples are hard drives, CD/DVD drives, flash drives, external drives, and so on (Lee, 2017, para. 18). With this type of storage there is less concern over security as it is stored on a physical device instead of over a network connection. However, with this type of storage the institution will need to keep up with the changing technology to ensure that all the data can be accessed and moved if needed (Lee, 2017, para. 20).

The last type of storage is called Redundant Array of Independent (or Inexpensive) Disks (RAID). RAID is a data storage virtualization technology that uses multiple physical disk drive components into one or more logical units for data redundancy and performance improvement. The data is distributed across the drives according to RAID level, which depend on the required level of redundancy and performance. Each level is called RAID followed by a number, for example, RAID 0 or RAID 1. The various levels provide a different balance among the key goals of digital preservation: reliability, availability, performance, and capacity (Patterson, Gibson, & Katz, n.d., p.6).

Museums want a place where they can store, organize, catalog, preserve, maintain, access, and distribute their digital assets. To start they should be developing plans for

preservation of digital assets throughout their lifecycle. As well as keeping any eye on what is going on around them with new standards, tools, and software. When a museum wants to store their digital assets, they must create administrative, descriptive, technical, structural, and preservation metadata by using the proper standards (Digital Curation Center, n.d., Curation Lifecycle Model). The museum will have to decide what data should be excluded from the cloud service, such as, data for which an outside party owns the copyright, administrative record copies of internal documents that the museum has legal responsibility, personal and confidential information regarding donors/staff/visitors, any data that is covered by a policy prohibiting storage outside of the museum's control, and so on. After that they will be able to evaluate and transfer to ensure their digital assets with attached metadata to long-term storage for preservation. Museums need to be able to make sure that the data remains authentic, reliable, and usable while maintaining its integrity. The metadata will help with those along with checksums, data cleaning, making sure file formats are accessible and more (Digital Curation Center, n.d., Curation Lifecycle Model). The ability to migrate data away from obsolete formats into a new format that will continue to work.

When it comes to storage, museums want to make sure that they have the capacity for their current digital assets and for future digital assets. With storage, museums also want to make sure that it is secure, since it can often contain sensitive information. When looking at cataloging needs for digital assets, museums want to ensure that the necessary fields are available so that they can properly document information related to the digital assets. It is important to document rights, restrictions, and security requirements for the digital assets so there is a clear understanding of access (Smithsonian Institution, 2010, p. 11). By looking at the

ISO and TRAC (Trusted Digital Repositories and Audit Checklist) the museums will be able to understand what they need from a repository that would hold their data and what they need to be doing to ensure long-term preservation.

## Research Method

By conducting interviews along with looking at case studies that institutions or service providers have made available, I was able to get an understanding of where cloud services stand with museums. Cloud service providers include some of their partners that they have worked with which provided some contacts along with asking a few museums that may not use the cloud. The first questions asked was if the institution uses cloud services in any way and depending on their answer and followed up with a how and why question. I asked the museums what type of set up do they have (hybrid, NAS, SAS, DAS, etc.). The last thing asked was if they thought a cloud service would be a practical option for the museum and their digital assets. In addition to the interview, there are some case studies of museums using cloud services, all of which gave great insight.

At the Wisconsin Veterans Museum, the Processing Archivist and the Oral Historian were able to answer the questions. The museum does not currently use cloud storage, but it has been something that they have considered. Currently the IT team is most comfortable using storage that they can setup on servers. They also are concerned with the security of the cloud. Since they are a public, state-entity, they do not want to take any risks with security and their digital assets. For their storage, they have two NAS servers and the collections are copied to tape by the IT team.

Since the Wisconsin Veterans Museum had to cut their Digital Archivist position the growth of the digital collections has slowed down a bit. This in addition to their concerns easily explain why they have not ventured into utilizing cloud services. The concern over security seems to be the most common throughout the literature and discussions with the museums. While security with cloud services have improved greatly the larger providers have more threats attempted on them due to their size. With being a state-entity, it may not be worth the risk for most of their digital collections.

At the Philadelphia Museum of Art, they are using Shared Shelf to provide their docents with access to images. The museum wanted a way to make their images and information more available to others. The docents use Shared Shelf to study and prepare for future exhibits within the museum. Docents are in the gallery spaces with tablets so that they can pull up Shared Shelf and show the visitors additional images and provide better context to the art. The museum developed a crosswalk from their TMS to Shared Shelf. Since Shared Shelf has an ingest tool that will take in data from an excel sheet the museum is able to load information relatively quickly into the storage. So far everything has been working out well with the cloud service.

The Philadelphia Museum of Art gives a good example of how a museum can use a cloud service for supplemental storage and not as a primary storage. They have their own storage for digital assets elsewhere and take specific information from that to upload into a cloud service. With Shared Shelf's helpful ingest tool the museum staff do not have to spend a lot of extra time adding the information and images on to the cloud. The museum does not have to worry as much about the security concerns since the information will be the duplicate

image along with descriptive information. The amount of storage that will be needed is not as large since it is limited information that will be on Shared Shelf. The system that the museum has works well at the moment and will give them the opportunity to continue sharing with larger audiences.

At the Minnesota Historical Society, I spoke with the Digital Archivist. The MNHS has digitized and born-digital content not only in their museum collection but also in their manuscript, government records, sound/visual, oral history, and newspaper collections as well. The institution has a large amount of unique data, such as inventories and indexes to help researchers access the collections, research guides, topic-specific websites, and more. It was within the past couple of years that they have decided to use cloud services in the form of contract with Preservica. MNHS uses Preservica to store high-value digital made files in their accessioned collections. The use cloud storage in conjunction with local servers and both nearline and offline backup that is managed locally. The MNHS use their cloud service for a specific part of their digital collection, knowing that storing all would not be worth the resources that would have to be put into it.

At the National Gallery of Art in Washington, I spoke with Peter Dueker, Head of Web and Imaging Services. The museum uses cloud services for a variety of applications including financial management, training and their new enterprise DAM. The reason for going with a cloud service, regarding the DAM, was that is provided a flexibility for storage, redundancy and geographic dispersion of digital assets, and a lower IT maintenance burden. They are using Amazon web services and Amazon storage (mostly S3) and their new DAMS is NetX. They have found a couple negatives to working with a cloud service. The first is performance since the

museum needs to have the infrastructure to support high bandwidth applications. It is becoming less of an issue for them, but Peter mentions that for smaller museums it could be a problem. The other negative is the cost of the services. With on premise storage you make the initial investment but with cloud services you must keep paying every month, year, amount of storage used. For the National Gallery of Art in Washington, the cloud was the right decision and seems to be beneficial to the museum.

The first case study is of Harvard University, with its thirty-nine different repositories from departments all over campus, using Shared Shelf. Initially they had their own system, but they could not keep up with it and eventually switched over to Shared Shelf. They knew ArtStor had a community vision for image collection management, which was just what they needed with the number of contributors. With Shared Shelf, Harvard could use VRA Core, controlled vocabularies, and Cataloging Cultural Objects guidelines to ensure their digital assets were up to standard (Luther, n.d., p. 3). Shared Shelf allows for faculty to share resources and templates with each other to improve digitization policies and standards within the institution. Harvard also found it beneficial that they can restrict content so that high resolution images cannot be used from the website to protect against any copyright problems. One last benefit they mentioned was the ability to select role-based permissions so that not everyone can edit, delete, or add digital assets to the system (Luther, n.d., p. 4). Shared Shelf seems to be working out well for Harvard and has given them the opportunity to customize and share with each other all their digital assets.

The other case study was of the Walters Art Museum in Baltimore, Maryland as they worked with ResourceSpace. The major problem the museum had was that they have every

type of object one could imagine within their collection. By using ResourceSpace the museum has been able to add 300,000 images to the system. They also got help by being able to add a TMS feature into the base code, which then allowed for easy importing. Walters Museum of Art was able to pull data from the TMS while uploading onto ResourceSpace (ResouceSpace, n.d., p 3). Every night the systems sync with each other to detect any changes. This also helps to ensure the integrity of the digital assets. Additional checks and balances is use by the pending submission and pending preview features. It prevents people from editing or adding things to ResourceSpace and so the changes need to be approved before made final (ResourceSpace, n.d., p. 5). The case study mentions a few times how ResourceSpace has made it so easy to share information and to access the digital assets from anywhere.

## Analysis

By looking at the interviews and case studies the cloud services are proving to be useful within the museum field. These examples show that there is no right way to use cloud services. It can be used as the primary storage of digital assets or it may be better to use it to provide public access to information, so it can be used as a learning tool, like the Philadelphia Museum of Art did. The museums analyzed all liked the ability to customize and have a system that works best for their digital assets, budget, and staff. The reoccurring concern was always security, which is why the Philadelphia Museum of Art only puts limited information in the cloud and why the Wisconsin Veterans Museum has not yet made the move to the cloud. Overall there is a great interest in what the cloud can do for the museum, it is more a matter of time and resources before it is utilized even more. Until then there are still other options that involve the cloud within institutions.

As previously mentioned, there is the option of hybrid cloud computing. This could mean combining the several types of cloud together, public, private, and/or community. This allows for more customization that will fit the needs of the museum and their digital assets. By combining several of the deployment options the museum can create a more secure place for sensitive data for more of a cost but then find a cheaper option for the information that the intend to share or do not mind if the public can access it.

Another hybrid option would be using the cloud in conjunction with an on-premises storage. There is the 3-2-1 method which has the institution creating three copies of the data. Keeping the data on at least two types of storage devices and store one of these offsite, or on the cloud (Hope, Thornhill, & Carr, 2017). Having various levels to storage can be beneficial and cost effective for museums. A museum could use a cloud service for storing images and descriptive information on artifacts since there would be no legal restrictions on the digital assets, so if someone were to hack into the cloud service there would be no legal ramifications. While storing the more sensitive data, such as donor information, emails, or internal documents may have to be produced in a legal matter, so it would be safer on a hard drive. Again, the museum will be able to tailor their digital management to fit their needs and have the advantages of using both cloud and on-premises services.

After looking at the literature, interviews with museums, and case studies, should museums start using cloud services? It is impossible to give a definite yes or no since it will come down to the needs of the individual museum. With the variety of options regarding cloud computing it would be difficult to find a museum that would not be able to use the cloud in some way. The way the cloud can help and what digital assets are stored within the cloud will

change for every museum. The cloud is adaptable and does not need to be the primary digital management service within a museum, as we can see there are many options out there.

Museum should start by thinking about the capabilities that they need to have versus the specific technology or product. Then establish the museums 'must haves' from the provider and understand what are the 'wants' (Beagrie, Charlesworth, & Miller, 2015, p. 12). By doing this along with identifying the amount of digital assets that the museum currently has and its projected growth, the museum is in a better place to make a decision on computing services (Hope, Thornhill, & Carr, 2017). When it comes to deciding on the cloud there are some additional questions that museums should ask themselves. Such as what do they want to put in the cloud, how do they want to use it (storage, additional backup, museum store), what information and records work the best in the cloud, are the necessary functions being met, does the museum have the staff and resources to move to the cloud, and lastly is security concerns (Kussmann, 2011, p. 6).

Cloud computing continues to grow in popularity with the public and business, so why not museums too? With the options of public, private, hybrid, community, service, platform, backup, and infrastructure there are many possibilities for the cloud to be beneficial to the museum field. The amount of cloud providers keeps growing and many of which are willing to work with the institution to make sure they have what they need. The cloud can be cheaper, more flexible, allow for sharing and collaboration, fast deployment, reduce IT maintenance and much more. Some ways around the risks that have been mentioned is by creating multiple copies, whether it be done by the museum or the cloud, and storing in multiple locations. It is also important that there is transparency in contracts about the services and protections

provided. The cloud providers should be clear with the museums on what they are providing and if they are not, that is an additional risk a museum will have to consider. Museums also need to be aware of the additional costs that may arise from increasing bandwidth, address intellectual property issues, copyright, potential for crashes, and security. By understanding the advantages, disadvantages, and their own collection of digital assets the museum can evaluate the cloud providers and make an educated decision on what would work best for them.

References

Arches. (2017). What is Arches? *ArchesProject*. Retrieved from https://www.archesproject.org/what-is-arches/

Artstor. (n.d.). Manage, catalog, and share. *Shared Shelf*. Retrieved from http://www.artstor.org/sharedshelf

Beagrie, N., Charlesworth, A., & Miller, P. (2015). How Cloud Storage can address the needs of public archives in the UK. *The National Archives Guidance on Cloud Storage and Digital Preservation*. Retrieved from http://www.nationalarchives.gov.uk/documents/archives/cloud-storage-guidance.pdf.

Berenz, E., Burns, A. R., & Stylianopoulos, L. W. (2013). Cataloging in the Cloud: Shared Shelf and ArchaeoCore. *VRA Bulletin*, 38 (2), Article 2. Retrieved from http://online.vraweb.org/vrab/vol39/iss2/2

Betts, A. (2015, April 2). NAS vs The Cloud: Which Remote Storage is Right for You. Retrieved from http://www.makeuseof.com/tag/nas-vs-the-cloud-which-remote-storage-is-right-for-you/.

Capterra. (n.d.). ResourceSpace. *Capterra.* Retrieved from https://www.capterra.com/p/126534/ResourceSpace/

Chen, Y., Paxson, V., & Katz, R. H. (2010). What's New About Cloud Computing Security?. *Electrical Engineering and Computer Sciences University of California at Berkeley.* Retrieved from http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-5.html

Clarke, K. (2013, October 1). Leveraging the Cloud for Museum Collections. Retrieved from https://www.slideshare.net/kacyclarke/mpma-2013-leveraging-the-cloud-for-museum-collections.

Cron, B. (2017). Metadata Requirements for Permanent Electronic Records in the Cloud. *The National Archives Records Express*. Retrieved from https://records-express.blogs.archives.gov/2017/11/06/metadata-requirements-for-permanent-electronic-records-in-the-cloud/

Digital Curation Center. (n.d.) Curation Lifecycle Model. *DCC*. Retrieved from http://www.dcc.ac.uk/resources/curation-lifecycle-model

Fastmetrics. (2017). What is Cloud Computing and How Does it Work?. *Fastmetrics Business Blog*. Retrieved from https://www.fastmetrics.com/blog/tech/what-is-cloud-computing/

Ferriero, D. (2010). NARA Bulletin 2010-05. *National Archives.* Retrieved from https://www.archives.gov/records-mgmt/bulletins/2010/2010-05.html

Goldner, M. (2010). Winds of Change: Libraries and Cloud Computing. *BFP*, 34, 270-275. DOI10.1515/bfup.2010.042.

Griffth, E. (2016, May 3). What Is Cloud Computing? Retrieved from https://www.pcmag.com/article2/0,2817,2372163,00.asp.

Hope, M., Thornhill, T., & Carr, J. (2016). Digital Preservation Storage Choices: On-premise, Cloud, Hybrid. [Presentation] Presented at *Practical Digital Preservation*. Retrieved from https://attendee.gotowebinar.com/recording/420238121347779843

Instrumental, Inc. (2013). Report on Digital Preservation and Cloud Services. *Minnesota Historical Society*. Retrieved from www.mnhs.org/preserve/.../Instrumental_MHSReportFinal_Public_v2.pdf

Kaussmann, C. (2011). Cloud Computing: An Introduction. *Minnesota Historical Society*. Retrieved from http://www.mnhs.org/ndiipp

Lee, J. (2017, June 1). Network Storage Explained: Cloud vs. NAS vs. SAN vs. DAS. Retrieved from http://www.makeuseof.com/tag/network-storage-explained-cloud-nas-san-das/

Ludden, J. (2014). An Introduction to Digital Strategies for Museums. *Museums and the Web*. Retrieved from http://mwa2014.museumsandtheweb.com/paper/an-introduction-to-digital-strategies-for-museums/

Luther, J. (n.d.). Digital Media Management = Shared Shelf. *Shared Shelf*. Retrieved from http://www.artstor.org/sharedshelf

Mattoo, I. A., (2017). Security Issues and Challenges in Cloud Computing: A Conceptual Analysis and Review. *International Journal of Advanced Research in Computer Science*, 8.2, 46-48. ISSN No. 0976-5697.

Mell, P. & Grance, T., (2011). The NIST Definition of Cloud Computing. *National Institute of Standards and Technology.* Retrieved from https://www.nist.gov/publications/nist-definition-cloud-computing

Moad, C., et al., Museums and Cloud Computing: Ready for Primetime, or Just Vaporware?. In J. Trant and D. Bearman (eds). *Museums and the Web 2009: Proceedings*. Toronto:

Archives & Museum Informatics. Published March 31, 2009. Consulted September 7, 2017. http://www.archimuse.com/mw2009/papers/moad/moad.html.

Patterson, D.A., Gibson, G., & Katz, R. H. (n.d). A Case for Redundant Arrays of Inexpensive Disks (RAID). *Computer Science Division University of California*. Retrieved from www2.eecs.berkeley.edu/Pubs/TechTpts/1987/CSD-87-391.pdf

Platz, B. (2017). Taking it to the Cloud. *Museum Store Association.* Retrieved from https://museumstoreassociation.org/2017/07/taking-it-to-the-cloud/

Radack, S. (2012). Cloud Computing: A Review of Features, Benefits, and Risks, and Recommendations for Secure, Efficient Implementation. *Information Technology Bulletin*. Retrieved from https://www.nist.gov/publications/cloud-computing-review-features-benefits-and-risks-and-recommendations-secure-efficient

ResourceSpace (n.d.). The Walter Museum. *Case Studies.* Retrieved from https://www.resourcespace.com/casestudies

Rouse, M. (2010). Backup as a Service. *TechTarget*. Retrieved from http://searchdatabackup.techtarget.com/definition/backup-as-a-service-BaaS

Simmon, E. (n.d.). Draft – Evaluation of Cloud Computing Services Based on NIST 800-145. *National Institute of Standards and Technology*. Retrieved from https://www.nist.gov/document/nist-sp-500-322-evaluation-cloud-computing-services-based-nist-800-145.

Smith, A. (2014). Collections and the Cloud. *Minnesota History Museums*. Retrieved from http://www.minnesotahistorymuseums.org/wp-content/uploads/2014/07/Collections-and-the-Cloud-Adam-Smith.pdf.

Smithsonian Institute. (2010). Digitization Strategic Plan. *Smithsonian Institute*. Retrieved from https://www.si.edu/content/pdf/about/2010_si_digitization_plan.pdf.

# Appendix A

Ellen Brooks – Oral Historian at the Wisconsin Veterans Museum
Brittany Strobel – Processing Archivist at the Wisconsin Veterans Museum
Evan Towle – Now with ArtStor, previously Librarian for Digital Collections and Services at the Philadelphia Museum of Art
Sarah Barsness – Digital Archivist at the Minnesota Historical Society
Peter Dueker – Head of Web and Imaging Services at the National Gallery of Art, Washington

Appendix B

Examples of General Cloud Providers

| Provider/Product | Choice of Locations | Speed of Access | Degree of Adoption | Costs | Security | Data Migration Out |
|---|---|---|---|---|---|---|
| Amazon Web Services (AWS) Glacier | EEA (Ireland and Germany) and Global | Typically within 3-5 hours | High | No initial costs. Billed for usage by the hour | Comprehensive accreditations | Download standard formats by API, and move large data volumes on disk |
| Amazon Web Services (AWS) S3 | EEA (Ireland and Germany) and Global | Immediate, by widely adopted API | Very High | No initial costs. Billed for usage by the hour | Comprehensive accreditations | Download standard formats by API, and move large data volumes on disk |
| CloudSigma | Switzerland (EEA equivalent) and USA | Immediate, by API | High | No initial costs. Billed for usage in 5-minute increments | Suitable accreditations, some through hosting provider Interaxion | Download standard formats by API |

| | | | | | | |
|---|---|---|---|---|---|---|
| GreenQloud | EEA (Iceland) and USA | Immediate, by AWS – compatible API | High | No initial costs. Billed for usage by the hour | Suitable accreditations, some through hosting partner Verne Global | Download standard formats by API |
| Microsoft Windows Azure | EEA and Global | Immediate, by API | High | No initial costs. Billed for usage by the minute | Comprehensive accreditations | Download standard formats by API. US option to move large data volumes on disk not yet available in Europe |
| Rackspace | UK and Global | Immediate, by Open-Stack API | High | No initial costs. Billed for usage by the hour | Comprehensive accreditations | Download standard formats by API |

Source: Berenz, E., Burns, A. R., & Stylianopoulos, L. W. (2013). Cataloging in the Cloud: Shared Shelf and ArchaeoCore. *VRA Bulletin*, 38 (2), Article 2. Retrieved from http://online.vraweb.org/vrab/vol39/iss2/2

Appendix C

| Provider / Service | Data Integrity | Reliability | Scalability | Retention and Portability | Availability | Data Ownership | Preservation Functionality | Total 3-Yr. Costs |
|---|---|---|---|---|---|---|---|---|
| Amazon S3 | Limited checksums | Average | Almost unlimited | Not easy to move | Average | Similar to others | None | Medium |
| Amazon Glacier | Limited checksums | Multiple tape copies | Almost unlimited | Not easy to move | Lower since on tape | Similar to others | None | Low |
| Google Cloud Storage | No checksums | Average | Almost unlimited | Not easy to move | None in contract | Contract concerns | None | Medium |
| Tessella's Preservica | Checksums and CRC | Multiple cloud copies | Same as S3 | Multiple providers | Multiple cloud copies | Similar to others | Developed for this | ? |
| VISI ReliaCloud | Limited checksums | Average | Cannot support MHS | Somewhat limited | Average | Similar to others | Some claimed | High |
| SDSC Cloud Storage | Automatic verification | Average but no tapes | Almost unlimited | Claimed to be easy | Concerns about disks | Similar to others | None | Medium |
| DuraSpace DuraCloud | User run checksums | Multiple cloud copies | Same as S3 | Multiple providers | Same as S3 | Similar to others | Some claimed | Medium |
| IBM SmartCloud | Limited checksums | Average | Unknown | Unknown | None in contract | Similar to others | None | ? |
| FujiFilm Permivault | Custom plans | On-site and cloud | Almost unlimited | Somewhat limited | On-site copy | Similar to others | Limited | Low |
| FujiFilm Permivault Client | Custom plans | Cloud only | Almost unlimited | Somewhat limited | Average | Similar to others | Limited | Low |
| Code 42 CrashPlan Pro | Limited checksums | Average | More limited than S3 | Somewhat limited | Average | Similar to others | None | Low |

| | |
|---|---|
| | Very good to best capability |
| | Average capability |
| | Poor or no capability |
| | Unknown data |

Source:

Instrumental, Inc. (2013). Report on Digital Preservation and Cloud Services. *Minnesota Historical Society*. Retrieved from www.mnhs.org/preserve/.../Instrumental_MHSReportFinal_Public_v2.pdf