

# **Keeping it Spinning: A Background Check of Virtual Storage Providers**

Anne Roach, AG<sup>®</sup>, CG<sup>SM</sup>

## **INTRODUCTION**

Whether developing software using virtual storage, backing up genealogical or historic data or research, or storing virtual communication for historic purposes, the world of genealogy relies heavily on the reliability of virtual storage services and will continue to do so with virtual genealogy software. The need to understand the details of virtual storage is more critical than ever before. Virtual storage providers are responsible for informing end users about their physical storage facilities to help users make educated decisions about where data or data backups are stored. The ultimate responsibility for making an educated virtual storage decision, however, rests on the user, business owner, or developer, who makes the decision to trust critical information to the virtual cloud.

In a world where more and more data is sent to the virtual cloud city for storage, it is critical to understand exactly how secure and accessible virtual storage really is. As the field of genealogy becomes more technologically advanced, the need to preserve historic records and personal genealogy files increases. Whether viewing digitized databases online, creating information in an online family history database, or backing up personal files from a home computer using an online service, virtual storage impacts everyone involved in the field of genealogy.

Trusting critical data, genealogical or otherwise, to a virtual server without understanding details about the virtual storage facility may lead to unrecoverable data loss. An uneducated decision about virtual storage is not unlike purchasing a vehicle without examining anything but the exterior paint, or purchasing it because of the anti-theft system without considering the vehicle's performance, reliability, or safety. While virtual storage may be obtained anywhere and anytime by online users, the physical limitations of data still apply. Many important factors are relevant to selecting a virtual storage provider, however this paper will focus on the information that is not readily available to consumers or developers when choosing a virtual storage provider, including available redundancies, storage system architecture, and physical location vulnerabilities.

## **DETERMINING THE RESEARCH FOCUS**

The concept of virtual storage leaves the developer and the consumer without much knowledge about the security of their data beyond the standard items described online. This usually entails elements of virtual protection, including password security elements, 128-bit SSL (Secure Sockets Layer) encryption, or with luck (or a slightly higher investment) one may employ 256-bit SSL or AES (Advanced Encryption Standard) encryption to protect data being transferred online from being viewed or caught in the

criminal fishing net. Ask a virtual storage user where the data is actually stored and he or she may answer that the information is “online” at worst, or slightly better, may give the site through which the user uploaded his or her precious data. Even an educated individual may only be able to give a physical location for the servers that keep the virtual data online. Ask the virtual storage user about the data’s security, and the likely answer will include the online security features mentioned above with little or no regard to physical limitations.

First, the term ‘Virtual Storage Provider,’ for the purpose of this paper refers to those entities that are involved in the physical creation of virtual storage containers, and make those containers available for use by consumers, businesses, enterprises, and developers. The difficulty rested in locating entities that fit this description. Locating true virtual storage providers may seem simple at first; however, there are many factors that determine whether a company is a true virtual storage provider, or simply company renting space from a larger virtual storage entity. This process involves distinguishing between several complex layers of virtual storage that cannot be completely covered in this paper. Simply stated, there are several layers of virtual storage between the actual server and the end user. While there are many more complexities to virtual storage ownership and management, a simple list is included below:

1. The owner of the ground upon which the data center resides
2. The owner of the physical building in which the data center resides. (Some data centers lease portions of larger buildings that house other businesses.)
3. The manager of the facility in which the servers reside
4. The owner of the racks, servers, and other equipment used to store the end user’s information
5. The company that maintains the equipment upon which the end user’s data resides. (This may or may not be the same as the public company that provides the space.)
6. The owner of the company that interfaces with the end user and offers the end user storage space free or for a fee

This paper primarily focuses on items 4, 5, and 6 in this list: those directly responsible for the end user’s data.

Second, the purpose for which the storage is being used must be considered. For example, several e-mail providers are, in essence, virtual storage providers with a limited range of use. Some of these companies, including ‘gmail,’ have extended the use of their virtual storage to include additional services for consumers, including Google Docs and Page Creator. Because of the limited extent of these services, e-mail providers and similar limited-use storage providers were not included in this study.

As mentioned above, several virtual storage entities are “stores” that purchase space from another company, or lease servers in a storage server farm, often known as “Secure Data Centers.” These facilities are often separate from the physical location of the virtual storage entity, and the servers are usually maintained and serviced by employees of the data center, not by employees of the virtual storage entity. This created an obstacle for locating the needed information from individual companies. Often those individuals with

information about the precise data center were unknown to employees who interface with the public, therefore, the request for information turned up negative research result.

Additionally, new virtual storage entities appear online on a regular basis, so while this paper includes several major virtual storage providers, it is not all-inclusive and does not claim to be. Despite attempts to obtain the desired information from several virtual storage providers, some requests for information were refused, unavailable, or in some cases, requests for information received no response. In several cases, as mentioned above, an individual with the critical information simply did not exist according to the accessible company informants. Despite these obstacles, much valuable information was located to help developers and consumers make educated decisions about virtual storage beyond the normal information included in Web site advertisements and security feature disclaimers.

Major elements of physical vulnerabilities were identified to determine the best possible standards for virtual storage providers. While various standards exist for data centers, not all data centers apparently follow the same standards. Cooling for data centers was not included in this study. The ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) guidelines are more likely to be followed due to building codes regulated by local government entities than are practical physical security elements. While data centers have engineering specifications and standards, no standard currently exists that requires a virtual storage provider to follow every possible precaution to make certain that even in the case of a regional natural disaster, information is not lost. SAS-70 (Statement on Auditing Standards) does check the viability of a data center to ensure that proper financial, auditing, and operations practices are followed, but is not required as an industry standard, and does not ensure a long lifespan of data.<sup>1</sup> No regulations exist requiring a virtual storage provider to make sure that clients have a copy of their stored data before that company goes out of business or before the lease on their servers ends.

## **RESEARCH MODEL**

The questions asked each virtual storage provider are included below, however, not every question had an available answer due to knowledge limitations between each provider. Many providers claimed that they had never been asked several of these questions:

---

<sup>1</sup> SAS-70 reporting, [www.sas70.com](http://www.sas70.com): “In a Type I report, the service auditor will express an opinion on (1) whether the service organization's description of its controls presents fairly, in all material respects, the relevant aspects of the service organization's controls that had been placed in operation as of a specific date, and (2) whether the controls were suitably designed to achieve specified control objectives. In a Type II report, the service auditor will express an opinion on the same items noted above in a Type I report, and (3) whether the controls that were tested were operating with sufficient effectiveness to provide reasonable, but not absolute, assurance that the control objectives were achieved during the period specified.”

1. Do you own or lease your servers?
2. Are your servers redundant?
3. Where are your servers/server clusters located?
4. Which vendors do you use for your servers?
5. Do you use a tape library to back up your servers?
6. What is your total storage capacity?
7. How are your servers monitored?
8. How quickly would a damaged server be replaced?

### **1. Do you own or lease your servers?**

The first question determined if the virtual storage provider was using a data center or not. Unfortunately, some providers stated that they owned the servers even if they were leasing them. After clarification, one provider explained that the servers were under a corporate lease and that when the lease ended, the company would own the servers. This question is also critical to determine the security of the data. If servers are leased and a virtual storage provider is unable to pay for the lease, or the lease cannot be renewed, the data contained on the machines must be migrated to a different location or machine en masse. This equals down-time for the virtual storage user or a possible loss of data.

### **2. Are your servers redundant?**

This question was intended to determine if the virtual storage provider was duplicating the information. In some cases, the redundancy only existed within the same facility, and some providers felt that using RAID in a single location compensated for the need to create redundancy. At least two providers stated that their system provided for two or more copies of the data within their servers.

### **3. Where are your servers/server clusters located?**

Location redundancy is another critical element to virtual storage. If all server clusters are located in the same city in California, for example, and a natural disaster occurs, the information may become unavailable for a long period of time or may be lost entirely. Despite the odds of a major disaster taking place in one area, it is helpful to know if a virtual storage provider has considered location redundancy as part of their security plan. Some providers have multiple locations of servers, but do not expend the money necessary for location redundancy. It takes more bandwidth to transfer data over a larger distance and bandwidth creates more overhead for the company.

### **4. Which vendors do you use for your servers?**

Many providers used only one vendor for all of their server clusters. While the primary provider of these servers is a noted and reliable company, using only one vendor for server clusters creates a problem only if the vendor fails to exist as a business entity or has widespread production problems. While this element of physical security and vulnerability is often overlooked, it may determine whether or not the virtual storage user maintains access to the virtual data. Only one of the interviewed virtual storage providers claimed vendor redundancy.

### **5. Do you use a tape library to back up your servers?**

The philosophy surrounding the use of tape libraries varies. Many companies feel that as long as the information is spinning, and has server redundancy, no other copy of the data is required. When working with data on an enterprise, or even a consumer level, however, legal issues may arise that requires a company or individual to use digital data as evidence. At present, the only digital data admissible in a court of law is digital linear tape because of the linear mode of recording data. Despite the expense, tape libraries also provide a storage format redundancy that allows data to be recovered even in a major data disaster when the tapes are stored in a separate location from the servers.

#### **6. What is your total storage capacity?**

Despite the physical limitations of storage space, many companies refused to admit that their virtual storage had a limit. Their logic was understandable to an extent, because in their own minds, the virtual storage was unlimited. Despite efforts to define a physical limitation on their current storage capacity to date, many providers resisted this question emphatically, basing their statement on the fact that at any moment, additional servers could be located, purchased, and added. Physical storage limitations are a security factor, however, determined upon the continued availability of new servers.

#### **7. How are your servers monitored?**

Some businesses require 24/7 assistance, including health care businesses, for example. With virtual data storage, there are no closing hours or opening hours, and this places virtual storage providers in the same category of 24/7 monitoring and customer service. Part of the problem may be resolved by using monitoring software. Monitoring software may locate a problem with a server, but, if the software fails, this failure may be the crack that breaks the dyke and begins the flood of lost information. While not all monitoring system information was available from virtual storage providers, at least one provider has a redundant-redundancy system to consistently keep information online with a backup to the provider's backup software system. Human monitoring or availability is also a critical element to the health of a virtual storage system.

#### **8. How quickly would a damaged server be replaced?**

Depending on the configuration of a storage system, only one or two server failures will be tolerated before data is lost. Immediate equipment replacement will ensure the life of virtual storage when combined with the other elements discussed above. While virtual storage seems the invincible method for keeping data alive, the physical limitations must also be considered. While this research did not discuss all of the aspects of physical vulnerabilities, including smoke or fire damage, cooling systems, power redundancy, and physical security, these critical questions will help virtual storage users locate a provider with reliable service. Several providers claimed that they had never been asked these questions, even those companies that market virtual servers to developers. Along with other virtual storage security elements, providers may begin to make the physical security information available to users and create a more secure environment for virtual storage in years to come.

## **RESEARCH RESULTS**

**iBackup: [www.ibackup.com](http://www.ibackup.com)**

Pro-Software's virtual storage service, better known as iBackup, serves a clientele of at least 550,000 strong. With several owned server clusters located in California, as well as New Jersey and London, the company not only has location specific redundancy plans, including power redundancy with onsite generators, but location redundancy as well. The company claims a minimum capacity of thousands of terabytes, depending on the data center in question. No tape libraries are used to back up the servers, however, redundancy plans are in place to keep virtual information online. The primary vendor for iBackup is Dell. Little other information was available for iBackup.

**Amazon S3: [aws.amazon.com/s3](http://aws.amazon.com/s3)**

Amazon S3 is one of the major providers for virtual storage, both to small business owners and to developers. Serving as a major backbone to the virtual storage industry, the need for details about physical security elements of virtual storage is especially critical for this company. Amazon S3 serves as the virtual storage provider to popular storage services like JungleDisk, ElephantDrive and Filicio. Amazon Web Services Business Developer and representative Moussa Dao explained that the information stored on Amazon servers is replicated 2-3 times for security and redundancy. Servers are carefully maintained, and monitored 24/7 to provide 99.9% end-user access. The data centers are equipped with power redundancy, and are located on the east and west coast of the United States, and in Europe. Despite the location redundancy, the information is stored closest to the user to minimize bandwidth use. This means that if one data center fails, the information is immediately transferred to a second data center in very close proximity to the first center to minimize bandwidth use. The information is not transferred or mirrored to a data center in a vastly different location.

Amazon provides unlimited storage at a reasonable price, allowing businesses to not only use virtual storage, but allows companies and developers to use virtual servers as well. Dao did, however, mention at least three physical buildings in which the servers are stored, and explained that these have very high levels of security by limiting access to the actual machines. Surprisingly, Amazon's representative was unable to disclose much of the necessary information. When asked if he had been asked these questions before, the representative claimed that he had not. When asked if there was anyone else who could answer these questions, the representative claimed that he was unaware of anyone else who could answer the questions that he could not, or would not answer.

**Nirvanix: [www.nirvanix.com](http://www.nirvanix.com)**

A competitor to Amazon, Nirvanix offers virtual storage for consumers, small business owners, and developers. While Amazon's virtual storage system infrastructure was designed as an outcropping of its massive e-commerce network, Nirvanix was developed with the specific intent of providing a "massively-scalable media storage and delivery network."<sup>2</sup>

---

<sup>2</sup> <http://www.nirvanix.com/company.aspx?=&topnav> Accessed 13 January 2008.

Nirvanix included much of the requested information in security white papers. With location redundancy on the North American east and west coasts, Europe, and Asia, Nirvanix also maintains the virtual storage content closest to the storage of the nodes. The servers are entirely owned by the company, using redundant vendors Cisco and Dell. The software used to maintain the system is agnostic to the hardware underneath, allowing for flexibility with additional vendors in the future as needed.

The company has planned power redundancy with dual power feeding into each individual rack in SAS-70 compliancy. The servers are maintained and staffed to keep all of the servers up and hot, calculating into milliseconds of downtime to the perspective end-user. The company's current capacity for storage is over 1 petabyte, with the ability to scale to exabytes within the current architecture of the company. One virtual storage vendor, Freedrive, a recently acquired client of Nirvanix, has expanded from zero to over 200,000 customers in less than a year. Representative Geoff Tudor stated that the system was scaling very well to handle that much growth. Despite the company's youth in the major virtual storage industry, Nirvanix provided a convincingly strong argument for security, including additional information to provide users with the best educated decision about the company. The architecture and physical security standards were not only readily available, but apparently well thought out. While the company does not use tape libraries to back up the information, it does include dual paths coming into a rack and a backup system that kicks in if the current monitoring system fails. In other words, as Tudor put it: "Our redundancy system even has redundancy!" In terms of cost, Nirvanix provides unlimited storage for about the same cost as S3.

**FilesAnywhere: [filesanywhere.com](http://filesanywhere.com)**

FilesAnywhere was not entirely willing to respond to all of the questions asked, as described by representative Jonathan Grubb, "Some of the confidential [information] is proprietary and thus cannot be disclosed." He did say that the company leased space for their data centers, but owned the servers located in the data center. Regarding location redundancy, Grubb responded, "No, we do provide synchronization software to all users so that they have a copy on their end as well." In other words, the company relies on the end user as the location redundancy system. Regarding maintenance, the company guaranteed 99.9% uptime.

**DriveHQ: [www.drivehq.com](http://www.drivehq.com)**

Drive HQ is a veteran in the field of virtual storage, and has been around for 5 years, since 2003, prior to the appearance of Mozy and Amazon S3. The company is a fast-growing company whose primary clientele consists of businesses, developers, and high-end users. Drive HQ owns all of their servers, and is co-located with Hurricane Electric in Fremont, California. To reduce cost, the company does not have location redundancy. The company hosts several businesses, and claims unlimited storage space. The responding representative, John Zhang began his response to the request by stating that the information was "confidential which I cannot release to you unless we sign some kind of business agreement," but was reasonably thorough in his answers.

When asked about tape library use, representative Zhang stated, “We have more advanced backup than regular tape libraries.” The company does have vendor redundancy, and monitors the system with automatic and manual monitoring. According to Zhang, “Major problems can be altered in real-time. We also have server health monitoring software, plus a network monitoring team.” The company uses Raid 50 storage, and has a hot mirror system to avoid loss of data availability. Their track record for reliability is over 99.9% according to the representative.

**MyOtherDrive: [www.myotherdrive.com](http://www.myotherdrive.com)**

MyOtherDrive representative John DeRegnaucourt explained that he preferred not to answer some of the questions, stating, “You have a lot of detailed questions below, some of which [is] not information we provide to the general public.” He stated that the company owns their drives, located in a data center that he preferred not to name. The servers are located in a single data center with no location redundancy. When asked about a backup system, he asked to skip that question. The company’s maintenance system informs the monitoring party via e-mail, although DeRegnaucourt did not include other monitoring details. The company has no current power redundancy such as a generator, but is looking at moving the servers to a new location that does include power redundancy. He claimed unlimited storage capacity currently. The most surprising part of this interview involved the vendors for his servers: he claimed they make their own servers. Router and switch vendor is Cisco, but the company representative preferred not to respond to all other questions.

**Mozy: [www.mozy.com](http://www.mozy.com)<sup>3</sup>**

Josh Coates, founder of Mozy, explained that Mozy’s primary purpose is to back up data and provide a secondary location at a remote location. The company hopes to offer cloud storage in the future, similar to Amazon S3, and mentioned that the company will run this architecture differently as a primary data storage service. Coates anticipates this change by the end of the year. At present, Mozy owns all of their servers and that data is redundant on their servers. Three of Mozy’s data centers are located in Utah, with plans to open a fourth center in Ireland this summer, however, the data is not redundant between locations.

Tape library backup is not included in Mozy’s current structure, however, the storage structure includes various vendors for a true COTS (commodity off the shelf) system to prevent the system from being limited by proprietary equipment. Mozy maintains and monitors all of their own equipment. The servers are monitored using an automated monitoring system as well as operations staff members. Non-critically damaged equipment is replaced on a weekly or bi-weekly basis, and, according to Coates, “we use an algorithm that allows an enormously high MTTR [mean-time-to-restore] threshold.” Given the size of the storage system, approximately 6 petabytes to date, a high MTTR can only mean greater security for the end user.

---

<sup>3</sup> Note: Mozy’s information was contributed after the March Technology Workshop, which is why their information is not included in the original slides.

**AllMyData: [www.allmydata.com](http://www.allmydata.com)**

AllMyData offers unlimited virtual storage and access, but for a fee beyond the initial 1 GB account. Representative Peter Secor also seemed surprised by the questions and asked if the purpose was for commercial deployment. He explained that the company had developed their own storage technology using an open source platform. They presently own their servers, but those servers are located and managed by United Layer collocated in Data Centers owned by Digital Realty Trust, owner of about 70 percent of all data center facilities in the United States. AllMyData's location redundancies are limited to the Bay Area in California, but the company is actively planning new locations on the United States East Coast, and in Europe and Asia.

Their primary client is the small business or consumer, 10,000 in number, with a storage capacity of 100 TB. The company does include vendor redundancy, but does not use tape libraries to back up information. According to Peter Secor, their well designed system would allow for over two-thirds of the servers to go down with the ability to recover 100 percent of the stored data due to internal redundancy and external redundancies described above. "Because we expect them to fail, we maintain full availability and can replace them at our leisure without affecting customer availability of data," said Secor. The company uses off-the-shelf equipment to allow for quick and easy expansion of their current storage capacity.

**United Layer: [www.unitedlayer.com](http://www.unitedlayer.com)**

United Layer was the only data center management company mentioned by name throughout the research among user-interface companies, and because of this fact, this company was included to provide additional information about AllMyData. Contact Richard Donaldson was more than willing to share any information he could, not only about the services provided by his company, but any question about data centers in general. He willingly sent as much useful information as possible about data centers, including the company's 'Technocratic Oath,' in which UnitedLayer employees agree to remember that they do not practice their "craft on just routers, switches, or servers," but on human beings, and that they will "gladly share such knowledge as is [theirs] with those who are to follow."

The company offers a variety of services from providing cabinets in which a virtual storage provider, small business owner, developer, or enterprise may install their own servers, to providing ready-to-use leased servers to the same. The company owns the servers they lease, and included more detailed infrastructure information about each data center than any of the user-interface virtual storage providers, including exact vendors used for switches (Foundry), routers (Cisco), servers (Dell), cooling systems, humidity control, fire suppression systems, and other physical security elements. Their data centers are currently located in San Francisco and Los Angeles, and have redundancy in Chicago and Ashburn, Virginia, using Digital Realty Trust facilities. The company does not use tape libraries, only disks, and has real time monitoring so they "know when any component begins to panic prior to failure." Donaldson also verified that no consistent entity or provisions are in place to regulate the security elements of virtual storage providers.

### **Resistance to Providing Information**

Companies that were less than eager to share physical security features of their systems included Carbonite, located in Boston, who stated that they did not give out the requested information. DataDepositBox explained that the company owned the servers, and later explained that they were under a corporate lease, but had redundant power, and onsite redundancy within the single location and single physical building. Box.net representative replied, “We’ll respectfully pass on this opportunity.”

Humyo corresponded with several e-mails before sending links to their security white pages and information online. The only physical security element included in this information was the climate-controlled former Bank of England gold vault in which the servers are located. The company did not respond when they were informed that the information contained in the white papers and online were not adequate to answer the requested questions. Other companies failed to respond to the request for information, or a representative with the necessary information could not be located such as with the AOL-based virtual storage service, XDrive. XDrive only listed a physical address for the company, and despite attempts to locate a representative for XDrive by contacting AOL, no known representative was identified.

Non-responsive companies included Live SkyDrive, Winkflash, MediaMax, FileGenie, Writeboard, StrongSpace, and Omnidrive. Diino answered the initial phone call, stated that the servers were owned, but after the other questions were presented, asked to end the call and requested the researcher to call back in ten minutes. When the second and third calls were made, no one answered the calls. No response was received when an e-mail was later sent.

Based on Internet searches and online blogs, some misunderstanding about virtual storage companies also existed. MediaMax, Google, and LiveSkyDrive sent generated or live responses, showing that the request for information was received, and that additional correspondence would follow, but no other response arrived. Some companies appeared to offer virtual storage, but instead only offered backup software including Openomy and Remote-Backup. Lighthouse, on the other hand, was a server architect that offered assistance and consultation services to building virtual storage solutions.

Other companies had merged or evolved into larger providers including Streamload (now MediaMax) Fabrik/SimpleTech Backup (powered by Mozy), and Writely (now Google Docs). A last-minute find, AT&T Online Vault was eager to participate in the study and asked for more time to answer the request, but did not respond within the three days the company representative requested, and has not responded since.

<b>Provider</b>	<b>Site</b>	<b>Requested Info?</b>	<b>Responded?</b>
iBackup	www.ibackup.com	Y	Y
Live SkyDrive	Skydrive.live.com	Y	N
Google Docs	Docs.google.com	Y	N

Lighthouse	N/A	Architect only	N/A
Winkflash	Winkflash.com	Y	N
MediaMax	www.mediamax.com	Y	N
Amazon S3	Aws.amazon.com	Y	Y
Filegenie	www.filegenie.com	Y	N
Writeboard	www.writeboard.com	Y	N
Writely	www.writely.com	(See Google Docs)	N/A
Allmydata	www.allmydata.com	Y	Y
Mozy <sup>4</sup>	www.mozy.com	Y	Y
strongspace	www.strongspace.com	Y	N
Streamload	www.mediamax.com	(See Media Max)	N/A
Xdrive	www.xdrive.com	Contact Info Physical address only	-
Openomy	www.openomy.com	Software Company	N/A
Omnidrive	www.omnidrive.com	Y	N
Box.net	www.box.net	Y	Refused
Carbonite	www.carbonite.com	Y	Partial/Refused
Humyo	www.humyo.com	Y	Partial
Datadepositbox	www.datadepositbox.com	Y	Refused
Remote-backup	Remote-backup.com	Yes - Software Company	N/A
Diino	www.diino.com	Y	N
Myotherdrive.com	www.myotherdrive.com	Y	Y
Backuplite	N/A	Couldn't locate	N/A
Nvbackup	N/A	N/A- Palm Backup	N/A
Innerbackup	N/A	N/A- Palm Backup	N/A
driveHQ	www.drivehq.com	Y	Y
Nirvanix	www.nirvanix.com	Y	Y
Fabrik/SimpleTech Backup	www.fabrik.com	Mozy Powered	N/A
FilesAnywhere	www.filesanywhere.com	Y	Y

Figure 1: Table of surveyed virtual storage providers

## NEW FAMILYSEARCH:

New FamilySearch offers an online pedigree where genealogists may work collaboratively together by uploading and adding information to a specialized data system that only allows for storage of genealogy-specific data. Because of this fact, New

<sup>4</sup> Note: Mozy's information was contributed after the March Technology Workshop, which is why their information is not included in the original slides.

FamilySearch does not fall into the category of a virtual storage provider; however, because of their critical role in the field of Genealogy, specifically for Latter-day Saints, the question of reliability of this virtual genealogical database was included in this paper.

The responsibility of New FamilySearch is first and foremost to provide names for The Church of Jesus Christ of Latter-day Saints' 125 temples. The Family and Church History Department under the Corporation of the President stores the records all of the church ordinances, and takes this responsibility very seriously. Safety and accessibility overrides a need to reduce overhead by limiting bandwidth use, so data redundancy includes duplicate copies of data in more than one location, over 2500 miles apart.

Currently the common pedigree located on new.familysearch.org has two simultaneously syncing copies of data. This will be increased to three simultaneously syncing copies before the Wasatch Front Temple districts are added to the system. The New FamilySearch online pedigree has been established to provide consistent access to the data with only 2-3 minute down-time per instance and should calculate to no more than 6 minutes of down time per year. In other words, the data is available 99.999% of the time.

New FamilySearch representatives Dave Beus and John Sjogren were eager to share information about the data security elements of the New FamilySearch online pedigree. The primary system, located in Virginia, consists of Sun Microsystems M9000 servers that work much more like a mainframe than a cluster to provide the best data reliability and accessibility. Each M9000 is independently connected to a Hewlett Packard XP 12000 Storage Array. These independent storage environments sync data back and forth with each other at a rate of 2000 MB per second. Data transferred between each unit exists in an independent failure domain, translating into no corrupt data transfer between these units. This configuration allows for any piece of hardware and replaced to be removed at any time, much like a cluster configuration.

The XP 12000 Storage Array is backed up in duplicate: by an offline tape library system, the Sun Storage Tek SL8500 consisting of 10000 tapes, 8 robots that move tapes from storage slot to drive, and back again at a rate of approximately 60 mph, as well as a virtual tape library that allows for a complete restore of lost or damaged data within 3-4 hours.

Just minutes behind the Virginia system, a secondary storage location in Utah contains a mirrored copy of the Virginia data using Sun Microsystem's M8000. The M8000 is also backed up in duplicate by a virtual tape library, and by an SL8500 tape library in Utah, with a second tape library located in a separate location in Utah using the SL500. With several redundancies already present, the Family and Church History Department also has a separate system that is a duplicate of the Virginia storage configuration that is located in Utah and may be used as the primary system at any time if the Virginia system is rendered unusable.

At present, the system mentioned above has 2.2 PB of data spinning, with over 10 PB of tape backup. The system is monitored by a group of on call system support engineers

7/24, 365 days per year with a response time of five minutes or less. Not only is the server configuration redundant, but the human support system is also redundant with multiple system support engineers on call at all times. Additionally, the storage hardware in the configuration ‘calls home’ to the vendor twice a day. If no call is received, a vendor technician is dispatched to service the system, located no more than fifteen minutes away.

Availability to data has also been carefully considered with the New FamilySearch online pedigree. All data is available online through redundant networks, which are also connected to redundant power sources. New FamilySearch has gone to great lengths to protect the data contributed to the online pedigree and to ensure that the data stays online and available to the end user.

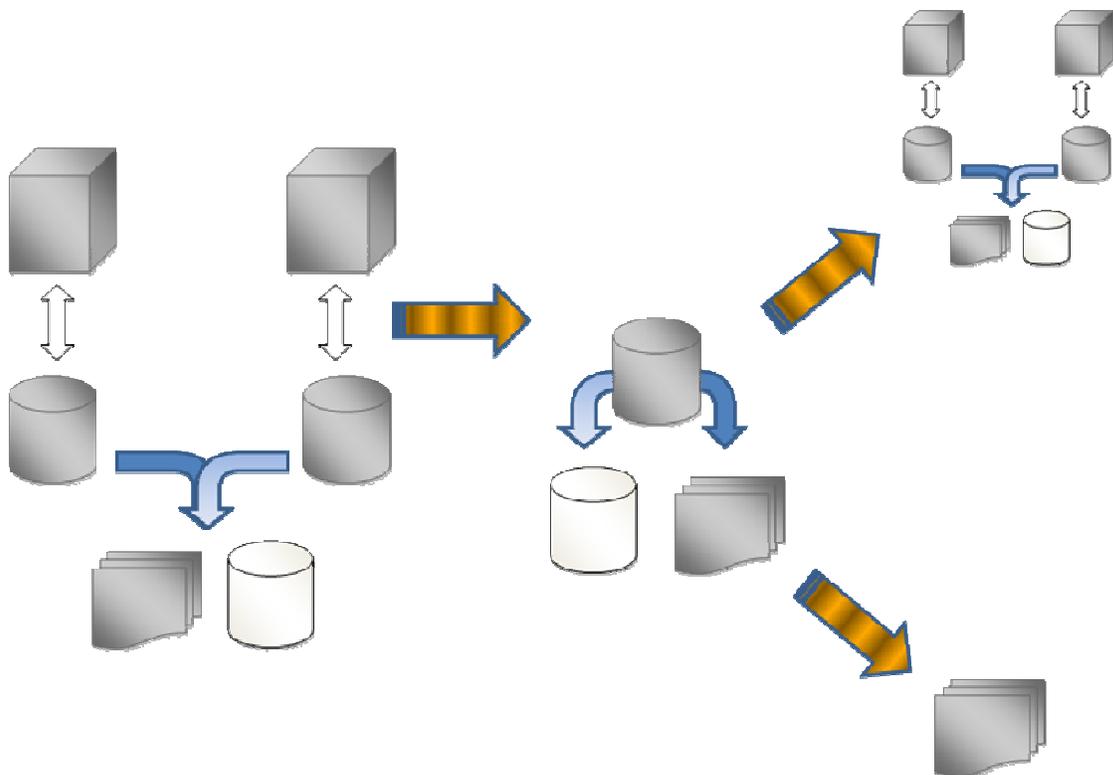


Figure 2: Visual model of storage redundancy for the New FamilySearch online pedigree

### CONCLUSION:

While new storage providers pop up every day in the Virtual Storage community, there are no immediate designations about what role each company plays without additional research. Some companies claim that they manage and maintain their own servers when they are actually far from the physical location of the user-interface company. This disconnect can create a lack of responsibility and a vulnerability to the end-user. With little control over the physical storage elements of the equipment, and with no solid regulations in place to determine the highest standards of virtual storage, several

thousands of cubic feet of data could be easily lost by failing companies that do not own or manage their own equipment, or by a natural disaster where no location redundancy exists. Power redundancy, data redundancy, and physical security elements may not be clearly communicated to the end-user, creating potentially poor decisions by the end user.

The chances of losing data in a personal backup and in a virtual storage backup simultaneously is very slim, however, some companies store data close to the user to avoid paying for the extra bandwidth it takes to transfer the data a safe distance away from the end user's location. For example, if the end user lives in San Francisco, and the end user's data is also stored in San Francisco, the data may be lost in natural disaster or become inaccessible in a major power outage without power redundancy.

Virtual Storage companies have the responsibility to inform the end-user about assumptions about user copies, such as FilesAnywhere, a virtual storage provider with no location redundancy. Some companies use the guarantees of the data center managers to protect their own clients by stating that if data is lost, money will be refunded to the end user, or the provider may include a separate guarantee, however, these guarantees may be hollow when the company has little involvement in the protection and maintenance of the information without informing the client about the actual data guardian. On the other hand, when virtual storage providers invest proper care toward a user's data, this may bolster trust in the end user, and help alleviate virtual storage disaster concerns.

Taking the time to ensure genealogical data is secure and accessible in any form, including blogs, e-mail, genealogical data, and other forms of communication, will create a well defined path for future researchers to follow, if they are able to see the efforts of past researchers and family members. Taking responsibility for virtual storage before data is lost or unavailable will not only help future generations avoid duplication, but will allow them to continue locating the sources necessary to extend research lines without wasting time searching the same records that have been searched before, and help the work of genealogy move forward.