COMPUTER CORNER

What *is* The Cloud and Why Should I Care?

Strategy for Genealogists — Part 2

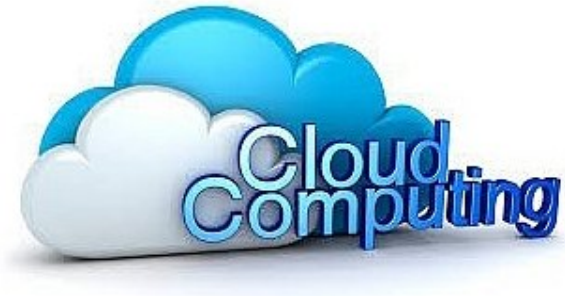
Guest Column by Dick Eastman

Many AGS members receive Eastman's Online Genealogy Newsletter, *the free genealogy blog* by email*; Dick Eastman provides other services via a paid subscription to his Plus Edition. Included recently was his series on Cloud Computing.*

In the August Quarterly*, we presented Part 1, which explained what the cloud is, outlining the width and breadth of this new computing technology advancement. In this* Quarterly*, Mr. Eastman details using genealogy applications in the cloud, and how this concept can benefit genealogists. Again, we re-print with the permission of the author.*

Mike Blackledge

Probably the first cloud application to appear was email service; the second to become popular was that of making airline reservations. A third cloud-based application soon appeared: off-site file storage. Today, as disk drive prices have dropped, dozens of cloud-based services provide the capability to store **DUPLICATE** copies of files from a user's personal hard drive. Ideally, the user should have one copy stored locally on his or her hard drive, and then a duplicate copy of the same file should also be stored on one or more remote file servers. The more popular cloud-based file storage services of today include Dropbox, Google Drive, Microsoft OneDrive, and SpiderOak. Dozens of other companies provide file protection services in the cloud as well. [Mike's note: just this year, OneDrive has reduced their **free** storage from 15 GB to 5 GB].



Various backup services also store files but with extra capabilities of storing multiple versions of the files. These are useful if you need to retrieve a file as it existed days or weeks ago. Cloud-based backup services include Mozy, Carbonite, BackBlaze, and CrashPlan.

These days, other applications of all sorts have moved to the cloud. Google Drive (formerly called Google Docs) now provides a word processing application (loosely similar to Microsoft Word), a spreadsheet application (loosely similar to Microsoft Excel), a presentation application (loosely similar to PowerPoint), and other functionality in the cloud. Other vendors do the same. All programs, as well as the data, are stored in the cloud. The user simply opens a web browser and accesses everything he or she needs. Another common variation is storing a small, low-powered program, or "app," on the computer in place of using a web browser. This is especially popular in handheld tablets and smartphones. The powerful servers in the cloud provide all the heavyduty computing tasks and data retrieval. The small app in the handheld or desktop computer simply provides a customized screen display and automatic sending and retrieving of information.

Genealogy applications in the cloud

All of the previously mentioned services qualify as cloud applications. To be sure, genealogists can use cloud-based backup programs to protect their valuable files or use a cloud-based word processor for writing tasks. However, let's look at a few cloud applications designed especially for use by genealogists.

People-centric public collaborative genealogy web sites

Several genealogy web sites allow users to add, modify, or delete information about ancestors via a web browser. They all are designed to let users collaborate, or pool, their knowledge together. Not all of them qualify as cloud applications, however.

MyHeritage.com, *WeRelate.org*, *WikiTree.com*, *Geni.com, FamilySearch.org, Ancestry.com, Findmypast.com, OneGreatFamily.com*, and a number of similar services collect information from many users and attempt to merge all the records into one large, inter-connected family tree on each service. The long-term goal of most of these sites is to collect all the known information about each person who ever lived and left records behind. By merging together all the information contributed by users and, in some cases, adding images of original records or links to original records stored on other sites, each service hopes to provide one large family tree of all mankind.

All data contributed by users can be "groomed" by others with many of these online services, although not by all. In the case of erroneous information, a later user may collaborate by entering his or her version of the information.

All of the above collaborative genealogy sites are unique. No two of them are exactly alike, and no two of them operate in exactly the same manner. For instance, one genealogy web site may allow you to upload a GEDCOM file of hundreds, or even thousands, of ancestors. Another might not allow GEDCOM uploads at all and instead will require manually entering information about one person at a time. One genealogy site might allow you to change information entered by another user while a different site may not.

Perhaps a bigger – and more important – difference is how each site handles conflicting information. If the online data about one individual is incorrect, or if it is disputed by two or more users with different records of the event in question, some genealogy services simply allow the new user to REPLACE information entered by the original user. Of course, this frequently leads to "online wars" over whose information is correct.

A better and more elegant approach (in my view) is to APPEND new information, not replace an existing entry. There may be two, three, or even more versions of a single event. Each user is able to enter what he or she believes is the truth and also is invited to upload corroborating evidence, such as an image of a birth certificate, death record, marriage certificate, census record, military pension application. In this case, each person who wishes to collaborate with others gets to present information to "plead" his or her case, similar to presenting evidence in court. However, the only judge in this case is the future genealogist who reads all the entries on the person's record and then makes up his or her own mind as to which version is correct, if any.

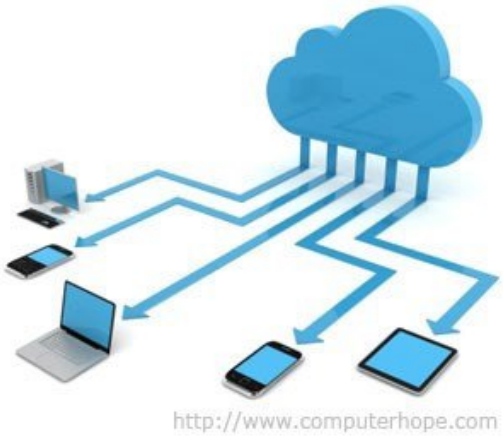
Some genealogy collaborative web sites allow for introduction of contradictory evidence. However, some use the more simplistic approach of "the latest entry is the correct one." As with all genealogy information found in books and online, the reader must be aware that these are CLAIMS, not proof. However, claims that are accompanied by supporting evidence (images of a birth certificate, death record, marriage certificate, census record, military pension application, etc.) will receive more credence. The reader is always left to accept or reject each claim, based upon separate research else-where.

While different from each other, most of the collaborative genealogy web sites do share a few attributes. For instance, all the previously entered information about any one individual is displayed on one web page. All other persons known to be associated with that individual (spouses, parents, siblings, children, employers, godparents, and others) are listed on the page as hyperlinks. Clicking on a listed

person's name changes the page to display all the known information about the person whose name was just clicked. In this manner, you can "walk" up or down a family tree rapidly.

On many services, users can contribute pictures of the individual or digital images of supporting documentation. The pictures normally are displayed on the same web page as the individual's information. In the same manner, source citations may be displayed on the same page. Which is better: a statement that the person appeared in the 1850 census on a certain volume and page number or an image of that census page showing the enumerator's handwriting for the individual of interest? Most genealogists will prefer the image.

In cases where the contributor doesn't have an image of the source record, or if copyright issues prohibit the uploading of an image created by someone else, a link to the original image or book or document or written source citation of the original record may be used. Future users can click on the link to view the cited source.



Again, all of these genealogy web sites are based upon data from individual genealogists who work together (or collaborate) to create databases that far exceed the capabilities of any one genealogist working alone. For details of how each genealogy web site operates, you need to look at the help files on each site.

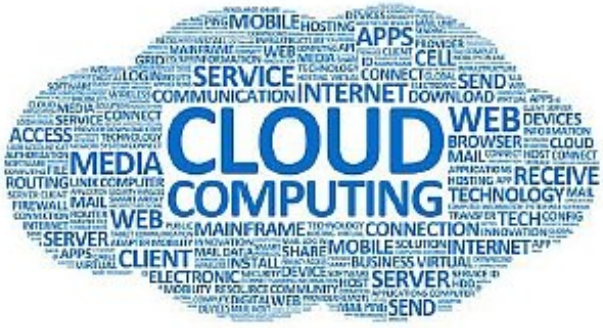
Which site is the best one for you? There is no easy answer. While I might prefer the operation of one web site over another, you may have a different preference. One factor is that one web site might contain more information about your ancestors than does another. In short, I recommend you try all of them to see which one suits you best. Indeed, there is no rule about restricting your usage to just one genealogy web site; you can use as many as you wish! I would suggest you contribute your known and documented family tree to each of the sites. The more places you use to publish the information, the greater the chances that your info will either help another genealogist or connect you to more info about your own family, if not both.

All of these are excellent examples of cloud-based, COLLABORATIVE services. They provide excellent knowledge management services that benefit the maximum possible genealogists.

Create your own genealogy web site

The previously mentioned online services are mostly owned by large corporations or non-profit organizations that manage the services and dictate how the software is designed and used. Some of them allow a user to keep his or her information private. However, the expectation is that most users will contribute their information publicly with the hope of assisting other genealogists and possibly finding others who are researching the same families. Indeed, these online services do provide a great service. However, one size does not fit all.

Some genealogists may prefer to use their information in a different manner. Perhaps they want to keep information private or to share it with only a few members of their own family. Some non-profits, especially family surname societies, may wish to restrict access to their own members. Still others may want to share publicly but in a different manner from the larger sites. As is true in many activities in life, if you don't like the way others do it, you can always do it yourself.



A discussion of creating your own genealogy web site typically will include both cloud-based and non-cloud-based solutions. The software used to create personally owned genealogy web sites typically can be installed in either cloud computing environments or on a single server. However, the solutions discussed here do emulate cloud computing in one manner: the software used to run the genealogy database is installed on the server(s) as well as the data. As with cloud-based applications, no special software needs to be installed in the user's computer. The products to be discussed here will work on Windows, Macintosh, Linux, as well as iPad and Android tablet computers. They also will work on iPhones and other smartphones although the small screen size of smartphones may be a limiting factor.

Again, these products can be installed in a true cloud, using multiple servers in multiple locations. However, it seems to be much more common to install them on a single server, even though the required software does not reside in the user's computer. Because the software resides on web server(s), not in the user's computer, these online applications appear to be cloud-based whether or not they are in-stalled on multiple servers in multiple data centers around the world.

You or your society can create an online database for storing and displaying genealogy data. You can make this visible to an audience of one (yourself) or to a list of people that you define. Access to the restricted information normally requires a user name and password. The creator of the data-base controls password distribution and access to the data.

Products for creating genealogy web sites are very popular amongst family societies, such as an organization devoted to researching all the descendants of one couple or surname societies devoted to researching the ancestry of everyone born with a particular surname.

The two best known online genealogy programs used by individuals are software products called **The Next Generation of Genealogy Site Building** (often referred to as "TNG") and **WebTrees**. These are programs that must be installed on web servers, not in individual Windows or Macintosh computers. Once the software is installed, data may be entered manually or by uploading GEDCOM files. Data may be restricted in any of a number of ways. Access might be available to only one person (the owner) or restricted to society members or to a few cousins or, optionally, to everyone in the online world. Different levels of access also might be available, such as giving read-only access to a large group and restricting the ability to add or change information to one person or to a small group.

Somebody somewhere needs to possess technical knowledge to obtain the required software and install it on a web server. However, the owner of the system does not need to be that technical expert. Several companies will install the software for a web site owner on their own servers and then charge that owner $5 or $6 a month for access. These same hosting services will normally install any new software up-grades and will troubleshoot problems, if any. The web site owner only needs to be concerned about one thing: the in-formation. When using someone else to install and maintain the software, the web site owner needs only modest technical skills.

Again, these two programs look like cloud applications. The software is in the cloud, not installed in the user's Windows, Mac, or other hardware. When the user wishes to use the application, he or she connects to the server(s) and runs the software directly from that server.

In reality, these two programs can be installed on single servers in a data center someplace or in a cloud environment on multiple servers spread across multiple locations around the world. Since these are typically privately owned or owned by a non-profit organization and because there usually is no need to have 99.9999% uptime, most copies of TNG and of WebTrees are installed on single servers, not in a cloud. As such, it is the responsibility of the person or organization that owns the web site to know where their pro-gram and data reside so that they can provide for backup if their web site resides on a single server that is not otherwise protected.

More information about these host-your-own products can be found at *http://www.tngsitebuilding.com* and at *https://www.webtrees.net/index.php/en*.

A Special Word about Ancestry.com:

Defining the online giant called *Ancestry.com* in a single sentence is probably an impossible task. Indeed, it is not a single service. Instead, it is a COLLECTION of somewhat related services. A few of the *Ancestry.com* services may be true cloud services while the majority are not. The majority of services on the *Ancestry.com* web site are not spread across multiple data centers using technology that instantly transfers users to a different server in a different location when one server fails.

In its earliest days of the late 1990s, *Ancestry.com* was NOT a cloud service. Indeed, the phrase "cloud computing" had not yet been invented. *Ancestry.com* started as a data retrieval service, providing ONE-WAY transfers of requested information. The data available in those days included transcribed genealogy data. As the years went by, this one-way transfer of genealogy information expanded to include images of census and other valuable records, digitized family history books, and much more. *Ancestry.com* also purchased other services, including *RootsWeb*, a message board and file storage service. While all these services continue to be valuable, they do not qualify for the definition of cloud computing because they are located in a single data center owned by Ancestry.com, not installed in redundant servers in multiple locations.

Other genealogy-related web sites

Of course, more content than "databases of people" is available in the online genealogy world today. Some of the other sites may be cloud-based (installed on multiple servers in multiple data centers in widely-dispersed locations) while others are not.

In the **FamilySearch Research Wiki** at *https:// www.familysearch.org/learn/wiki/*, you can learn how to do genealogical research or share your knowledge with others. You will not find information about people in this web site. Instead, you will find articles about getting started in family tree research, selecting genealogy software, where to find records, understanding census forms, where to find Family History Centers, and much more. This online site is a virtual encyclopedia, containing thousands of "pages," and yet it is available even in a tiny smartphone. That many printed pages will not fit into a small device, but accessing digital pages is easily handled by even the smallest device.

The **Encyclopedia of Genealogy** is just what it says: an encyclopedia, not a database of people. (**Disclaimer:** I created and started the Encyclopedia of Genealogy although most information added since those days has been contributed by users, not by myself. It is not a cloud-based service.) The Encyclopedia of Genealogy serves as a compendium of genealogical tools and techniques. It provides reference information about everything in genealogy except people. Look to the Encyclopedia of Genealogy to provide explanations of how to look up your family tree and explanations of terms found in genealogy research, including obsolete medical and legal terms. It will describe locations where records may be found. It also will describe how to research Italian, German, Polish, French-Canadian, Jewish, Black, Indian, and other ancestors. It lists information about many genealogy societies. In short, the Encyclopedia of Genealogy will serve as your standard genealogy reference manual. The Encyclopedia of Genealogy may be found at *http:// www.eogen.com.*

You can find many more genealogy web sites in the many lists on Cyndi's List at *http://www.cyndislist.com*.

Summation

The cloud is a bit difficult to describe with precision. Everything in the cloud is web-based, but not everything on the web can be considered cloud computing. There are dozens of "borderline" web sites that may or may not fit the exact definition of cloud computing: entrusting remote services with a user's data, software, and computation services. That last phrase, "computation services," is what typically differentiates the cloud. Lots of web sites will provide data. However, cloud-based services typically store your data and then add extra computing power above and beyond what your own computer can do alone by supplying software pro-grams. The end user's computer simply provides a window into the powerful server(s) located in a distant data center.