The Release of FamilySearch GEDCOM 7.0 with GEDZip and the Future of GEDCOM as a Universal File Specification for Sharing Family Tree Data

Syllabus presented by James L. Tanner, member of the FamilySearch GEDCOM Steering Committee
For the BYU Family History Technology Workshop
February 27, 2023

Introduction

GEDCOM is an acronym for GEnealogical Data Communication. GEDCOM is also the name of a standard programming specification usually referred to as the "GEDCOM Standard or GEDCOM" created by [The Church of Jesus Christ of Latter-day Saints](#) and [FamilySearch](#) as the specifications of a file standard used for exchanging genealogical data between different desktop genealogical family tree software and websites. The GEDCOM Standard Specifications are designed to be used by programmers when modifying their own programs and websites to share information between different genealogy software and websites. FamilySearch is still the owner of both the GEDCOM Copyright and the Trademark.

FamilySearch GEDZip has been developed to complement the updated GEDCOM Standard to transmit a GEDCOM document together with a set of external files. A FamilySearch GEDZip file is a Zip archive, as defined by [http://www.pkware.com/appnote and standardized by ISO/IEC 21320-1:2015](#). See [https://gedcom.io/gedzip/](#)

A Very Short History of the GEDCOM Standard

The GEDCOM Standard was developed because of a problem that became apparent with the rapid developments in computer technology beginning with the introduction of the first "desktop" computers beginning in about 1974 with the Altair 8800. As desktop computers became more powerful, two main competing operating systems developed: Microsoft Windows and Apple OS. As far as genealogical data was concerned there was no practical way to connect two computers together or exchange genealogical data between the competing systems. It became apparent almost immediately with the development of more sophisticated genealogy software that there needed to be a way to transfer the data from one computer to another, i.e., from one desktop computer to another and from one operating system to another such as from DOS/Windows to the Apple OS. In 1984, The Church of Jesus Christ of Latter-day Saints released the first version of GEDCOM or **GE**nealogical **D**ata **COM**munication. From 1984 to 1996 different versions of the GEDCOM Standard paralleled the technological advances in computers.

Beginning in 1982, I was the owner and operator of an Apple Computer dealership in Mesa, Arizona. We sold a variety of computer models from different manufacturers, not just Apple. About that same time, I began my continuing interest in genealogical research. As a result of that interest, I became aware of the GEDCOM Standard shortly after Personal Ancestral File (PAF) 2.0 was released in April of 1986. At that time, the GEDCOM Standard had been available since

1984. Another version of GEDCOM for PAF 2.1 was released in February 1987. PAF 2.1 was also released for the Macintosh in 1987 with supported and early specification of GEDCOM 4.0.

Development of the GEDCOM Standard by the Church stopped with the release of a draft version 5.5.1 on 2 October 1999. The draft version was finally released as Version 5.5.1 on 15 November 2019. Some third-party developers published updates to Version 5.5.1 and a Version 6.0 XML draft was also circulated back in 2002. Online discussion about an upgrade and some meetings were held at the annual RootsTech Conference. In 2020, a Steering Committee was formed to continue the development of the GEDCOM Standard to add features that were being implemented in many of the online and desktop genealogy programs.

Challenges of developing a version of the GEDCOM Standard after 1999

Although there had been a rapid and phenomenal development of processing power to 1999, data storage was lagging, and it was not until the advent of the internet that data storage became a major issue. Some illustrations might help in understanding the need and ability to update the GEDCOM Standard.

In 1976 when the Apple I was introduced it cost $667. Adjusted for inflation, today it would cost $3,066.35. When I bought an Apple II computer in 1977, it cost $1,298 which comes out to $5,602.86 which is much more than I spent for a new 10-core iMac this past year. One more example, the Apple Macintosh was introduced in 1984 for a price of $2,495. Adjusted for inflation, that is the equivalent of $6,281.49. See U.S. Inflation Calculator.

I think another helpful illustration of the change involves a single digital photo. Back in 1981 a gigabyte of storage cost about $500,000. See "Hard Drive Cost Per Gigabyte." I can now buy a 16 Terabyte hard drives for about $279.00 or about $17 for a terabyte or 1000 gigabytes of storage. So, my actual cost of storing a gigabyte is $17 divided by 1000 or about $.02 per gigabyte. Oh, by the way, a 3.5" floppy disk could store 1.44 megabytes which is less than the memory size of one of my digital photos.

The other main issue is speed. Again, I am not anywhere near the average but here in Provo, Utah we have Google Fiber Internet and I have a very high-speed connection.

These rapid advances in technology allowed the developers of genealogical software and websites to incorporate digitized documents, photographs, and sophisticated search options. However, many genealogists who use these features are generally interested in local ownership and control of their personal family history information and despite the proliferation of online family tree websites, there is still a significant reliance on local, desktop programs. As time passes, the ways these differences were implemented by programmers by the genealogy websites and desktop programs formed an increasing barrier to the exchange of information. As result, there is a continual loss of valuable information.

All the previous versions of the GEDCOM Standard encountered the same implementation differences. The variations in the data between websites and desktop programs were not accounted for in the last version of the GEDCOM Standard or the GEDCOM Standard was not

followed by the developers of the different programs. These differences resulted in data loss during an exchange.

The internet's ability to localize the world showed that traditional genealogical methodology was insufficient to reflect the different social and cultural structures that were now reflected on the internet. Western European based genealogical websites and programs needed to adapt to a globalization of genealogical data. Any attempt at a genealogical standard had to account for the differences in recording names, calendar dates, places, references to time, family relationships, and kinship systems.

By 2019, it was time to update the GEDCOM Standard to Version 7.0 with GEDZip and begin to incorporate additional standards for the sharing genealogical data across the internet as reflected by changes. It was also important to make sure that any new version would not "lose ground" by failing to incorporate the essential features of all the previous standards.

The development of FamilySearch GEDCOM Version 7.0 and GEDZip.

Obviously, in drafting an update to GEDCOM 5.5.1, the Steering Committee had to take into consideration that the previous versions of the GEDCOM Standard had become a *de facto* standard in the worldwide genealogical community. The passage of time alone imposed a significant burden due to technological changes. All the previous versions of the GEDCOM Standard were created by collaboration with key developers and influencers in the greater genealogical community and this was continued with the development of the GEDCOM 7.0 Standard. During the development, access to the code and all changes has been made available to the genealogical community through FamilySearch/GEDCOM.io and reviews in GitHub.com. Input was and continues to be solicited from interested developers and influencers and the issues identified were and are being constantly reviewed in weekly meetings.

The focus of the new upgrade was and continues to be resolving inconsistencies, poor selected examples, poor organization, spelling, outdated tags, making popular custom tags into standard tags and other efforts to streamline and make the GEDCOM Standard more universally effective for data transfer. There is also an effort to make the documentation including notes easier to understand.

As of the date of this syllabus, if you were to subscribe to one of the major online genealogy family tree/database websites such as Ancestry.com, you would be able to upload your basic genealogical data from your desktop software program using GEDCOM but that would not include any of your digital images including photos. You could upload your photos one by one, but then you would have to tag or attach them individually to your new family tree. This challenge was the driving force for adding the ability to download and upload digital documents and photographs attached as sources or memories to a family tree. This was resolved by the development of FamilySearch GEDZip to enable transmittal a GEDCOM document together with a set of external files using GEDCOM fil.

The FamilySearch GEDCOM Version 7.0 with GEDZip was released on 27 May 2021. Incremental additions and changes to the Standard continue to be made.

We now find that individuals use GEDCOM to share portions of their family tree for collaboration, reports, charts, special analysis, and other innovative purposes.

Collaboration, sharing, and preservation

As long as separate and distinct systems of recording genealogical data exist, there will still be a relevant need for a GEDCOM Standard and there will also be concomitant need for continued upgrading and improvement. In addition, valuable utilities and tools can be built to support a Data File standard.

For the average genealogist using a relatively recently upgraded computer, the new GEDCOM Standard 7.0 release candidate will only be available as the genealogy software companies and websites implement its use to enable the genealogists to use it. First, it is a standard. That means that developers and programmers must decide to incorporate the standard in their software so that genealogists can exchange copies of the documents and images attached or reference in the software or websites they are using.

The idea behind updating the GEDCOM Standard to Version 7.0 release candidate and adding the ability to support external images using GEDZip is that a genealogist can upload or exchange files that include all those records, documents, and photos already attached and with copies included.

GEDZip is an added feature of FamilySearch GEDCOM version 7.0. A standard Zip archive file is formatted to combine existing family tree data with external images and other files. The GEDZip file can be unzipped after transfer, making the genealogical data, and associated external files available together in a different software product. The exact procedures for downloading or uploading external files will vary according to the individual software product's implementation of the FamilySearch GEDCOM specification. GEDZip packages **multimedia** files and the **GEDCOM** file in a single compressed ZIP archive thus increasing the value of both GEDZip and GEDCOM. The file extension for a GEDZip file is: ".gdz".

To view the present implementation progress, See
https://www.familysearch.org/en/GEDCOM/implementation-progress#home

For information about the progress see the FamilySearch GEDCOM Changelog.
https://gedcom.io/changelog/

There is still significant work that needs to be done, including the following:
- More relationship types between persons
- Better Name and Place parts
- Expanded Sourcing
- Hypothesis to share valuable points of collaborations that are not yet record supported facts
- Improved ability to add and reference proprietary tags

The GEDCOM Steering Committee is certainly open to any suggestions both to correct existing issues and for improvements in the future. We welcome your participation in the GEDCOM community, and we are particularly interested in those who can volunteer their time to review the GEDCOM standard and, in addition, for programmers who wish to help with further coding. We are most interested is someone to help with development of a SAX-like parser for GEDCOM 7.0. For more information about volunteering see https://gedcom.io/

If you would like to join a team for future development to focus on relationships, places, sources and citations, and hypotheses, you can send an email to GEDCOM@familysearch.org putting the project name in the subject.

Sources

"A Brief History of the Internet." n.d. *Internet Society* (blog). Accessed February 4, 2023.
    https://www.internetsociety.org/internet/history-internet/brief-history-internet/.
Clarke, Gordon. 2021. "GEDCOM 7.0 Standard Promotional Strategy for FamilySearch Management."
    Circulated Document.
"FamilySearch GEDCOM - Genealogical Data Exchange." n.d. FamilySearch GEDCOM. Accessed
    February 4, 2023. https://gedcom.io/.
"FamilySearch GEDCOM • Genealogy." n.d. Genealogy. Accessed February 4, 2023.
    http://familysearch.org/en/GEDCOM/gedcom-the-standard-file-format-for-family-history.
"FamilySearch GEDZip." n.d. FamilySearch GEDCOM. Accessed February 4, 2023.
    https://gedcom.io/gedzip/.
"GEDCOM." 2023. In *Wikipedia*.
    https://en.wikipedia.org/w/index.php?title=GEDCOM&oldid=1137433695.

General information can be found at GEDCOM.info.
Helpful Sources
General Info: GEDCOM.info
Technical Specs, Tools and Guides: GEDCOM.io
Community: GEDCOM General Google Group and GitHub Public GEDCOM Repository
Email: GEDCOM@FamilySearch.org