

Genealogy Basics

GEDCOM 101 (Part 1)

What exactly is a GEDCOM and how do I use it?

One of the biggest advantages to using the Internet for genealogy research is the ability it provides to exchange information with other researchers. One of the most common methods used for this information exchange is the GEDCOM, an acronym for **GE**nealogical **D**ata **COM**munication. In simple terms it is a method of formatting your family tree data into a text file which can be easily read and converted by any genealogy software program. The GEDCOM specification was originally developed in 1985 and is owned and managed by the Family History Department of the Church of Jesus Christ of Latter Day Saints. The current version of the GEDCOM specification is 5.5 (as of November 1, 2000) but the LDS Church will soon be releasing version 6.0XML.

A GEDCOM specification uses a set of TAGS to describe the information in your family file, such as INDI for individual, FAM for family, BIRT for birth and DATE for a date. Many beginners make the mistake of trying to open and read the file with a word processor. Theoretically, this can be done, but it is a very tedious task. GEDCOMS are best suited for opening with a family tree software program or a special GEDCOM viewer (see related resources). Otherwise, they basically just look like a bunch of gibberish.

Anatomy of a GEDCOM File

If have ever opened a GEDCOM file using your word processor, you have probably been faced with a seeming jumble of numbers, abbreviations, and bits and pieces of data. There are no blank lines and no indentations in a GEDCOM file. That's because it is a specification for exchanging information from one computer to another, and was never really intended to be read as a text file.

GEDCOMS basically take your family information and put it in an outline format. **Records** in a GEDCOM file are arranged in groups of lines that hold information about one individual (INDI) or one family (FAM) and each line in an individual record has a **level number**. The first line of every record is numbered zero (0) to show that it is the beginning of a new record. Within that record, different level numbers are subdivisions of the next level above it. For example, the birth of an individual may be given level number one (1) and further information about the birth (date, place, etc.) would be given level two (2).

After the level number, you will see a descriptive tag, which refers to the type of data contained in that line. Most tags are obvious: BIRT for birth and PLAC for place, but some are a little more obscure, such as BARM for Bar Mitzvah.

A simple example of GEDCOM **records** (my explanations are in blue):

Editor's note: for our black and white publication, the comments are grey-shaded - sdj

0 @I2@ INDI <- The level 0 means this is a new record. The INDI tag means the record is for an individual.

1 NAME Charles Phillip /Ingalls/ <-name of individual

1 SEX M <- level 1 info denotes data about the individual in level 0, Charles Ingalls

1 BIRT

2 DATE 10 JAN 1836 <- this is level 2, because it refers to the birth in level 1 above
2 PLAC Cuba, Allegheny, NY
1 DEAT
2 DATE 08 JUN 1902
2 PLAC De Smet, Kingsbury, Dakota Territory
1 FAMC @F2@
1 FAMS @F3@
0 @I3@ INDI <-A new record (level 0) for a new individual
1 NAME Caroline Lake /Quiner/
1 SEX F
1 BIRT
2 DATE 12 DEC 1839
2 PLAC Milwaukee Co., WI
1 DEAT
2 DATE 20 APR 1923
2 PLAC De Smet, Kingsbury, Dakota Territory
1 FAMC @F21@ <-FAMC indicates the FAM record where this person appears as a Child
1 FAMS @F3@ <-FAMS indicates the FAM record where this person appears as a Spouse

Tags can also serve as pointers (@I2@), which indicate a related individual, family or source within the same GEDCOM file. For example, a family record (FAM) will contain pointers to the individual records (INDI) for the husband, wife and children.

Here is the family record which contains Charles and Caroline, the two individuals discussed above:

0 @F3@ FAM <- this is family record #3, pointed to from the above individual records
1 HUSB @I2@ <- this is a pointer to individual record (INDI) I2, for Charles Phillip Ingalls
1 WIFE @I3@
1 MARR
2 DATE 01 FEB 1860
2 PLAC Concord, Jefferson, WI
1 CHIL @I1@ <- these point to the individual records for the children
1 CHIL @I42@
1 CHIL @I44@
1 CHIL @I45@
1 CHIL @I47@

As you can see, a GEDCOM is basically a connected database of records with pointers which keep all of the relationships straight. While you should now be able to decipher a GEDCOM with a text editor, you will still find it much easier to read with the appropriate software.

Source: <http://genealogy.about.com/library/weekly/aa110100a.htm>

Editor's note: Stay tuned for Part 2 in the next issue - sdj