

January 23, 2002

Dear GEDCOM Developer:

This letter accompanies a draft version of the GEDCOM XML Specification, Release 6.0. This draft is not a complete specification. For example, descriptions of the meaning and expected contents of tags are not included. However, the Document Type Definition (DTD) and extensive examples are included. We feel that is enough to give a reasonable understanding of the specification.

If you have any comments or suggestions concerning this specification, please E-mail them to [gedcom@gedcom.org](mailto:gedcom@gedcom.org). In the past we have received useful and significant comments from the genealogical community, and we welcome your suggestions.

Bear in mind that this is a draft document. It has been reviewed by our development staff internally, but we expect that suggestions from the larger genealogical community will lead to significant improvements and changes. We do not recommend developing to this specification at this time. Hopefully, we can incorporate changes quickly and have a more stable specification in the near future.

There have been previous GEDCOM 6.0 drafts available at various times. This draft is quite different, and supercedes those. They were more elaborate and ambitious, and were not well received, either within or outside our organization.

Thank you for your continuing interest and support.

# GEDCOM XML

## Specification

Release 6.0

*Draft*

Prepared by the  
Family History Department  
The Church of Jesus Christ of Latter-day Saints

December 28, 2001

### Suggestions and Correspondence:

Email:  
GEDCOM@gedcom.org

Telephone:  
801-240-3442

Mail:  
Family History Department  
GEDCOM Coordinator—3T  
50 East North Temple Street  
Salt Lake City, UT 84150  
USA

Copyright © 2000 by The Church of Jesus Christ of Latter-day Saints. This document may be copied for purposes of review, provided this notice is included. All other rights reserved. This draft version should not be used for programming genealogical software.

## TABLE OF CONTENTS

Introduction .....	5
Purpose of the GEDCOM XML Specification .....	6
Chapter 1	
Comments and Examples .....	7
Extensibility .....	7
Design Objectives .....	7
Consistency Between Traditional GEDCOM and XML GEDCOM .....	7
Some Implementation Choices .....	8
Items in Traditional GEDCOM Which are Not Carried Over to XML .....	8
Linked Record Structure of GEDCOM XML .....	9
Data Structure Diagram .....	10
Use of XML Document Capabilities .....	11
Sample GEDCOM XML .....	14
Chapter 2	
GEDCOM XML Definitions .....	23
Introduction .....	23
GEDCOM XML Document Type Definition (DTD) .....	23



## Introduction

GEDCOM was developed by the Family History Department of The Church of Jesus Christ of Latter-day Saints to provide a flexible, uniform format for exchanging computerized genealogical data. GEDCOM is an acronym for **GEnealogical Data Communication.**

To illustrate the advantages of GEDCOM, consider the data below, which is in a traditional fixed record format:

John Henry12 OCT 1954Los Angeles, Calif.

We might assume this is data about a person named John Henry, and a date and place associated with him. Perhaps when and where he received a drivers license Or it might be about the publication of a song called “John Henry.” GEDCOM adds tags, or labels, to the data which give an indication of its meaning. It, also, indicates how data values are related. The same data in a GEDCOM file might look like:

```
0 INDI
1 NAME John Henry
1 BIRT
2 DATE 12 OCT 1954
2 PLAC Los Angeles, Calif.
```

or indented to make it more readable:

```
0 INDI
  1 NAME John Henry
  1 BIRT
    2 DATE 12 OCT 1954
    2 PLAC Los Angeles, Calif.
```

One familiar with GEDCOM knows that this is data about an individual, his birth date and birth place. The “level” numbers identify the date and place as being related to the birth. This is a more robust and flexible way to communicate the information than with the original fixed format record. Data can be rearranged, and new data added without losing meaning. For example the original data can still be easily identified in:

```
0 INDI
  1 ALIAS James Horning
  1 BIRT
    2 PLAC Los Angeles, Calif.
    2 DATE 12 OCT 1954
  1 NAME John Henry
```

Fixed format records don’t have this level of flexibility.

With the widespread use of the internet, its users recognized this same need for robust, structured data exchange. In response, the World Wide Web Consortium (W3C) developed the Extensible Markup Language (XML). It serves the same purpose and has the same capabilities as GEDCOM. It uses tags, but adds matching ending tags. It indicates the relationships or structure of data by nesting instead of level numbers. In XML, our original data might look like:

```

<INDI>
  <NAME>John Henry</NAME>
  <BIRT>
    <DATE>12 OCT 1954</DATE>
    <PLAC>Los Angeles, Calif.</PLAC>
  </BIRT>
</INDI>

```

The line breaks and indentation are used for human readability, and would normally not be included. The data would likely be transmitted as:

```

<INDI>  <NAME>John Henry</NAME><BIRT><DATE>12 OCT 1954</DATE><PLAC>Los Angeles,
        Calif.
</PLAC></BIRT></INDI>

```

While GEDCOM and XML are both capable of transmitting genealogical data, XML has some advantages. First it is a widely recognized standard. With traditional GEDCOM, new users had to get past understanding its peculiar syntax before they could understand its content. With an XML version of GEDCOM, a new user will likely already understand the syntax, and understanding the content is the only concern.

Being widely accepted, XML brings with it readily available utilities, tools and interfaces. For example, there are parsers, internet interfaces, and graphic design tools. XML also has established standards in areas where we have struggled in the past. This includes handling international character sets, and “whitespace.”

XML has roots in the automated document world, which allows opportunity to better handle textual genealogical data. This will be discussed in more detail in the next chapter.

In short, XML has significant advantages over traditional GEDCOM. To move GEDCOM to an XML format is an obvious step in the right direction.

## **Purpose of the GEDCOM XML Specification**

This specification is a technical document written so that GEDCOM XML can be understood and utilized by computer program developers and by technically sophisticated users. An understanding of XML is assumed.

This document is available on the Internet at the following at:

www.FamilySearch.org > Family Search Questions > GEDCOM

## **Traditional GEDCOM**

For those interested in the traditional GEDCOM format, the GEDCOM 5.5 document can , also, be found in the site mentioned above.

# Chapter 1

## Comments and Examples

The primary purpose of this chapter is to help the reader understand GEDCOM XML content, use, and syntax before getting into the technical detail. It should give a good picture of the forest before becoming lost in the trees. Design objectives, examples, and data diagrams will be used to explain the GEDCOM XML specification.

### Extensibility

The primary strength of XML is in the “X”, that is, its extensibility. It allows application specific XML vocabularies. That is, it allows the definition of a well defined set of tags and structure suitable for a specific type of application. The *GEDCOM XML Specification* is an XML vocabulary for genealogical data.

### Design Objectives

XML serves two purposes:

- To facilitate the exchange of data in a structured, somewhat meaningful format, and
- To allow the processing and presentation of data with Web tools.

GEDCOM was invented to address the first item.

Addressing the second requires additional items such as style sheets and perhaps application code which understands genealogical data.

### Emphasis on the Exchange of Structured Data

Our objective in the design of the XML version of GEDCOM is to serve the original intent of GEDCOM, that is, to facilitate the exchange of structured data between genealogy applications. The complexity and subtlety of genealogical data is best handled by genealogical specific applications, as opposed to generalized Web tools. This specification only addresses the exchange of data between genealogical applications. It does not address the processing and presentation of genealogical data on the Web.

### Consistency Between Traditional GEDCOM and XML GEDCOM

Traditional GEDCOM has evolved over some 15 years or so. Technology, and use and understanding of genealogical data, has changed in that time. The specification in this document is based on newer concepts. It is not upward compatible with traditional GEDCOM. It is able to convey largely the same content as traditional GEDCOM, but in a cleaner, less ambiguous, and more robust structure. Some of the content of traditional GEDCOM which was marginal and not used has been dropped.

## Some Implementation Choices

### Specifying Linkage

An important part of GEDCOM is its ability to link records according to family lineage and other data relationships.

XML's standard linkage method using the ID and IDREF attributes, is equivalent to traditional GEDCOM's linkage method, and will be used in its place.

In traditional GEDCOM, links are bidirectional. For example, there is a CHIL tag in the FAM record that connects a family to a child, and a FAMC tag in the INDI record which connects a child to a family. There are also HUSB and WIFE tags in the FAM record, and in the opposite direction a FAMS tag in the INDI record which handles both spouses. To specify a link in both directions is, of course, redundant and unnecessary. Some programs produce traditional GEDCOM with links in one direction, some the other, and some give both. That makes it difficult to process GEDCOM from a variety of sources, and where both directions are specified, they may be inconsistent. In GEDCOM XML, all links are unidirectional, and can be specified in only one way.

### Attributes vs. Elements

In XML, information can be contained in elements or attributes. Elements can contain complex structure, while attributes can not. Consequently, data contained in an attribute cannot be easily expanded into a more structured form as requirements change. To illustrate, if a person's name was given as an attribute, it might appear as:

```
Name="Sgt. Henry James Clay, Jr."
```

If it was later decided that the parts of the name should be explicitly shown, there would be no simple way to do that. However, if the name was given as an element:

```
<IndivName>Sgt. Henry James Clay, Jr.</IndivName>,
```

theoretically, it could be expanded in an XML compliant manner to show additional structure, such as:

```
<IndivName><PersonalTitle>Sgt.</PersonalTitle><GivenName>Henry James  
</GivenName><SurName>Clay</SurName><NameSuffix>Jr.</NameSuffix></IndivName>.
```

Programs which did not understand the new tags could still get the same data as before by just ignoring the new tags. So elements are the more robust way of representing data.

Because of these considerations, GEDCOM XML uses attributes sparingly. One notable use of attributes is in the ID/IDREF record linkage mentioned earlier. Attributes are, also, used for some simple, unstructured data such as Type, Living, etc.

## Items in Traditional GEDCOM Which are Not Carried Over to XML

### Character Set

In the past ANSEL has been specified as the preferred character set to be used in GEDCOM. In GEDCOM XML, the UNICODE character set is used.



## Linked Record Structure of GEDCOM XML

A GEDCOM XML file contains a collection of related “records” which are “linked” together. It will normally contain Individual and Event records with each individual record linked to the events records in which he participated. Each individual will, also, be linked to ordinance records for his ordinances, and so forth. As was mentioned earlier, the XML ID/IDREF mechanism is used for linking related records. The following is an example of a husband and wife linked to a marriage event:

```
<IndividualRec Id="I0001">
  <IndivName>John Smith</IndivName>
</IndividualRec>

<IndividualRec Id="I0002">
  <IndivName>Jane Doe</IndivName>
</IndividualRec>

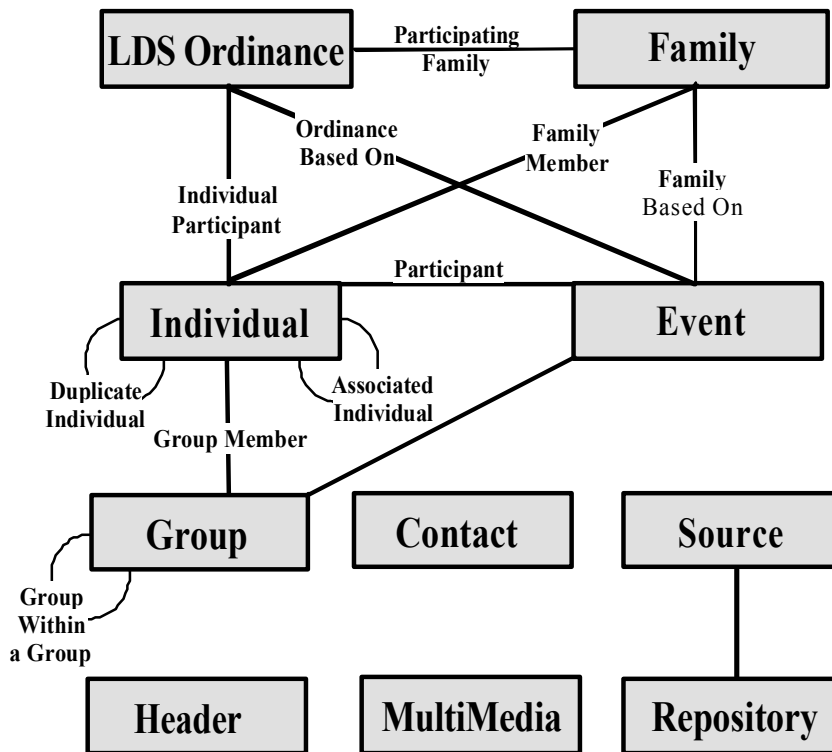
<EventRec Id="E0001" Type="marriage">
  <Participant>
    <Link Target=" IndividualRec" Ref="I0001"/>
    <Role>husband</Role>
  </Participant>
  <Participant>
    <Link Target=" IndividualRec" Ref="I0002"/>
    <Role>wife</Role>
  </Participant>
  <Date>7 NOV 1891</Date>
  <Place>Cove, Cache, Utah, USA</Place>
</EventRec>
```

In this example, the Event record is "linked" to the two Individuals who were married. The record ID's must be unique within the GEDCOM file. Using a different prefix for each record type can make this easier. A genealogy application which imports a GEDCOM file, would generally not keep the record ID's from the GEDCOM file, but would assign primary key values consistent with its database structure. In fact, keeping the GEDCOM ID's as primary keys in the application database would likely lead to duplicate keys after multiple imports.

## Data Structure Diagram

The data diagram below shows the various GEDCOM record types and how they are related (or linked) to one another.

### GEDCOM XML Data Model



**Source, Contact and MultiMedia records are referenced by most other records, so all the connections are not shown**

## Use of XML Document Capabilities

There is a significant difference between "data" and "text." To illustrate,

Name: John Henry  
Birth Date: 12 OCT 1954  
Birth Place: Los Angeles, Calif.

is an example of "data." It contains discreet pieces of structured, standardized data. On the other hand an example of text is:

"The prominent citizen of San Francisco, John Henry, was actually born in Los Angeles. According to his sister, Jane Franklin, he was born on the morning of October 12, 1954 at Grand View hospital."

Data oriented applications, such as genealogy programs, easily handle data, but have had more difficulty handling text adequately. But text, such as excerpts from sources, are important in genealogy. The markup languages, SGML, HTML, XML, etc., come from the document processing world, so by their nature they are text aware. They bring to data systems better techniques for handling text. In GEDCOM XML we will take advantage of this. If the text above is an extract from a cited source, it can be coded as:

```
<Extract>  
  <IndivDoc>The prominent citizen of <ResidenceDoc>San Francisco</ResidenceDoc>,  
  <NameDoc>John <SurNameDoc>Henry </SurNameDoc></NameDoc>, was actually born in  
  <BirthPlaceDoc>Los Angeles</BirthPlaceDoc>. According to his  
  <RelativeDoc><Relationship>sister</Relationship>, Jane Franklin</RelativeDoc>, he was born on  
  the morning of <BirthDateDoc> October 12, 1954</BirthDateDoc> at Grand View  
  hospital.</IndivDoc>  
</Extract>
```

By marking the text in this fashion, we could search for individuals, with specific criteria, within text excerpts. For example, "relatives of a Jane Franklin who lived in San Francisco." Obviously, such searches would be more difficult to formulate and would be less precise than searches on standardized, structured data. But they could be valuable for research..

Notice that the tag "IndivDoc" is used instead of "Individual." Different tags are needed for an individual within a document context than within structured data because allowable nested tags within them are quite different.

This specification does not address how to design a user interface capable of collecting textual information and marking it as shown in the example. But, if an application is clever enough to collect such information, GEDCOM XML will be able to faithfully preserve it.

Addresses have, also, been hard to handle. For example, consider

Mr. John Henry  
HIQ Corp.  
1234 Pleasant Dr., Suite 26  
Denver. CO, USA  
84095-0034

One approach is to break the address into discreet fields, and reconstruct the address when it is used. However, addresses can vary a great deal in content and structure, particularly when international addresses are needed. In this case, it is useful to use HTML tags for formatting. In GEDCOM XML, HTML tags are allowed. In this example, we will use HTML tags to cause line breaks. The address can be tagged as:

```
<MailAddress>Mr. John Henry<html:BR/>HIQ Corp.<html:BR/>1234 Pleasant Dr., Suite 26
<html:BR/> Denver, CO, USA<html:BR/>84095-0034</MailAddress>
```

For the sake of clarity, we will take some liberties throughout this document in displaying examples. In elements of XML where text is allowed, line breaks and tabs are preserved. However, in examples we will use tabs and line breaks which are not a part of the actual XML content. In this light the above example could be shown as:

```
<MailAddress>
  Mr. John Henry<html:BR/>
  HIQ Corp.<html:BR/>
  1234 Pleasant Dr., Suite 26 <html:BR/>
  Denver, CO, USA<html:BR/>
  84095-0034
</MailAddress>
```

But only the formatting explicitly specified by HTML tags are actually a part of the XML content.

This example could be further tagged to show the parts of the address as follows:

```
<MailAddress>
  <Addressee>Mr. John Henry </Addressee><html:BR/>
  <OrganizationName>HIQ Corp.</OrganizationName><html:BR/>
  1234 <Street>Pleasant Dr.</Street>, Suite 26 <html:BR/>
  <City>Denver</City>, <State>CO</State>, <Country>USA</Country><html:BR/>
  <PostalCode> 84095-0034</PostalCode>
</MailAddress>
```

This tagged address preserves both the format for mailing and identifies the data elements which might be useful for sorting, searching, building directories, etc.

The tagging above, however, is still biased toward United States geography. By using a neutral tag for parts of the address, and geographic level numbers, multinational data can be handled. The cultural name of the address (place) parts can be given in an attribute. This gives the following:

```
<MailAddress>
  <Addressee>Mr. John Henry </Addressee><html:BR/>
  <OrganizationName>HIQ Corp.</OrganizationName><html:BR/>
  1234 <PlacePart Type="Street" Level="6">Pleasant Dr.</PlacePart>, Suite 26 <html:BR/>
  <PlacePart Type="City" Level="4">Denver</PlacePart>,
  <PlacePart Type="State" Level="2">CO</PlacePart>,
  <PlacePart Type="Country" Level="1">USA</PlacePart><html:BR/>
  <PlacePart Type="PostalCode" Level="5"> 84095-0034</PlacePart>
</MailAddress>
```

This last example is how GEDCOM XML handles addresses. The smaller level numbers represent the larger geographic entities. Countries are always level 1, and cities or towns are always level 4. In the United States, states are level 2 and counties are level 3. In other countries, that will vary. Postal codes are level 5. Higher levels may be used for localities, farms, streets, etc., but are not standardized.

This specification does not suggest how an application will store the data in a database. (For example, an application may store the formatted address and the discreet data values of interest in separate fields.) But regardless of how it is used, GEDCOM XML can transmit both the format and content in a nonredundant manner.

Source citations are another challenging item to handle as discreet data fields, again because of the wide variety of formats. Also, they normally required a font change since titles are italicized. HTML tags will be useful here, also. Consider:

James Gardner, "Study of Early Birds," Vol. IV of *Birds of the Southwest* (1997), p. 14-16 in Fred Fredericks, *A Summary of Great Bird Studies* (San Jose: Big Johns Publications, 1986). Fredericks' work contains a complete reprint of the article by Gardner.

This can be marked up as:

**<CitationText>**

**<AuthorDoc>**James Gardner**</AuthorDoc>**, "**<ArticleDoc>**Study of Early Birds**</ArticleDoc>**," Vol. IV of **<TitleDoc>** **<html:I>**Birds of the Southwest**</html:I></TitleDoc>** (1997), p. 14-16 in **<AuthorDoc>**Fred Fredericks**</AuthorDoc>**, **<TitleDoc><html:I>**A Summary of Great Bird Studies**</html:I></TitleDoc>** (San Jose: Big Johns Publications, 1986). Fredericks' work contains a complete reprint of the article by Gardner.

**</CitationText>**

Though not illustrated in the earlier extract example, HTML tags could be used in extracts to preserve the original format, such as centered headings, bolding, italics, underlining, etc.

## Sample GEDCOM XML

The following is a sample GEDCOM XML file which illustrates most common content and structures. To see all the allowed content and structure, refer to the GEDCOM XML Document Type Definition in the following chapter.

For the sake of clarity, we will use line breaks, tabs, and bolding which are not a part of the actual XML data stream. In elements which allow text, any formatting which is a part of the XML data is explicitly coded with HTML tags. Ellipses, "...", are used to indicate data which may or should be included, but is omitted to keep the example as simple as possible.

```
<?xml version="1.0"?>

<!DOCTYPE GEDCOM SYSTEM "http://gedcom.org/dtd/gedxml60.dtd">

<GEDCOM xmlns:html="http://www.w3c.org/TR/REC-html40/">

<!-- Header Record -->

  <HeaderRec>
    <FileCreation Date="2 Oct 2000" Time="15:20:2.3">
      <Product>
        <ProductId>DAS</ProductId>
        <Version>6.3</Version>
        <Name>
          Deluxe Ancestral System
        </Name>
        <Supplier>
          <Link Target="ContactRec" Ref="CN001"/>
        </Supplier>
      </Product>
      <Copyright>Copyright 1999 So . . .</Copyright>
    </FileCreation>
    <Citation>
      <!-- This Citation is useful if the data in a file is from a single source -->
      . . .
    </Citation>
    <Submitter>...
      <Link Target="ContactRec" Ref="CN002"/>
    </Submitter>
    <Note>
      This material was prepared for the
      <IndivDoc><NameDoc><GivenNameDoc>John</GivenNameDoc>
      <SurNameDoc>Tate</SurNameDoc></NameDoc></IndivDoc> and <IndivDoc><NameDoc>
      <GivenNameDoc>Mary</GivenNameDoc> <SurNameDoc>Hamblin</SurNameDoc>
      </NameDoc></IndivDoc> Family Organization.
    </Note>
  </HeaderRec>
```

<!-- Family Records -->

```
<FamilyRec Id="FM001">
  <HusbFath>
    <Link Target="IndividualRec" Ref="IN001"/>
  </HusbFath>
  <WifeMoth>
    <Link Target="IndividualRec" Ref="IN002"/>
    <FamilyNbr>2 </FamilyNbr>          <!-- Her second marriage or family -->
  </WifeMoth>
  <Child>
    <Link Target="IndividualRec" Ref="IN003"/>
    <ChildNbr>1</ChildNbr>          <!-- Child's order in family, if birth dates unknown -->
  </Child>
  <Child>
    <Link Target="IndividualRec" Ref=". . ."/>
    <ChildNbr>2</ChildNbr>
    <RelToFath>adopted</RelToFath>
  </Child>
  <BasedOn>          <!-- Events justifying the creation of the family, and its members -->
    <Event>
      <Link Target="EventRec" Ref="EV001"/>
    </Event>
    <Event>
      <Link Target="EventRec" Ref="EV002"/>
    </Event>
    <Note>Comments on why this family was formed, and why particular members were
    included.</Note>
  </BasedOn>
  <ExternalID Type="AFN" Id="4S3469Q"/>
  <ExternalID Type="submitter" Id="F8945"/>
  <!--
    This is the ID used by the system which produced this GEDCOM file. It can be used to communicate
    changes, differing opinions, etc, to the file submitter.
  -->
  <Submitter>. . .</Submitter>
  <Note>. . .</Note>
  <Citation>
    <Source>
      <Link Target=" SourceRec" Ref="SR001"/>
    </Source>
    <!--
      A citation can be stored in free form, as shown below, or can be fabricated from the discrete
      fields which follow, combined with the Source and Repository records information
    -->
    <WhereInSource>File No. 7895-09</WhereInSource>
    <WhenRecorded>10 June 1903</WhenRecorded>
    <References> <!-- Used to cite the primary source, when this is a secondary source citation. -->
      <Citation>. . .</Citation>
    </References>
```

```

<CitationText>
  <TitleDoc>Marriage Certificate </TitleDoc> for <IndivDoc><NameDoc>John
  <SurNameDoc>Smith </SurNameDoc></NameDoc></IndivDoc> and
  <IndivDoc><NameDoc>Jane <SurNameDoc>Doe </SurNameDoc></NameDoc></IndivDoc>,
  <WhenRecDoc>10 June 1903 </WhenRecDoc>, <WhereInDoc>File No. 7895-09
  </WhereInDoc>, California State Board of Health, P. O. Box 2406, Sacramento, Calif.
  98065. Certified copy in possession of Larry T. Smith, Sandy, Utah.
</CitationText>
<Extract>Contains text extracted from the source. Same markup as Notes.</Extract>
<Note>Certified copy in possession of Larry T. Smith, Sandy, Utah.</Note>
</Citation>
<MultiMedia>
  <Link Target="MultiMediaRec" Ref="MM001"/>
  <Caption>We Attend the Kunzle Family Reunion</Caption>
  <WhereInMM>
    <Start>5:15</Start>
    <Stop>10:25</Stop>
    <Units>minutes</Units>
    <!--
      In a graphic, start and stop are the location of the upper left and lower right corners' x,y
      coordinates in inches, pixels, etc.
    -->
  </WhereInMM>
  <Note>Our family is featured at about 5 minutes into the video.</Note>
</MultiMedia>
<Changed Date="23 APR 1976" Time="13:25:12">
  <Note>Record created</Note>
</Changed>
<Changed Date=".." Time="..">
  <!-- The Contact here is the person responsible for the change -->
  <Contact>
    <Link Target="ContactRec" Ref=".." />
  </Contact>
  <Note>Adopted child added</Note>
</Changed>
</FamilyRec>

```

**<!-- Individual Records -->**

```

<IndividualRec Id="IN001">
  ...
</IndividualRec>

<IndividualRec Id="IN002">
  <IndivName Type="married">
    <PersonalTitle>Duchess </PersonalTitle><GivenName>Neta </GivenName>
    <MaidenName>Eskelson </MaidenName><von> <SurName>Allen</SurName>
    <IndNameVariation TranslationType="phonetic" Method="kana">
      ...
    </IndNameVariation>
  </IndivName>

```



```

    <IndNameVariation TranslationType="romanized" Method="romanji">
      ...
    </IndNameVariation>
  </IndivName>
  <IndivName Type="maiden" xml:lang="de">
    <GivenName>Neta </GivenName><SurName>Eskelson </SurName>
  </IndivName>
  <IndivName Type="also known as">
    ...
  </IndivName>
  <IndivName Type="nickname">
    ...
  </IndivName>
  <Gender>F</Gender>
  <DeathStatus>dead</DeathStatus>
  <AssocIndiv>
    <Link Target="IndividualRec" Ref="..." />
    <Association>first ancestor</Association>
    <!--
      How the associated person is related to this person. For example, the linked individual is my great
      uncle. The example shown is an oriental cultural requirement.
    -->
    <Note>...</Note>
    <Citation>...</Citation>
  </AssocIndiv>
  <DupIndiv>
    <Link Target="IndividualRec" Ref="..." />
    <Note>Explanation of why these individuals are considered duplicates</Note>
    <Citation>...</Citation>
  </DupIndiv>
  <ExternalID Type="..." Id="..." />
  <Submitter>...</Submitter>
  <Note>...</Note>
  <Citation>...</Citation>
  <MultiMedia>...</MultiMedia>
  <Changed>...</Changed>
</IndividualRec>

<IndividualRec Id="IN003">
  ...
</IndividualRec>

```

**<!-- Event Records -->**

```

<EventRec Id="EV001" Type="marriage" VitalType="marriage">
  <Participant>
    <Link Target="IndividualRec" Ref="IN001"/>
    <Role>husband</Role>
  </Participant>
  <Participant>

```

```

    <Link Target="IndividualRec" Ref="IN002"/>
    <Role>wife</Role>
</Participant>
<Date Calendar="Herrodicus">
    abt <Day>7</Day><Month>NOV </Month><Year>1024</Year>
</Date>
<Place>
    <PlacePart Type="Town" Level="4">Cove</PlacePart>
    ,
    <PlacePart Type="County" Level="3">Cache</PlacePart>
    ,
    <PlacePart Type="State" Level="2">Utah</PlacePart>
    ,
    <PlacePart Type="Country" Level="1">USA</PlacePart>
<!--
    This would print as Cove, Cache, Utah, USA. In the data stream, each comma is followed by a
    blank, which can't be seen here. The line breaks which we have used for clarity are not in the
    actual data stream. This could, also, be formatted as a complete address.
-->
<Coordinates>18.153N 178.150E</Coordinates>
<PlaceVariation TranslationType="phonetic" Method="kana">...</PlaceVariation>
<PlaceVariation TranslationType="romanized" Method="romanji">...</PlaceVariation>
</Place>
<Religion>Reformed Christian</Religion>
<ExternalID Type="..." Id="...">
<Submitter>...</Submitter>
<Note>...</Note>
<Citation>...</Citation>
<MultiMedia>...</MultiMedia>
<Changed>...</Changed>
</EventRec>

<EventRec Id="EV002" Type="christening" VitalType="aftbirth">
<!-- VitalType of "aftbirth" indicates an event that is accepted as proof of birth, and normally
happens shortly after birth -->
<Participant>
    <Link Target="IndividualRec" Ref="IN001"/>
    <Role>father</Role>
</Participant>
<Participant>
    <Link Target="IndividualRec" Ref="IN002"/>
    <Role>mother</Role>
</Participant>
<Participant>
    <Link Target="IndividualRec" Ref="IN003"/>
    <Role>child</Role>
</Participant>
...
</EventRec>

```

<!-- LDS Ordinance Records -->

```
<LDSOrdRec Id="OR001" Type="SP">
  <FamParticipant>
    <Link Target="FamilyRec" Ref="FM001"/>
    <HusbFathLiving>N</HusbFathLiving>
    <WifeMothLiving>N</WifeMothLiving>
  </FamParticipant>
  <IndivParticipant>
    <Link Target="IndividualRec" Ref="IN003"/>
    <Living>N</Living>
  </IndivParticipant>
  <OrdStat Code="submitted">
    <Date>...</Date>
  </OrdStat>
  <OrdStat Code="cleared">
    <Date>...</Date>
  </OrdStat>
  <OrdStat Code="completed">
    <Date>
      <Day>7</Day> <Month>APR</Month> <Year>1982</Year>
    </Date>
  </OrdStat>
  <!--
  Other examples:
  <OrdStat Code="dns"/>    Do not seal
  <OrdStat Code="bic"/>   Born in the covenant
  -->
  <TempleCode>CHICA</TempleCode>
  <Place>...</Place>
  <BasedOn>
    <Event>...</Event>
    <Note>...</Note>
  </BasedOn>
  <ExternalID Type="..." Id="...">
  <Submitter>...</Submitter>
  <Note>...</Note>
  <Citation>...</Citation>
  <MultiMedia>...</MultiMedia>
  <Changed>...</Changed>
</LDSOrdRec>
```

<!-- Contact Records -->

```
<ContactRec Id="CN001" Type="business">
  <Name>Genealogical Technology Corp.</Name>
  <MailAddress>
    <OrganizationName>GTC</OrganizationName><html:BR/>
    Bldg 10, mail stop 2A<html:BR/>
  </MailAddress>
</ContactRec>
```

```

    345 S. State Street<html:BR/>
    <PlacePart Level="4" Type="City">Salt Lake City</PlacePart>
    ,
    <PlacePart Level="2" Type="State">UT</PlacePart><html:BR/>
    <PlacePart Level="1" Type="Country">USA</PlacePart><html:BR/>
    <PlacePart Level="5" Type="Postal Code">84050-3400 </PlacePart>
</MailAddress>
<!--
    Because of the html:BR tags this would print as:
    GTC
    Bldg 10, mail stop 2A
    345 S. State Street
    Salt Lake City, UT
    USA
    84050-3400
-->
</ContactRec>

<ContactRec Id="CN002" Type="person">
  <Name>Joseph Tate</Name>
  <MailAddress>
    <Addressee>Mr. J. T. Tate</Addressee>
    1025 Pleasant Dr.<html:BR/>
    <PlacePart Level="4" Type="Town">Smithtown</PlacePart>
    ,
    <PlacePart Level="2" Type="State">CA</PlacePart><html:BR/>
    <PlacePart Level="1" Type="Country">USA</PlacePart><html:BR/>
    <PlacePart Level="5" Type="Postal Code">93502</PlacePart>
  </MailAddress>
  <Phone Type="home">940 345-6789</Phone>
  <Email>joeqpublic@mycorp.com</Email>
  <URI>http://www.mypage.org</URI>
  <Note>. . .</Note>
  <Changed>. . .</Changed>
  <SameIndiv>      <!-- This links a contact with his Individual record, if there is one -->
    <Link Target=" IndividualRec" Ref=". . ."/>
  </SameIndiv>
</ContactRec>

```

<!-- Source Records -->

```

<SourceRec Id="SR001" Type="vital records">
  <Repository>
    <Link Target="RepositoryRec" Ref="RP001"/>
    <CallNbr>4568-09</CallNbr>
  </Repository>
  <Title>Marriage Registry</Title>
  <Article>Used for articles in periodicals, etc.</Article>
  <Author>California State Board of Health</Author>

```

```
<Publishing>...</Publishing>
<Note>...</Note>
<Changed>...</Changed>
</SourceRec>
```

**<!-- Repository Records -->**

```
<RepositoryRec Id="RP001" Type="government">
  <Name>California Department of Vital Records</Name>
  <MailAddress>P. O. Box 2406 ... </MailAddress>
  <Phone>...</Phone>
  <Email>...</Email>
  <URI>...</URI>
  <Note>...</Note>
  <Changed>...</Changed>
</RepositoryRec>
```

**<!-- Group Records -->**

```
<GroupRec Id="GR001" Type="household">
  <Name>Smith Home</Name>
  <Contact>
    <Link Target="ContactRec" Ref="...">
  </Contact>
  <Member>
    <Link Target="IndividualRec" Ref="IN001"/>
    <Role>head</Role>
  </Member>
  <Member>
    <Link Target="IndividualRec" Ref="IN002"/>
    <Role>spouse</Role>
  </Member>
  <Member>
    <Link Target="IndividualRec" Ref="IN003"/>
    <Role>son</Role>
  </Member>
  <Member>
    <Link Target="IndividualRec" Ref="...">
    <Role>nanny</Role>
  </Member>
  <ParentGroup>
    <Link Target="GroupRec" Ref="..."> <!-- Neighborhood, for example -->
  </ParentGroup>
  <MailAddress>...</MailAddress>
  <Phone>...</Phone>
  <Email>...</Email>
  <URI>...</URI>
  <ExternalID Type="..." Id="...">
  <Submitter>...</Submitter>
```

```
<Note>...</Note>
<Citation>...</Citation>
<MultiMedia>...</MultiMedia>
<Changed>...</Changed>
<!--
    The Group record can be used for any type of group. It could be used for unconventional familial
    groups such as tribes, communes, etc., which are not the basis for sealing ordinances.
-->
</GroupRec>
```

**<!-- Multimedia Records -->**

```
<MultiMediaRec Id="MM001" Type="video" xml:lang="de">
  <Title>Kunzle Family Reunion</Title>
  <Description>...</Description>
  <Author>...</Author>
  <URI>http://www.kunzlefam.org/famreun.mov </URI>
  <Format>mov</Format>
  <Publishing>...</Publishing>
  <Note>...</Note>
  <Changed>...</Changed>
</MultiMediaRec>

</GEDCOM>
```

## Chapter 2

# GEDCOM XML Definitions

### Introduction

This chapter contains the specification for GEDCOM XML in the Document Type Definition (DTD) format, and in XML Data Schema format.

In traditional GEDCOM the order of data was not considered important. For example, in a family, information about parents could come first, or children information may be first. We didn't care. XML definitions by nature do specify and enforce the order of data. Order can be relaxed, but it requires extra specification, makes the definition more confusing, and loses precise cardinality specification. For example, in a family you can specify that parents and children can be specified in any order, but then you cannot specify that there can only be one father and mother, but any number of children. The document definitions below specifies and enforces the order of GEDCOM tags.

### GEDCOM XML Document Type Definition (DTD)

```
<!-- GEDCOM Root -->
```

```
<!-- The first level of elements contained in the GEDCOM element will be referred to as "records", i.e. a Family record, Individual record, etc -->
```

```
<!ELEMENT GEDCOM  
  (HeaderRec,  
   FamilyRec*,  
   IndividualRec+,  
   EventRec*,  
   LDSOrdRec*,  
   ContactRec*,  
   SourceRec*,  
   RepositoryRec*,  
   GroupRec*,  
   MultiMediaRec*)>
```

```
<!-- Elements Common to Multiple Records -->
```

```
<!ENTITY % RecordCom  
  "ExternalID*,  
   Submitter?,  
   Note*,  
   Citation*,  
   MultiMedia*,  
   Changed*">
```

**<!-- Header Record -->**

```
<!ELEMENT HeaderRec
  (FileCreation,
  Citation?,
  Submitter,
  Note?)>
```

<!-- The Citation is useful when all the data in the file comes from a single source -->

**<!-- Family Record -->**

```
<!ELEMENT FamilyRec
  (HusbFath?,
  WifeMoth?,
  Child*,
  BasedOn?,
  %RecordCom;)>
```

```
<!ATTLIST FamilyRec
  Id ID #REQUIRED>
```

**<!-- Individual Record -->**

```
<!ELEMENT IndividualRec
  (IndivName*,
  Gender?,
  DeathStatus?,
  AssocIndiv*,
  DupIndiv*,
  %RecordCom;)>
```

```
<!ATTLIST IndividualRec
  Id ID #REQUIRED>
```

**<!-- Event Record -->**

```
<!ELEMENT EventRec
  (Participant+,
  Date?,
  Place?,
  Religion?,
  %RecordCom;)>
```

```
<!ATTLIST EventRec
  Id ID #REQUIRED
  Type CDATA #REQUIRED
  VitalType
    (marriage | befmarriage | aftmarriage |
    birth | befbirth | aftbirth |
    death | befdeath | aftdeath) #IMPLIED>
```



**<!-- LDSOrdinance Record -->**

```
<!ELEMENT LDSOrdRec
  (FamParticipant?,
  IndivParticipant?,
  OrdStat*,
  TempleCode?,
  Date?,
  Place?,
  BasedOn?,
  %RecordCom;)>
```

```
<!ATTLIST LDSOrdRec
  Id ID #REQUIRED
  Type (B | C | E | SM | SP | SS | TO | W) #REQUIRED>
```

**<!-- Contact Record -->**

```
<!ELEMENT ContactRec
  (Name,
  MailAddress*,
  Phone*,
  Email*,
  URI*,
  Note*,
  Changed*,
  SameIndiv*)>
```

```
<!ATTLIST ContactRec
  Id ID #REQUIRED
  Type CDATA #IMPLIED>
```

**<!-- Source Record -->**

```
<!ELEMENT SourceRec
  (Repository*,
  Title,
  Article?,
  Author?,
  Publishing?,
  Note*,
  Changed*)>
```

```
<!ATTLIST SourceRec
  Id ID #REQUIRED
  Type CDATA #IMPLIED>
```

**<!-- Repository Record -->**

```
<!ELEMENT RepositoryRec
  (Name,
  MailAddress*,
  Phone*,
  Email*,
  URI*,
  Note*,
  Changed*)>
```

```
<!ATTLIST RepositoryRec
  Id ID #REQUIRED
  Type CDATA #IMPLIED>
```

**<!-- Group Record -->**

```
<!ELEMENT GroupRec
  (Name,
  Contact*,
  Member*,
  ParentGroup*,
  MailAddress*,
  Phone*,
  Email*,
  URI*,
  %RecordCom;)>
```

```
<!ATTLIST GroupRec
  Id ID #REQUIRED
  Type CDATA #IMPLIED>
```

**<!-- MultiMedia Record -->**

```
<!ELEMENT MultiMediaRec
  (Title,
  Description?,
  Author?,
  URI,
  Format,
  Publishing?,
  Note*,
  Changed*)>
```

```
<!ATTLIST MultiMediaRec
  Id ID #REQUIRED
  Type CDATA #IMPLIED>
```

**<!-- Sub Record Elements in alphabetical order-->**

```
<!ELEMENT Addressee (#PCDATA)>
```

<!ELEMENT Article (#PCDATA)>

<!ELEMENT ArticleDoc (#PCDATA)>

<!ELEMENT Association (#PCDATA)>

<!ELEMENT AssocIndiv  
 (Link,  
 Association,  
 Note\*,  
 Citation\*)>

<!ELEMENT Author (#PCDATA)>

<!ELEMENT AuthorDoc (#PCDATA)>

<!ELEMENT BasedOn  
 (Event\*,  
 Note\*)>

<!ELEMENT BirthDateDoc (#PCDATA) >

<!ELEMENT BirthPlaceDoc (#PCDATA) >

<!ELEMENT CallNbr (#PCDATA)>

<!ELEMENT Caption (#PCDATA) >

<!ELEMENT Changed  
 (Contact?,  
 Note?\*)>

<!ATTLIST Changed  
 Date CDATA #REQUIRED  
 Time CDATA #REQUIRED>

<!ELEMENT Child  
 (Link,  
 ChildNbr?,  
 RelToFath?,  
 RelToMoth?,  
 BasedOn?\*)>

<!ELEMENT ChildNbr (#PCDATA)>

<!ELEMENT Citation  
    (Source,  
    WhereInSource?,  
    WhenRecorded?,  
    References\*,  
    CitationText\*,  
    Extract\*,  
    Note\*)>

<!ELEMENT CitationText  
    (#PCDATA |  
    IndivDoc |  
    AuthorDoc |  
    TitleDoc |  
    ArticleDoc |  
    PublishingDoc |  
    WhereInDoc |  
    WhenRecDoc |  
    ReferencesDoc)\*>

<!ELEMENT Contact  
    (Link)>

<!ELEMENT Coordinates (#PCDATA)>

<!ELEMENT Copyright (#PCDATA) >

<!ELEMENT Date  
    (#PCDATA |  
    Month |  
    Day |  
    Year)\*>

<!ATTLIST Date  
    Calendar CDATA #IMPLIED>

<!ELEMENT Day (#PCDATA)>

<!ELEMENT DeathDateDoc (#PCDATA) >

<!ELEMENT DeathPlaceDoc (#PCDATA) >

<!ELEMENT DeathStatus (#PCDATA)>

<!ELEMENT Description (#PCDATA)>

<!ELEMENT DupIndiv  
    (Link,  
    Note\*,  
    Citation\*)>

<!ELEMENT Email (#PCDATA)>

<!ELEMENT Event  
(Link)>

<!ELEMENT ExternalID EMPTY>

<!ATTLIST ExternalID  
Type CDATA #REQUIRED  
Id CDATA #REQUIRED>

<!ELEMENT Extract  
(#PCDATA |  
IndivDoc)\*>

<!ELEMENT FamilyNbr (#PCDATA)>

<!ELEMENT FamParticipant  
(Link,  
HusbFathLiving?,  
WifeMothLiving?)>

<!ELEMENT HusbFathLiving (#PCDATA)>

<!ELEMENT FileCreation  
(Product?,  
Copyright?)>

<!ATTLIST FileCreation  
Date CDATA #REQUIRED  
Time CDATA #IMPLIED>

<!ELEMENT Format (#PCDATA)>

<!ELEMENT Gender (#PCDATA)>

<!ELEMENT GivenName (#PCDATA)>

<!ELEMENT GivenNameDoc (#PCDATA) >

<!ELEMENT HusbFath  
(Link,  
FamilyNbr?)>

```
<!ELEMENT IndivDoc
  (#PCDATA |
  NameDoc |
  BirthDateDoc |
  BirthPlaceDoc |
  DeathDateDoc |
  DeathPlaceDoc |
  MarrDateDoc |
  MarrPlaceDoc |
  ResidenceDoc |
  RelativeDoc)*>
```

```
<!ELEMENT IndivName
  (#PCDATA |
  GivenName |
  SurName |
  MaidenName |
  PersonalTitle |
  NameSuffix |
  IndNameVariation)*>
```

```
<!ATTLIST IndivName
  Type CDATA #IMPLIED
  TranslationType CDATA #IMPLIED
  Method CDATA #IMPLIED >
```

```
<!ELEMENT IndNameVariation
  (#PCDATA |
  GivenName |
  SurName |
  MaidenName |
  PersonalTitle |
  NameSuffix)*>
```

```
<!ATTLIST IndNameVariation
  TranslationType CDATA #IMPLIED
  Method CDATA #IMPLIED>
```

```
<!ELEMENT IndivParticipant
  (Link,
  Living?)>
```

```
<!ELEMENT Link EMPTY>
```

```
<!ATTLIST Link
  Target CDATA #REQUIRED
  Ref IDREF #REQUIRED>
```

```
<!ELEMENT Living (#PCDATA)>
```

```
<!ELEMENT MaidenName (#PCDATA)>
```

```

<!ELEMENT MailAddress
  (#PCDATA |
  Addressee |
  OrganizationName |
  PlacePart)*>

<!ELEMENT MarrDateDoc (#PCDATA) >

<!ELEMENT MarrPlaceDoc (#PCDATA) >

<!ELEMENT Member
  (Link,
  Role?)*>

<!ELEMENT Month (#PCDATA)>

<!ELEMENT MultiMedia
  (Link,
  Caption?,
  WhereInMM*,
  Note*)>

<!ELEMENT Name (#PCDATA)>

<!ATTLIST Name
  Type CDATA #IMPLIED>

<!ELEMENT NameDoc
  (#PCDATA |
  SurNameDoc |
  GivenNameDoc)*>

<!ELEMENT NameSuffix (#PCDATA)>

<!ELEMENT Note
  (#PCDATA |
  IndivDoc)*>

<!ELEMENT OrdStat
  (Date?)*>

<!ATTLIST OrdStat
  Code (canceled | cleared | completed | dns | dns-can | pre-1970 |
  qualified | submitted | uncleared) #REQUIRED>

<!ELEMENT OrganizationName (#PCDATA)>

<!ELEMENT ParentGroup
  (Link)*>

```



<!ELEMENT Participant  
    (Link,  
    Role?)>

<!ELEMENT PersonalTitle (#PCDATA)>

<!ELEMENT Phone (#PCDATA)>

<!ATTLIST Phone  
    Type CDATA #IMPLIED>

<!ELEMENT Place  
    (#PCDATA |  
    PlacePart |  
    Coordinates |  
    PlaceVariation)\*>

<!ATTLIST Place  
    TranslationType CDATA #IMPLIED  
    Method CDATA #IMPLIED>

<!ELEMENT PlacePart (#PCDATA)>

<!ATTLIST PlacePart  
    Type CDATA #IMPLIED  
    Level ( 1 | 2 | 3 | 4 | 5 | 6 | 7 ) #IMPLIED>

<!ELEMENT PlaceVariation  
    (#PCDATA |  
    PlacePart)\*>

<!ATTLIST PlaceVariation  
    TranslationType CDATA #IMPLIED  
    Method CDATA #IMPLIED>

<!ELEMENT PublishingDoc (#PCDATA)>

<!ELEMENT Product  
    (ProductId?,  
    Version,  
    Name,  
    Supplier?)>

<!ELEMENT ProductId (#PCDATA)>

<!ELEMENT Publishing (#PCDATA)>

<!ELEMENT References  
    (Citation+)>

<!ELEMENT ReferencesDoc (#PCDATA)>

<!ELEMENT Relationship (#PCDATA) >

<!ELEMENT RelativeDoc  
(#PCDATA |  
Relationship |  
NameDoc)\*>

<!ELEMENT Religion (#PCDATA)>

<!ELEMENT RelToFath (#PCDATA) >

<!ELEMENT RelToMoth (#PCDATA) >

<!ELEMENT Repository  
(Link,  
CallNbr?)>

<!ELEMENT ResidenceDoc (#PCDATA) >

<!ELEMENT Role (#PCDATA)>

<!ELEMENT SameIndiv  
(Link,  
Note\*)>

<!ELEMENT Source  
(Link)>

<!ELEMENT Start (#PCDATA) >

<!ELEMENT Stop (#PCDATA) >

<!ELEMENT Submitter  
(Link)>

<!ELEMENT Supplier  
(Link)>

<!ELEMENT SurName (#PCDATA)>

<!ELEMENT SurNameDoc (#PCDATA) >

<!ELEMENT TempleCode (#PCDATA)>

<!ELEMENT Title (#PCDATA)>

<!ELEMENT TitleDoc (#PCDATA)>

<!ELEMENT Units (#PCDATA) >

<!ELEMENT URI (#PCDATA)>

<!ELEMENT Version (#PCDATA)>

<!ELEMENT WhenRecDoc (#PCDATA)>

<!ELEMENT WhenRecorded (#PCDATA)>

<!ELEMENT WhereInDoc (#PCDATA)>

<!ELEMENT WhereInMM  
(Start,  
Stop?,  
Units)>

<!ELEMENT WhereInSource (#PCDATA)>

<!ELEMENT WifeMoth  
(Link,  
FamilyNbr?)>

<!ELEMENT WifeMothLiving (#PCDATA)>

<!ELEMENT Year (#PCDATA)>