

# Comments on the GEDC Genealogical Data Model

Tom Wetmore – September 23, 2011

## Introduction

The GEDC genealogical data model is defined in terms of an XML schema. As seems now conventional with other XML schemas for large sets of data, the top-level elements in GEDC XML strings and files correspond to independent records, and lower level elements represent the substructures found in records.

The GEDC model is a modern genealogical data model that has most of the features that I believe are necessary for the next generation of genealogical applications. Some of these features are:

- Support for the evidence/record and conclusion/person models – however, though the main objects from the conclusion realm are represented as records in their own right, the evidence objects are represented as sub-structures within conclusion objects.
- Support for places as separate, hierarchical entities.
- Support for the wide range of complexity required for dates and personal names in genealogical data.
- Support for sources as hierarchical entities.

The GEDC model has ten record types. The seven mentioned here are:

- Person – Conclusion level person.
- Event – Conclusion level multi-role event.
- Family – Conclusion level family.
- Source – Source of evidence; hierarchical.
- Place – Geopolitical place; hierarchical.
- Association – Conclusion relationship between two persons.
- Repository – Location where sources are found.

## Person Records for Conclusion Persons

A Person record represents a conclusion person. A Person record contains name and sex values, a few possible id values (Id, GUID, externalId), a few special properties (personStatus, locked, private, marked), other attributes, notes, references to Source records and change indicators. The only record type that a Person record can refer to is a Source record, which it does with a SourceRef value. Note specifically that a Person record does not refer to any Events that the person plays a role (is an Actor) in (or to any other structure, e.g., Family or Group, that it may be a member of). As will be described in the next section, Event records refer to Person records through Actor links.

## Event Records and Actor Structure Links for Conclusion Events

On the surface an Event record represents a multi-role conclusion event that has been extracted from one or more items of evidence collected by the genealogist (more on evidence aspect in the discussion of Accounts and Mentions). An Event record may contain any number of Actor structures; each Actor is a “conclusion” link from an Event record to a Person record that specifies the role the Person has in the Event. The link may carry other properties such as age and marital status. These “link-based” properties are important and are a hallmark, in my opinion, of genealogical data models that properly deal with the relationships between Persons and Events; that is, there are attributes of Persons that are true only in the context of an Event, and these links provide the correct context for recording this information. In other models the Actor object is called a Role. The other records that Event records may refer to are Place records, indicating the places where the Event occurred, and Source records.

## Accounts and Mentions Hold Evidence

Account structures are found only in Event records. Each Account structure holds information extracted directly from a physical (or in this modern era, a digital representation of a physical) Source item that holds evidence that the genealogist believes documents the Event. An Event record may hold any number of Accounts, one for each item of evidence discovered that pertains to the Event. An Account may have a DateValue, may have a PlaceRef that references a Place record, and should have a single SourceRef that refers to the Source record that describes the source of the extracted information. The Account structures found in the Event records provide the evidence about the event, while the values in the main part of the Event record hold the researcher's conclusions about the Event. Since there may be some inconsistencies between the information found in different Accounts, these inconsistencies are resolved in the conclusion part of the Event record.

Each Account may contain any number of Mention structures. Each Mention structure holds the information extracted about a single person mentioned in the evidence. This information includes the person's name, and may include the person's marriage status, age and role (string describing the role). The Mention structure is analogous to the Persona record of New Family Search.

Of concern is the fact that Accounts and Mentions, which are the evidence level equivalents of Events and Persons, are not record-level objects in their own right, but are subsumed in the bodies of the records that they purport to provide the evidence for. One very awkward outcome of this design is the tenuous connection that exists between the Person conclusion records and the Mention structures that provide their actual evidence. There is no clear linking between a Person record and the Mentions it is based upon, only a circuitous route through the Event record using a sequencing number scheme that first binds a Mention to an Actor in the Event, which is then bound to a Person.

In my opinion placing Mentions inside Accounts and placing Accounts inside Events is one of the only major problems with the GEDC model, a problem that could be easily resolved by promoting the two structures into their own records. The Account then becomes an evidence level Event record, and the Mention becomes an evidence level Person record, the kind of record that is now popularly called a Persona. One easy way to see the need for this change is to imagine the stability of both Accounts within Events and Mentions within Accounts. As research progresses there can be many situations in which a researcher discovers he/she has made mistakes about what evidence applies to which events and persons, necessitating wholesale movements of sub-structures of some Event records into the bodies of other Event records.

## Family and Association Records

GEDC has conclusion records for Associations (two-way relationships between Persons), and Families (collections of Persons possibly including a husband/father role player, a wife/mother role player, and children role players). Each type of record may refer to Source records to provide their provenance, and Family records may refer to the Event records that hold the family level events of the Family. Nothing unconventional here. There are no evidence level equivalents of these records.

## Source and Repository Records

GEDC has conventional Source and Repository records. And the Source records are hierarchical. A good solution.

## Place Records and a Nice Linking Feature

In the GEDC model places are generally handled by a hierarchy of Place records. Each Place record may refer to a parent Place record that contains it. I believe this is the proper way to handle Places. Many modern genealogical applications take this approach and provide a "Name Expert" component that is pre-configured with a huge number of Place records.

GEDC has an interesting twist on the link concept, a twist that I also use to some advantage in the DeadEnds model. In the GEDC context a Link is a sub-structure in one record that allows that record to refer to another record. In most contexts a Link can be "wrapped" with other attributes that can

provide a context for the referring object within the referred object. But that is not the wrinkle I am concerned with here.

In GEDC a Link, instead of actually referring to another record, may instead include a string representation of that other object. For example, if some Event occurred at a place that might not fit cleanly into a Place hierarchy, the genealogist could choose not to actually create a new Place record for that place, but to simply refer to it by name in the Link. For example, instead of:

```
<link class="Place" uuid="abcdefabcdef"/>
```

where "abcdefabcdef" is the record id of a Place record, the link could be encoded without reference to another record as:

```
<link class="Place" text="somewhere in the Himalaya Mountains"/>
```

This isn't exactly the way these elements would be done in GEDC but it captures the concept. Obviously, the link in this second approach does not refer to an external record for its data, but hold its data directly.