Proposal to Adopt Certain Core Record Linkages

Authors	Tony Proctor
Created	2013-03-14
Title	Record_Linkage
Version	1.0
Type	Technical
Language	English
Description	Proposal for the major record linkages in the data model
Suggested Keywords	Record-linkage, Events, Persons, Sources, Personae

Contents

1. Abstract

This proposal suggests a way for structuring the fundamental linkages in the Data Model that will accommodate family history in addition to genealogy. These linkages are the ones that will relate persons, places, sources, and evidence.

2. Proposal

Family history, and probably history in general, requires support for events. Events are core concepts that link a place to a date (or even to a date range). Multiple people can be linked to a shared event and distinguished by their 'role', such as the groom, the bride, a witness, the bride's mother, etc.

Traditional genealogy (i.e. family trees and pedigree charts) has little use for events. It has dates for vital events such as birth, marriage, & death, but these are mostly represented as simple properties.

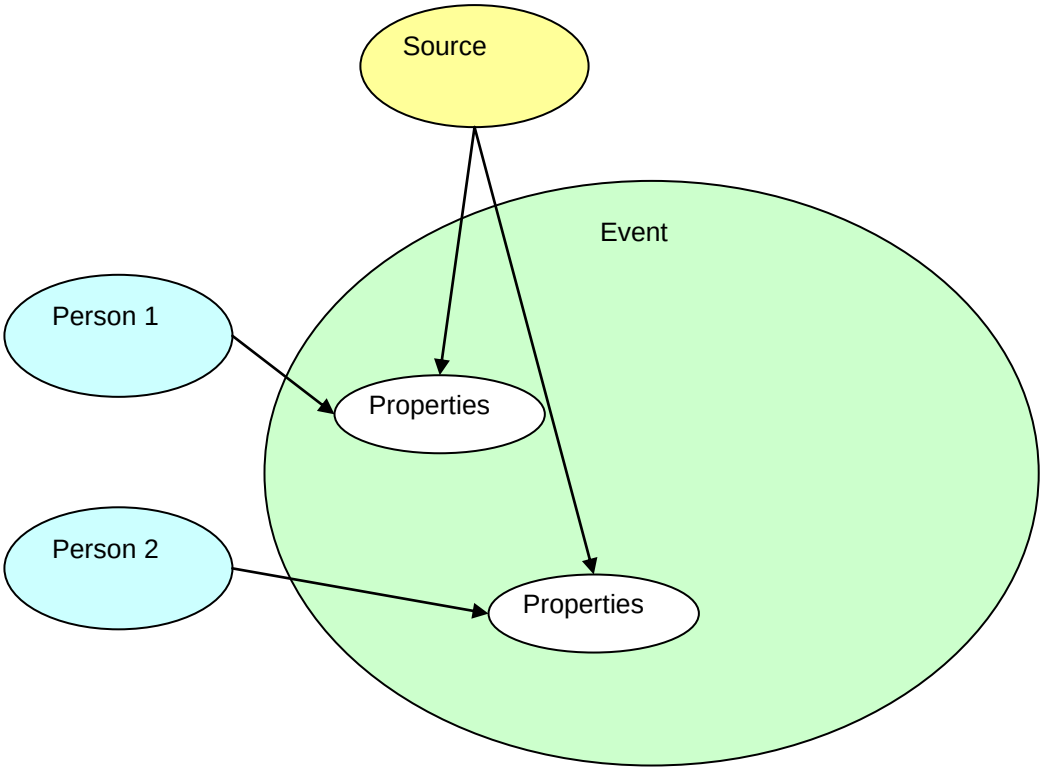
The more limited scope of traditional genealogy makes it difficult to associate evidence correctly. For instance, if there are multiple, non-agreeing records implying a date-of-birth then what are they attached to? Do they imply multiple dates of birth have to be recorded?

In event-based family history, are sources connected to events, or to people? What happens when the same source for an event contributes distinct properties for several people? What happens when the event is supported by multiple sources, each contributing distinct properties for each person?

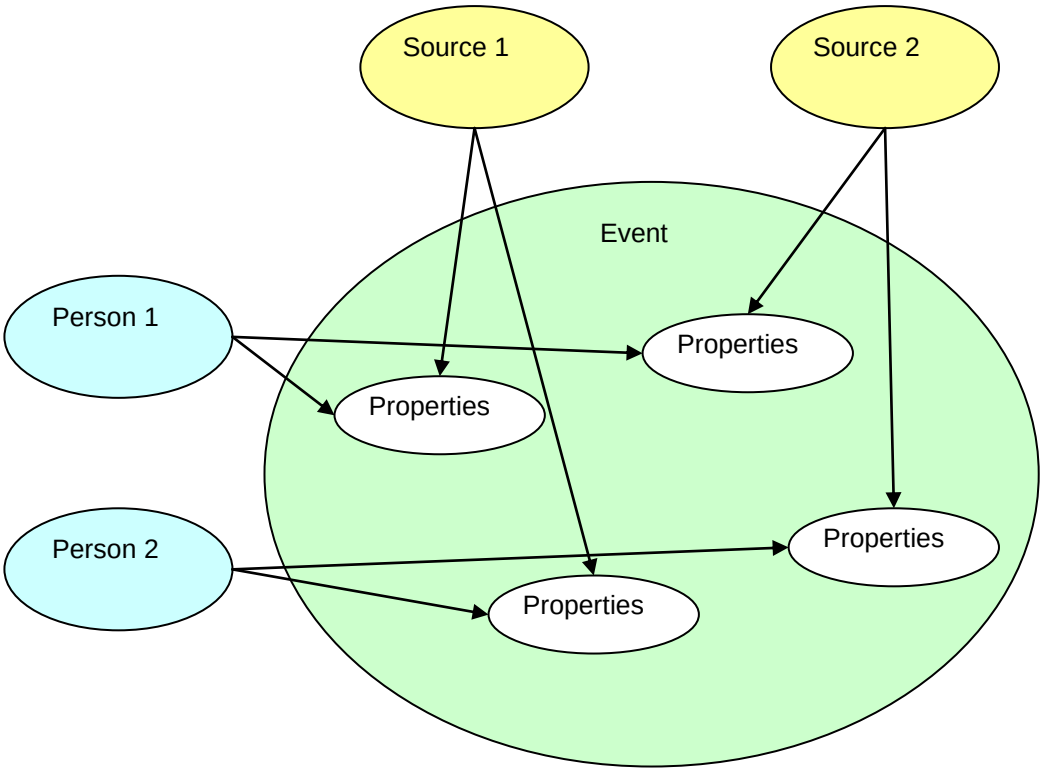
It is the contention of this proposal that the vast majority of sources support an event rather than just people or places. The people and places are associated with the event, and the person properties (e.g. age, occupation, name, etc) are associated with the person-to-event link rather than either the person or the event.

The following diagram shows a shared event with two people associated with it. The event is supported by a single source which yields distinct properties for each of the persons.

Note that a group of such properties is basically what is described as a 'persona' in some models. That is, evidence of some person yielded by one single source. These groups of properties also exist in the STEMMA model but are not described as personae because they do not have their own IDs and cannot be manipulated separately.



The next diagram extends this scenario to an event supported by two different sources, each of which yields distinct properties for the two persons.



Whereas the properties were associated with the person-to-event link in the first diagram, they are associated with the person-to-subevent in the second diagram (where 'subevent' is the part specifically supported by a given source).

3. Not Covered or Not Required

This proposal identifies where *personae* fit in to the linkages. As a group of properties, these are a fundamental part of this proposal. However, to be truly accepted as personae, they would have to be distinct entities with their own IDs. Although outside of this proposal, this extension is still in keeping with it. In summary, the merit of incorporating true personae into the model is to do with the inherent expressive power rather than any of the claimed procedural uses.

STEMMA treats places and persons in a very similar and streamlined fashion. This means that the case made here for person-properties and the person-to-event link also apply to places. The STEMMA philosophy here is that places are as equally important to history as are persons. This is likely to be a major difference from genealogical requirements. That generic treatment is not a part of this specific proposal.

4. Illustration

See STEMMA Examples in References section.

5. Use Cases

Consider a census return. This is a shared event and all the people in the same household would have a distinct role in that event. It should be obvious how this relates to the first of the diagrams above. The event is supported by common set of sources and so they're cited from the shared event, not from each of the associated persons.

The properties yielded by the census sources are represented in the person-to-event link. For instance:

```
<Person Key='pWilliamElliott'>

    <EventRef Key='eCensusElliott1851'>
        <Property Name='Name'> William Elliott </Property>
        <Property Name='Age'> 10 </Property>
        <Property Name='Occupation'> Scholar </Property>
        <Property Name='Role'> Son </Property>
    </EventRef>

</Person>
```

Hence, each associated person gets their own set of properties from the same event. However, note that these properties reflect evidence rather than fact or conclusion. This means they have to be recorded unchanged, and that a subsequent census event may have differing properties. The single-valued details recorded for a Person entity (e.g. date of birth) are therefore conclusions derived by consideration of the available evidence.

As a multi-source event, consider a marriage. As well as having the civil or religious registration, there may be announcements in the newspapers. Although these sources support the same event, they will yield different sets of properties for the people involved, and there may be some variation in the values.

STEMMA accomplishes this by referencing a division of the event supported by a particular source. For example:

```
<EventRef Key='eEventA'>
        <DetailRef Key='eEventA-Src1'>
                <Property Name='Age'>26</Property>
        </DetailRef>
        <DetailRef Key='eEventA-Src2'>
                <Property Name='Age'>27</Property>
        </DetailRef>
</EventRef>
```

6. References

STEMMA Examples. http://www.familyhistorydata.parallaxview.co/data-model. Sections 'Evidence and Timelines' and 'Multi-Source Events' (currently 4.1 and 4.2).