

CFPS 89

(Call for Papers Submission number 89)

Sources, events, personae, individuals and graphs

Submitted by: Cothey, Viv

Type: A comment on a submitted paper

Comment on: 72

Created: 2013-08-06

URL: Most recent version: <http://fhiso.org/files/cfp/cfps89.pdf>
This version: http://fhiso.org/files/cfp/cfps89_v1-0.pdf

Description: Some comments on several papers based on an examination of the informational relationships between sources, events, and biographical information. The paper identifies an evidence graph as the basis for an explicitly evidence based and culturally agnostic approach to genealogical information systems.

Keywords:

Sources, events, personae, individuals and graphs

Viv Cothey

6 August 2013

Abstract

This paper is to support some of the proposals already made but also to oppose others [4,21,22,24,25,34,63,70,72,73,78]. I advocate an explicitly evidence based and culturally agnostic approach to genealogical information systems. In particular I examine the informational relationships between sources, events and biographical (or nominal record) information. The paper is informed by graph theory and identifies a need for an *evidence graph*. Conclusional graphs can be derived together with presentational reports which may introduce user-dependent constructs and formats.

Introduction

Graphs (sometimes called *networks*) are used to represent collections of a pair of objects, one referred to as *nodes* or *vertices* and the other as *edges* or *arcs*. An edge can connect two nodes and represents a relationship between them. Both nodes and edges can have an associated set of properties depending on context. For example, in the *citation graph* each node represents a published article while an edge represents a citation from an article to the article being cited. Node properties here include the name of the author. White's P-Graph (White and Jorian, 1992) represents parental relations. Each node is either one or both parents or a child (not a parent). Edges link parents to their children.

Setting genealogical information systems as a particular instance within the general field of graphs provides opportunities to share tools providing graph visualisation and manipulation.

Sources and events

By sources I mean genealogical sources that contain some kind of biographical information and which mostly refer to a named individual, (note 1). This information provides a nominal record which usually includes a name together with a collection of attributes, characteristics, traits, relationships etc. Sources are documents (defined broadly to include film, broadcasts and oral history etc.). Sources can therefore always be defined by their proper bibliographic description. Referencing and citation formats and styles are a presentational issue which should not affect the bibliographic description (note 2), see also Proctor (2013c).

Many sources owe their existence to officialdom's predilection to document in writing events deemed to be of interest. For example taxation charges and payments, marriages, wills and census lists. Other sources quite literally refer to an event of note.

Although logically it might be considered that it is the event that is fundamental, it is also true that we can only know of an event when there is a document that provides us with evidence of its existence. From this point of view it is the source that is fundamental. Each event must have just one source.

A particular source may generate many different nominal records, call these personae, where biographically each persona is characterised as a participant in the event that the source records. Each persona must have just one source, that is a persona node connects to just one source node in what I am calling the *evidence graph*.

Suppose there is a second source that describes an event similar to that described by our first source and that we have two personae, one from the first source and one from the second. We may decide to conclude that the two personae comprise a single individual despite persona characteristics being inconsistent or incomplete. We should not conclude that the two events are identical since this would oblige us to merge the personae. The events, that is their existential evidence, are not identical since each relies on a different source. (Here it might be that one can develop a notion for events that is analogous to persona but I'm not sure there is any benefit.)

An empty source is one that has no (relevant) persona records. Identifying empty sources can be useful to a researcher.

Hence as regards *Proposal to adopt certain core record linkages* (Proctor, 2013d) while there is much to agree with, I disagree with linking persona characteristics to events; they should be linked to sources, and the notion of a single event being supported by two (or more) different sources should be avoided.

A more radical approach would be to dispense with “event” altogether although I accept it has become part of most genealogical information system. I disagree with *Proposal to accommodate structured events* (Proctor, 2013e). If anything like what is described is needed then better date structures etc. should be used. In *Nine necessities in a GEDCOM replacement* Louis Kessler (2013) moves in this direction. He also identifies issues with the notion of “family”, (see below).

Personae, individuals and places

Thomas Wetmore in *Persona records* (2013) makes the case for persona records. Here I am proposing “individual” to be a conclusional construct formed by an assemblage of personae records. Biographical information is aggregated and inconsistencies resolved. Any choices that are made should be explicit.

For example the value of the name property for an individual may well be chosen as a combination of persona names. The evolution and usage of an individual's name is explicit in the persona records. At some stage this will become a purely presentational issue, see also Proctor (2013a).

Edges between personae nodes describe relationships. In principle there is no restriction on what the relationship is but clearly some such as “parent” will be more genealogically important. A detailed review of relationships is not undertaken here but the case of “grandparent” is noted to illustrate thinking. Suppose we have a source recording persona-1 and persona-2 together with the relationship persona-2 being a grandparent of persona-1. Then in the evidence graph we have three persona nodes, 1, 2 and 3 (which may be anonymous) and two parent edges, 2-3 and 3-1, where 2 is parent of 3 etc. In this way more complex lineage relationships can be constructed.

Non-lineage relationships such as membership of a household or employee can be also be represented, see Johnston (2013a), (note 3).

The notion of a “family” especially if this entails “marriage” is a particular construction

that may be needed for presentational purposes. Since “sex” is a persona node property and “parent” is a relation edge in the evidence graph then “mother” and “father” can be concluded.

The evidence graph described here does not require a “place” node. Place can be a persona node property having a value which is the placename. However despite placenames, like personal names, being problematic in that how does one know that any two are intended to name the same entity, there is a strong preference in genealogical information systems for the creation of gazetteer like components. This is discussed by Johnston (2013b) while Tony Proctor (2013b) draws attention to some of the challenges that need to be faced. Place name references have long been a topic of study by the electronic text encoding community (TEI Consortium, 2013). Chapter 13 includes relevant guidance regarding placenames. Therefore other than noting that persona nodes could link to auxiliary “place” nodes if needed the further ramifications of this proposal are not pursued.

Evidence and conclusions

An evidence graph with additional individual nodes and individual relationship edges is a conclusion graph. Each individual node has edges to all the personae nodes which it incorporates. A persona node can link to only one individual node. Individual nodes inherit relationship edges to other individual nodes from their incorporated personae. Of course there is no guarantee that information will be consistent. Inconsistency may indicate either a conclusional error in incorporating a particular persona as a component for an individual or a record inconsistency. Hence there is a strict separation between evidence and conclusion see also Proctor (2013f). Alternative interpretations of the evidence personae may generate a different conclusion.

Figure 1 shows an artificial example of two sources, S1 and S2. Persona P1 is provided by S1 while personae P2 and P3 are provided by S2. S2 also provides the evidence for the relationship R1.

Figure 1 also shows two individuals In1 and In2 together with the inherited relation R2.

Hence if P1 is G Jones died 1920, P2 is George Jones, P3 is Fred, a son, born 1900 with R1 being P2 is the parent of P3, then we have the conclusional individuals In1, George Jones died 1920 and In2, Fred Jones born 1900 sex is male and the parent relation In1 is the parent of In2.

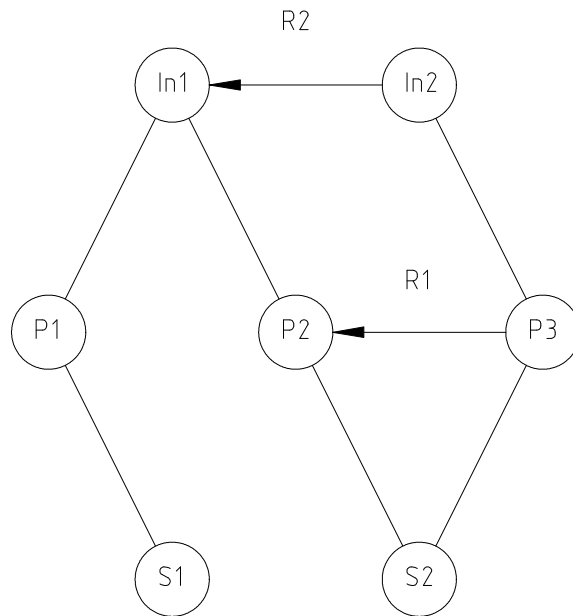


Figure 1: Evidence-conclusional graph

Tychonievitch (2013) argues for a rich commentary within the conclusion graph. This could be accommodated by the inclusion of, say, “research” nodes that connect to individual nodes.

Conclusion

The evidence graph contains both source nodes and personae nodes. Edges representing a citation/reference connect each persona to its source. Technically the evidence graph is bipartite.

The evidence graph can provide a culturally agnostic and format neutral basis for documenting and exchanging genealogical information. The next step would be to define more detail for nodal and edge properties.

Relationship edges between personae nodes represent the evidence based relationships. These will include a “parent” relationship from a persona child to its parent persona. There is no “family” node or “mother/father” relationship since “parent” is an edge property and “sex” is a node property. “Family” is a social specific construct that can be added by a developer if required.

A conclusion graph with individual nodes is derived from the evidence graph by aggregating personae nodes and their edges.

Presentational diagrams and reports with user relevant formats and styles are independent of the evidence graph.

Notes

1. “individuals” here could be fictional or non-human, for example race horses.
2. Both libraries and archives maintain an active interest in maintaining and developing national and international bibliographic standards. It is common practice for authors using bibliographic systems to generate different reference styles from the same bibliographic record according to a journal publisher’s specification.
3. TEI Consortium (2013) chapter 13 provides examples of encoding relationships in a way that is compatible with the evidence graph described here.

References

- Johnston W. (2013a). Beyond lineage-linked. *Family History Information Standards Organisation: call for papers submissions*, 73.
- Johnston W. (2013b). Places: functional requirements. *Family History Information Standards Organisation: call for papers submissions*, 70.
- Kessler, L. (2013). Nine necessities in a GEDCOM replacement. *Family History Information Standards Organisation: call for papers submissions*, 78.
- Proctor, T. (2013a). Proposal for handling personal names. *Family History Information Standards Organisation: call for papers submissions*, 21.
- Proctor, T. (2013b). Proposal for handling place references. *Family History Information Standards Organisation: call for papers submissions*, 22.
- Proctor, T. (2013c). Proposal for handling sources and citations. *Family History Information Standards Organisation: call for papers submissions*, 63.
- Proctor, T. (2013d). Proposal to adopt certain core record linkages. *Family History Information Standards Organisation: call for papers submissions*, 24.
- Proctor, T. (2013e). Proposal to adopt structured events. *Family History Information Standards Organisation: call for papers submissions*, 25.
- Proctor, T. (2013f). Requirement to separate evidence and conclusion. *Family History Information Standards Organisation: call for papers submissions*, 34.
- TEI Consortium, eds. (2013). *TEI P5: Guidelines for electronic text encoding and interchange*. 2.5.0. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (accessed 6 August 2013).
- Tychonievich, L. (2013). Modelling research not conclusions. *Family History Information Standards Organisation: call for papers submissions*, 4.
- Wetmore, T. (2013). Persona records. *Family History Information Standards Organisation: call for papers submissions*, 72.
- White, D.R and Jorian, P. (1992). Representing and analyzing kinship: a network approach. *Current Anthropology*, 33, 454-462.