

# A Story of Olde

## A case for structured narrative in genealogy

**Tony Proctor, 30 Jun 2013, edited 19 Apr 2017**  
Copyright © 2017 Tony Proctor. Licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

There are an increasing number of Web sites that try to encourage the use of narrative for our family history, typically for recollections and stories (e.g. [www.narrativepublishers.com](http://www.narrativepublishers.com)). So why is it that our current software products do not help?

In this article, I want to make a case for the proper integration of narrative into our core family history data. I will explain how this could be done, and the enormous advantages that it would unlock.

Much of the online data that we use emphasises discrete items of data such as names, dates, places, occupations, and ages. This is primarily because those items of data are used to index images of census pages, parish records, service records, etc. Where some narrative is present, such as in a service record or a will, it is usually left in the image rather than being transcribed. An example that bucks this trend is newspaper archives. Projects such as The British Newspaper Archive<sup>1</sup> provide an incredibly rich insight into the lives of our ancestors, sometimes mentioning specific individuals or families by name. However, not all of them allow you to copy-and-paste text from their transcriptions into your data, thus leaving you to be content with mere scans or to make your own transcriptions.

Most people reading this will be used to storing image files in their data. This is fine for such things as photographs but should we really be content with that where narrative is concerned? Elizabeth Shown Mills has suggested that genealogists should write narrative in a word-processor, although the context was actually an opposition to letting software try and generate it. With all the best intentions in the world, software cannot generate narrative from simple facts and figures; the result is robotic at best, and laughable at its worst.

Virtually all current software products accommodate text as a form of notes or annotation, but that's a long way from accommodating rich-text narrative. This may be because they are more focused on genealogy than family history,<sup>2</sup> but it is certainly the case that a focus on the use of a database means that rich-text narrative cannot be handled adequately.<sup>3</sup> Also, the GEDCOM<sup>4</sup> data format is acknowledged as being a representation of genealogical data, and it set in stone the notion of textual notes.

---

<sup>1</sup> The British Newspaper Archive. [www.britishnewspaperarchive.co.uk](http://www.britishnewspaperarchive.co.uk).

<sup>2</sup> The Society of Genealogists publishes a differentiation of genealogy and family history on their Web site at: <http://www.sog.org.uk/education/gandfh.shtml>.

<sup>3</sup> See "Do Genealogists Really Need a Database?", Blogger.com, *Parallax View*, 20 Oct 2013 (<http://parallax-viewpoint.blogspot.com/2013/10/do-genealogists-really-need-database.html>).

There are many additional uses for storing narrative text such as recording stories and family recollections, incorporating text from family letters, biographical text, history of a place, or explaining your rationale when reaching conclusions during your research. When dealing with family history, narrative isn't simply something that has to be generated in client reports, along the lines required for [BCG](#) certification; it is something that should be part of the core data that we store and work on.

So what is wrong with our data incorporating lots of separate word-processor files? I will now illustrate how that fails to truly integrate the associated text, and how it can be significantly improved upon. In order to make this illustration, I am using the example of a two-page family letter written by an 11-year-old girl, Tessie Mae Hamlett, to her father in Americus, a small town in Indiana, on 3<sup>rd</sup> January 1899. The images and transcription of the letter are reproduced here by kind permission of [forgottennewengland.com](#)<sup>5</sup>.

---

<sup>4</sup> GEDCOM (GEnealogical Data COMmunication) is a genealogical data format devised by The Church of Jesus Christ of Latter-day Saints (LDS Church) in 1984.

<sup>5</sup> <http://forgottennewengland.com/2011/12/15/a-window-into-the-past-ancestors-letters-as-genealogical-records/>.

Americus Ind

Jan 3 1899

Dear papa

I thought I would write to you this evening to let you know that we are as well as usably and hope this may find you the same. George and Ann was up to day ~~they~~ ~~they~~ ~~they~~ ~~they~~ Aunt susie is sick she has got a nice swelled on her hip they pulled her to day that little pig, it weighed 125 pounds when it was dressed.

Bessie got a letter from Bennie Phipps to day he got home all write. George said to tell you that he got a letter from worth judgment to day. Bessie is making her a apron and manson is fitting her shoes manson got mad at Bessie Saturday

but he wasn't mad long,  
We was at Kuskela's Saturday night  
we popped pop corn had lots of fun,  
Well papa when are you coming  
home.  
We have not heard from Jessie Emma  
since they went home they left  
here Friday morning about 6 o'clock in the  
morning and it was 10 o'clock when they  
got to go away.  
Well I guess this is all for this time  
so goodnight Summer soon good by  
I love May to papa.

**FORGOTTEN  
NEW ENGLAND**

Here is the plain-text edition of that transcription, including original grammatical errors and spelling mistakes. Although the paragraph breaks have been preserved, the individual line breaks have not — the text being “flowed” as in a modern document. At this point, a full diplomatic transcription does not add to the case being made.

*Dear papa:*

*I thought I would write to you this eveng to let you know that we are as well as usual and hope this may find you the same George and Ann was up to day. Aunt Susie is sick. she has got a ugly swelled on her hip they bullchurd to day that little pig. it wayed 125 pounds when it was dressed.*

*Bessie got a letter from Bennie Pepper to day he got home all write. George said to tell you that he got a letter from worth judgment to day. Bessie is making her a apron and manson is fixing her shoes manson got mad at Bessie saturday but he wasnt mad long.*

*We was at Kurkehains saturday night we poped pop corn had lots of fun. well papa when are you coming home.*

*We have not heard from Jessie Emmons since they went home they left here friday morning about 6 oclock in the morning and it was 10 oclock before they got to go away.*

*Well I guess this is all for this time so good night answer soon good by.*

*Tessie Mae to papa*

Note firstly that there are several person references here. If you wanted to search for them using a plain-text search, though, then you may not find them. Without knowing the context, how would you know what exact spelling to use, or what contracted form (e.g. “Bessie”), or that “Papa” was actually William Franklin Hamlett?

If the references are to family members then they should be connected to relevant persons in your core data so that clicking on the reference to will take you to that part of your tree for further information. If a person is not in your tree, e.g. “George and Ann”, then the reference can still be flagged but without connecting it to any person in your data.

There are also a number of date references. A clever bit of software might recognise the obvious date “3 Jan 1899” in the letterhead, but without the context how would it know that “Saturday” was 31 Dec 1898, or that “Friday” was 30 Dec 1898?

A similar argument can be made for place references, although this letter only includes an “Americus, Ind” reference in the header. The reference to “Kurkehains” was likely the household of someone of German origin with the surname Kuckein or Kurkine.

For those of you with a technical background, what is needed is some type of *meta-data* (i.e. ‘data about data’), and for narrative text this means we need a *mark-up language* where meta-data is placed inline with the readable text using a particular syntax. HTML (HyperText Markup Language) is a well-known example, but so are ordinary word-processor documents. It’s just that your word-processor software nicely hides that mark-up for you. What we need is a way to mark those references and add hidden information, in much the same way that a word-processor encodes

such detail as hyperlinks and bookmarks. The following diagram illustrates the effect of marking-up selected references — limited in number, here, to keep it simple — and of having that mark-up connect the references to their respective person, date, or other.

Dear papa:

I thought I would write to you this eveng to let you know that we are as well as usual and hope this may find you the same. George and Ann was up to day. Aunt Susie is sick. she has got a ugly swelled on her hip they bullchurd to day that little pig. it wayed 125 pounds when it was dressed.

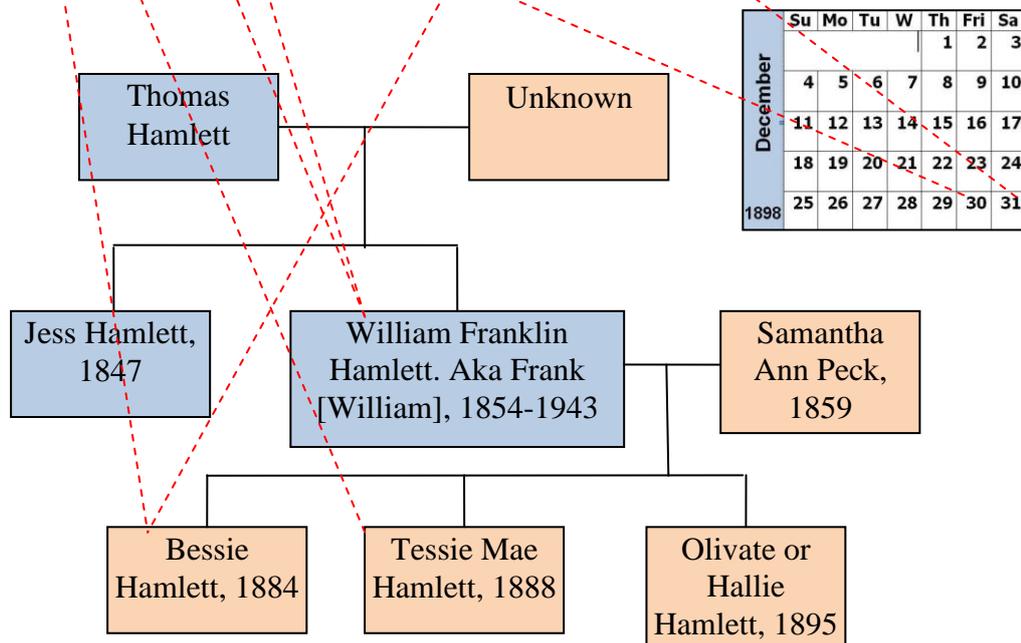
Bessie got a letter from Bennie Pepper to day he got home all write George said to tell you that he got a letter from worth judgment to day. Bessie is making her a apron and manson is fixing her shoes manson got mad at Bessie saturday but he wasnt mad long

We was at Kurkehains saturday night. We poped pop corn had lots of fun. Well papa when are you comming home.

We have not heard from Jessie Emmons since they went home they left here friday morning about 6 oclack in the morning and it was 10 oclock before they got to go away.

Well I guess this is all for this time so good night answer soon good by

Tessie Mae to papa



This was one of my dream features when, during 2010/2011, I began an R&D project to devise a way of representing family history data the way I wanted it. The first

edition of the resulting specification (STEMMA<sup>®</sup>, Source Text for Event and Ménage MApping), together with its research notes, were published on the Web in January 2012 at [www.parallaxview.co/familyhistorydata](http://www.parallaxview.co/familyhistorydata). I believe our tools would better support family history, and micro-history in general, if they provided something similar to STEMMA's *structured narrative*.

So let us examine how we might represent this in STEMMA. We will begin with the copy of the letter itself which would be described by a Resource entity:

```
<Resource Key='rLetter' Abstract='1'>
  <Type> Letter </Type>
  <Params>
    <Param Name='Page' Type='Integer' />
  </Params>
</Resource>

<Resource Key='rTessieMaeLetter'>
  <Title> Tessie Mae Hamlett's letter to her Papa </Title>
  <BaseResourceLnk Key='rLetter' />
  <URL> file:mydocuments/letters/TMHamlett-P{$Page}.jpg </URL>
  <DataControl>
    <Permission> By kind permission of Forgotten New England,
    http://forgottennewengland.com </Permission>
    <Text>
      Permission to use Tessie Mae's letter in the context of
      this article was provided by Ryan W. Owen on 30
      January 2013.
    </Text>
  </DataControl>

  <Text>
    ... to be completed below ...
  </Text>
</Resource>
```

This Resource entity describes both pages of the letter by virtue of the 'Page' parameter. Notice that we've left a space for a marked-up transcription of the letter. We'll return to this in a moment.

We now need a citation for this letter but we have two sources to cite: Tessie Mae herself, and Forgotten New England who hold the original and who supplied the scanned copies.

```
<Citation Key='cForgottenNewEngland'>
  <URI> http://stemma.parallaxview.co/source-type/blog </URI>
  <Params>
```

---

<sup>®</sup> STEMMA is a registered trademark of Tony Proctor.

```

    <Param Name='Title'> A Window into the Past: Ancestors'
    Letters as Genealogical Records </Param>
    <Param Name='Blog'> Forgotten New England </Param>
    <Param Name='URL'>
    http://forgottennewengland.com/2011/12/15/a-window-
    into-the-past-ancestors-letters-as-genealogical-records/
    </Param>
    <Param Name='Accessed' Type='Date'> 2013-01-30 </Param>
    <Param Name='Date' Type='Date'> 2011-12-15 </Param>
  </Params>
</Citation>

```

```

<Citation Key='cTessieMaeLetter'>
  <URI> http://stemma.parallaxview.co/source-type/correspondence
  </URI>
  <Params>
    <Param Name='From' Type='PersonRef' Key='pTMHamlett' />
    <Param Name='To' Type='PersonRef' Key='pWFHamlett' />
    <Param Name='Date' Type='Date'> 1899-01-03 </Param>
  </Params>
  <ParentCitationLnk Key='cForgottenNewEngland' />
</Citation>

```

```

<Source Key='sTessieMaeLetter'>
  <Frame>
    <CitationLnk Key='cTessieMaeLetter' />
    <ResourceLnk Key='rTessieMaeLetter' />
    <Credibility> Trusted </Credibility>
    <Reliability> Primary </Reliability>
    <Quality> Copy </Quality>
  </Frame>
</Source>

```

Here, we've created a citation chain (using ParentCitationLnk) since we're citing a copy of Tessie Mae's letter, and linking it to an associated scan, but we're also citing the source of the copy.

All we need now is to represent the event for the writing of the letter, and link it to the associated Citation entity and to a hierarchical Place entity for Americus.

```

<Person Key=' pTMHamlett'>
  <PersonName> Tessie Mae Hamlet </PersonName>
  ...
  <Eventlet>
    <Type> Social </Type>
    <SubType> Correspondence </SubType>
    <PlaceLnk Key='wAmericus' />
    <When Value=' 1899-01-03' />
    <SourceLnk Key='sTessieMaeLetter' />

```

```

    </Eventlet>
    ...
</Person>

<Place Key='wAmericus'>
  <Title> Americus </Title>
  <Type> Town </Type>
  <PlaceName> Americus </PlaceName>
  <ParentPlaceLnk Key='wIndiana' />
  <Text>
    Brief information on this town can be found at <Link
    URL='http://en.wikipedia.org/wiki/Americus,_Indiana'>Wiki
    - Americus</Link>
  </Text>
</Place>

<Place Key='wIndiana'>
  <Title> Indiana </Title>
  <Type> State </Type>
  <PlaceName> Indiana </PlaceName>
  <ParentPlaceLnk Key='wUS' />
  <Text>
    A potted history of Indiana state can be found at <Link
    URL='http://en.wikipedia.org/wiki/Indiana'>Wiki -
    Indiana</Link>
  </Text>
</Place>

<Place Key='wUS'>
  <Title> United States </Title>
  <Type> Country </Type>
  <PlaceName> US </PlaceName>
</Place>

```

However, a STEMMA Event doesn't just represent the people physically present but all those mentioned in its supporting source(s). This would include her relatives, and even those people who we couldn't identify or who don't exist in our family data. We would therefore replace the previous Eventlet that was embedded in a single Person entity with a standalone Event entity. That Event would identify each person or place referenced in the source and connect them to their corresponding entity, if one is defined for them.

```

<Event Key='eTessieMaeLetterWriting'>
  <Title> Tessie Mae's letter writing to her father </Title>
  <Type> Social </Type>
  <SubType> Correspondence </SubType>
  <PlaceLnk Key='wAmericus' />
  <When Value=' 1899-01-03' />
  <SourceLnk Key='sTessieMaeLetter'>

```

```

<PersonLnk Key='pTMHamlett'>
  <Property Name='Name'> Tessie Mae </Property>
</PersonLnk>
<PersonLnk>
  <Property Name='Name'> Susie </Property>
  <Property Name='Relationship' Key='pTMHamlett'>
    Aunt
  </Property>
</PersonLnk>
<PersonLnk Key='pBessieHamlet'>
  <Property Name='Name'> Bessie </Property>
</PersonLnk>
... etc ...

<!-- Event properties -->
<Property Name='When' Value='1899-01-03'>
Jan 3 1899 </Property>
<Property Name='Where' Key='wAmericus'>
Americus Ind </Property>
</SourceLnk>
</Event>

```

Let us now look at the transcription of the letter, as it might appear in the Resource entity:

```

<Text>
<ms>
<p align='R'>
Americus Ind<br/>
Jan 3 1899
</p>

<p>
Dear <PersonRef Key='pWFHamlett'>papa</PersonRef>:<br/>
<indent/>I thought I would write<br/>
to you this <Alt Value='evening'>eveng</Alt> to let you know that<br/>
we are as well as usual and hope this<br/>
may find you the same <PersonRef>George</PersonRef> and
<PersonRef>Ann</PersonRef><br/>
was up to day. <s style='*'><ucf>they * * * * </ucf></s><br/>
<s style='*'><ucf>* * </ucf></s><PersonRef >Aunt Susie</PersonRef> is
sick. she has<br/>
got a ugly <Alt Value='swelling'>swelled</Alt> on her hip they<br/>
<Alt Value='butchered'>bullchurd</Alt> <DateRef Value='1899-01-03'>to
day</DateRef> that little pig. it <Alt Value='weighed'>wayed</Alt><br/>
125 pounds when it was dressed.
</p>

<p>

```

<PersonRef Key='pBHamlett'>Bessie</PersonRef> got a letter from  
<PersonRef>Bennie Pepper</PersonRef><br/>  
to day he got home all <Alt Value='right'>write</Alt>. <br/>  
<PersonRef>George</PersonRef> said to tell you that he<br/>  
got a letter from worth judgment<br/>  
to day. <PersonRef Key='pBHamlett'>Bessie</PersonRef> is making her  
a<br/>  
apron and <PersonRef>manson</PersonRef> is fixing her shoes<br/>  
<PersonRef>manson</PersonRef> got mad at <PersonRef  
Key='pBHamlett'>Bessie</PersonRef> <DateRef Value='1898-12-  
31'>Saturday</DateRef><br/>  
<page/>  
but he wasnt mad long.  
</p>

<p>  
We was at Kurkehains <DateRef Value='1898-01-03'>saturday  
night</DateRef><br/>  
we popped pop corn had lots of fun.<br/>  
Well <PersonRef Key='pWFHamlett'>papa</PersonRef> when are you  
comming<br/>  
home.  
</p>

<p>  
We have not heard from <PersonRef>Jessie Emmons</PersonRef><br/>  
since they went home they left<br/>  
here </DateRef Value='1898-12-30'>friday morning</DateRef> about 6  
oclock in the<br/>  
morning and it was 10 oclock before they<br/>  
got to go away.  
</p>

<p>  
Well I guess this is all for this time<br/>  
so good night answer soon good by.<br/>  
<indent/><PersonRef Key='pTMHamlett'>Tessie Mae</PersonRef> to  
<PersonRef Key='pWFHamlett'>papa</PersonRef>  
</p>  
</ms>  
</Text>

This looks rather complicated but it makes perfect sense to the computer. Part of the reason is that all the person and date references are marked-up here, as well as more of the spelling errors. This transcription also records the crossing-out and the real pagination, paragraphs, and line breaks. Hence, it is a more accurate transcription than the simple plain-text edition, above.

It allows software to know where there are person and date references. It also distinguishes between persons that we can identify and persons that we cannot identify. For the former, this means a search can use both the name as-written and the name(s) associated with our conclusion-persons. We've also marked-up some clarifications, such as for the word "eveng" to indicate a simple misspelling. However, when this marked-up text is displayed on a computer screen then the power becomes self-evident.

Dear papa:

Evening

I thought I would write to you this eveng to let you know that we are as well as usual and hope this may find you the same George and Ann was up to day. Aunt Susie is sick. she has got a ugly swelled on her hip they bullchurd to day that little pig. it wayed 125 pounds when it was dressed.

Bessie got a letter from Bennie Pepper to day he got home all write. George said to tell you that he got a letter from worth judgment to day. Bessie is making her a apron and manson is fixing her shoes manson got mad at Bessie saturday but he wasnt mad long.

We was at Kurkehains saturday night we poped pop corn had lots of fun. Well papa when are you comming home.

30<sup>th</sup> December 1898

We have not heard from Jessie Emmons since they went home they left here friday morning about 6 oclock in the morning and it was 10 oclock before they got to go away.

Well I guess this is all for this time so good night answer soon good by.

Tessie Mae to papa

Those marked-up references allow the computer to highlight them, and optionally to make them selectable. The example here shows a pop-up window as it could appear when hovering over the uncertain word, or over an informal date reference. For a printed version, the same mark-up is capable to adding editorial brackets after an uncertain word, such as “eveng [*sic*, read: evening]” or simply “eveng [*sic*]”.

So, imagine that our genealogy tools all supported the concept of a word-processor specifically for genealogical and historical data? Imagine entering this transcription and having in-built support for marking references to people, and for correlating them with your core data. Imagine being able to flag a date, and to consult historical calendars. Imagine accessing an Internet-based authority with a database of place names, including spelling variations and boundary changes, in order to identify a place name. The result would be a live document (i.e. with links that can be followed) that is cross-referenced with your family tree, and which can be incorporated into timelines for reports.

Mark-up can also be used to link one section of text to another in order to generate reference notes, i.e. footnotes or endnotes. It can also link text to a datum in order to support an evidence-conclusion link. In conjunction with mark-up for generating inline citation references, the scheme can even generate complex citation references that include more than one cited work and analytical notes.

An unexpected consequence of having mark-up and structured narrative in the core data is that the data files can then be used as “genealogical documents”, analogous to Word or PDF documents. This appears to be an unexplored concept at the moment.

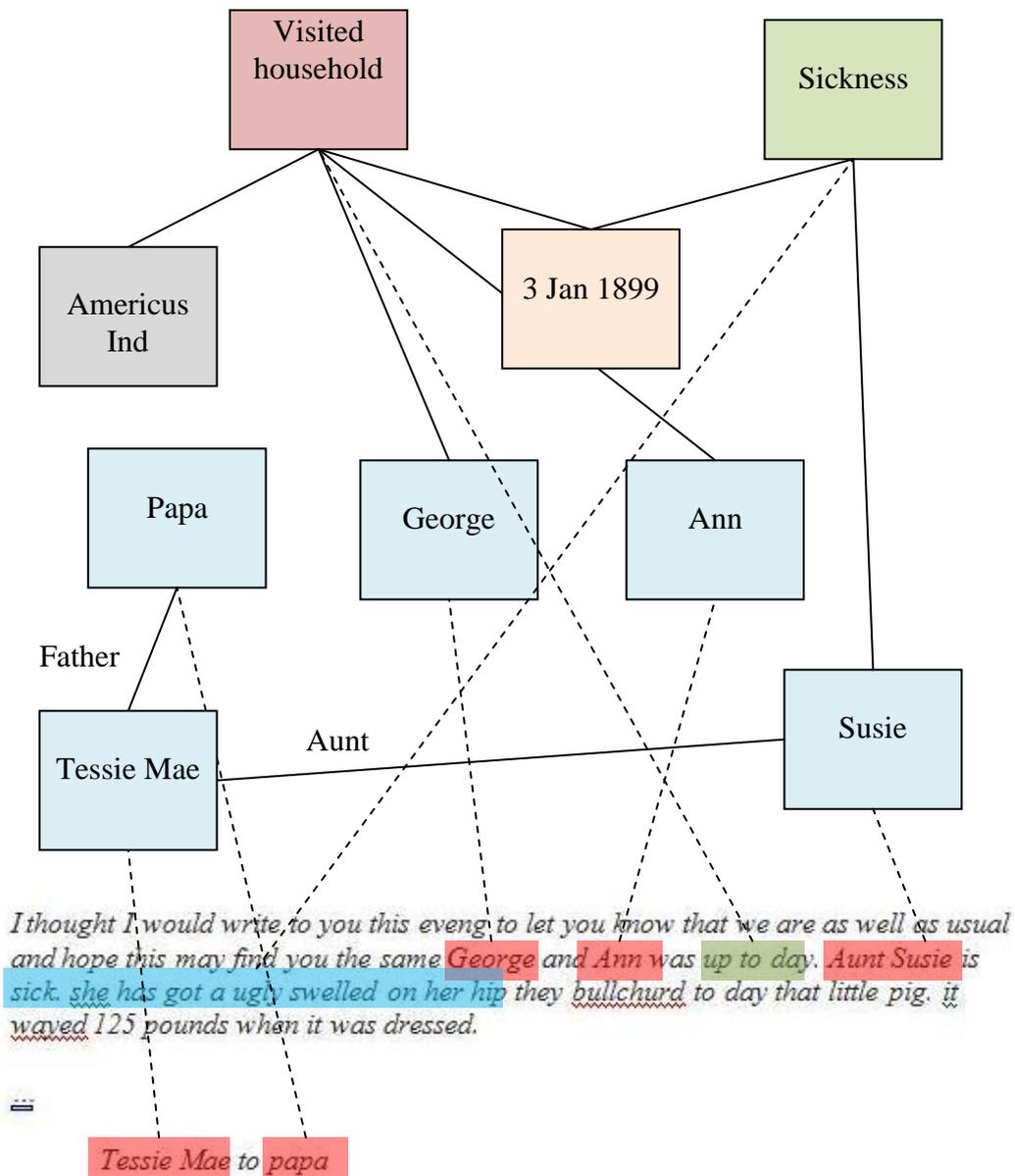
Imagine receiving a data file and being able to immediately look at it using a viewing tool similar to Word or Adobe Acrobat, and without having to store it in some specialised database first. It could present you with a family tree, or timeline, and allow you to navigate around the people and places using their hyperlinks, exploring their narrative content as you go.

## Analysis

So what about the analysis, I hear you ask. What I've shown is the connecting of textual references to conclusion items, as in a family tree. This is fine for use within a narrative essay, say for a family history story, but would it be typical when working from an original document or a transcription thereof? Where would the comparisons, correlations, resolution of conflicts, proof arguments, etc., be written?

Before STEMMA V4.0, the answer to this was unclear. In that version, though, a new Source entity was introduced that allowed a bottom-up assimilation of a source in order to make the correct identifications, and also to provide a trail that can be followed when drilling-down on a conclusion to see how it was arrived at.

The Source entity is mentioned in the code, above, but only in its briefest form where it identifies associated Citation and Resource entities. In its full form, it can support the implement of a [Graphic Organiser](#) that allows those references and their relationships to be analysed visually in a process known as [Link Analysis](#). I will present a simple link chart to illustrate this for Tessie's letter rather than provide more examples of STEMMA code.



Only after analysing the links, and possibly correlating them with details from other sources (using the related Matrix entity) then connections to conclusion Persons, Places, and Events can be made.