Does History Need Semantics?

• Something (“it”) happened
  – People participated in “it”, recorded “it” Documents describe events
  – Some people doubted “it” would “pan out” Historians do not always agree!
  – People rushed from far away to join “it” Events have a time and location
  – Many fortunes were made and lost due to “it” Events have context

• So what is “it”?
Identifiers

• At a minimum, we need to identify “It”
  - Were we talking about the same thing?

• We can now make statements about “It”
  - It created tens of thousands of new Californians
  - “Samuel Francis Cleveland” participated in it
    • My great great great grandfather
Genealogy And Identity

• In genealogy research, identity is key
  – Names are terrible identifiers
    • Ambiguous – Me or one of my namesakes?
    • Inconsistent between source documents
  – We need to link multiple documents
    • Is this the same person?
    • Birth, marriage, and death certificates; probate
    • Each document reveals new information
A Naïve Approach To Genealogy Semantics

- Simplest approach: a list of assertions
  - D has father C
  - C has father A
  - A has name “Felix Eubanks”
  - A was born in 1834 in Illinois
- Not enough for serious research
  - No provenance, no sourcing
An Improved Approach To Genealogy Semantics

- Document X describes Event Y
- Y has participants A, B, C...
- Y has Date and Location
- Researcher R makes assertions
  - $P = \text{“B is the mother of C”}$
  - $P$ is true, by evidence: Document X
  - C is unmarried, by evidence: living with parents

X,Y = 1880 Census
Richland Township
Madison County, AR
Provenance

- New information can come at any time
- Be prepared to revise
- Many interpretations

- In this case, someone might assume that Elisha is married to Rhoda. Elisha is the only one listed who is of the right age. However, her husband could be another unlisted son. Other evidence is needed to corroborate.
More Surprises

• The truth is...
  – Elisha really was unmarried at the time
  – Elisha was NOT the husband of Rhoda
  – Rhoda's husband was...
    • James Columbus Eubanks
    • Living elsewhere in 1880
    • Believed to have been working in California
Source Reconciliation

- Land records are another useful document type
  - This one is from an 1876 homestead for Felix

```
The United States of America

TO ALL TO WHOM THESE PRESENTS SHALL COME, GREETING,

Whereas, there has been deposited in the General Land Office of the United States a CERTIFICATE of the Register of the Land Office at Harrisburg A. H., whereby it appears that, pursuant to the Act of Congress approved 20th May, 1862, “To secure Homesteads to actual settlers on the public domain,” and acts supplemental thereto, the claim of Felix E. Evans has been established and duly consummated in conformity to law for the North East quarter of the South West quarter
```
Source Reconciliation
With Named Graphs

Graph LP1

- Graph X

- Land Patent

- Graph

- Image

- Aliquot Parts

- Location

- Type

- Patente

- Census

- Birth-year: 1834

- Birth-place: http://www.geonames.org/4120275

- Madison County, Arkansas

- Name: "Felix Eubanks"

- Father

- Mother

- Residence: 1880

- Location

- http://sws.geonames.org/4896861/about.rdf

- http://www.geonames.org/4120275
Some Related Vocabularies

- Event - Reified events with time & location
- Bio – Biographical events
- Relationship – Relationships between people
- FOAF – People, names, organizations, etc
- TriG – Named graphs
Data Sets

• Non-RDF
  – Footnote, Ancestry, FamilySearch, BLM GLO, USGenWeb, Google Books, many others

• RDF – Currently very few data sets
  – Library of Congress Recollections
  – New York Times (some years)
  – DBPedia
  – Geonames
To Do

• Much work to be done
  – Identify or develop vocabularies
  – Publish sample data
  – Develop best practices
  – Convince publishers to use Linked Data
  – Convert existing data
History and Genealogy
Semantics WG

- Working group to develop ontologies and best practices for publishing historical and genealogical data as RDF
  - http://groups.google.com/group/history-and-genealogy-semantics-wg/