

What comes after handwriting recognition?

Jon Morrey
BYU-FHTW 2020





The State of “Handwriting Recognition”

A.K.A “Automated Indexing”

80-80 Research Goal

Be able to automatically extract at least 80% of genealogical fields with 80% or greater word accuracy across our entire corpus (4B+ images)

Today



DIRECTORY.]

- Embley Mrs. Catherine 21 Blisworth st. Litherland
Geo. Huxley dispatch clerk 124 Salisbury rd 15
Jn. plumber 20 Briarwood rd 17
Wltr. Fred Courtney 231 Muirhead av 13
Embra Mark foreman 10 King's mount Oxton, B'head
Embry Mrs. 27 Mallaby st. Birkenhead
Rt.municipal clerk 22 Mount rd. Higher Tranmere
Emerson Ernest, J.P. 39 Ravenswood av. Rock Ferry.
T N Rock Ferry 581
Misses Florence & Ettie E. 30 Victoria drive West
Kirby, Wirral, Cheshire
Fras. Josph. cabnt. mkr. 12 Gloucester rd. Anfd.E
Geo. 16 Fairmead rd 11
Geo. 9 Smith drive 20
Leonard Watkins 42 Dencourt rd 11
Miss Margt. 5 Baytree rd. Higher Tranmere
Stephen F. electrcl. engrnr. 9 Grauge rd. West
Kirby, Wirral, Cheshire. Hoylake 789
Wm. W., M.B., Ch.B., B.A.O.Belf. physcn. and
surgn. 83 Merton rd. Bootle. T N Bootle 969
Emery Alfd. butcher 5 Hanbury rd
Alfd. commrcil. travllr. 47 Allangate rd 19
Alfd. who. warehsmn. Russell bldgs. School la 1.



ALPHABETICAL

- Emmett Harold Jas. Farquhar
dard Yeast Co. Ltd.) 1
Swan 1187
Jas. postman 15 Merton dr
Mrs. Jessie 58 Webb st 17
Jn. 67 Barlow la 4
Jn. 9 Railton rd 11
Jn. bacon curer 12 Clint rd
Jn. boat bldr. 6 Orrell rd.
Jn. pork butcher 93 Wai
61 County rd 4; 98 V
Granby st 8—res. 5 Stra
Jn. & Sons pork butchers
Broadway, West Derby
and cooked meat purve
328); and 225 Kensington
Jn. W. 20 Menlove gdns. 1
Leonard Leslie electrcl. e
Walton
Mrs. Louisa 42 Royden st.
Luther 58 Arkles la. Anfie
Mrs. Margt. 7 Dane st. ©W
Mrs. Norah, Westlake,

Today



Family Tree Search Memories Indexing Activities Temple

Help

Messages

Jon Morrey

Records Family Tree Genealogies Catalog Books Wiki

Film # 101448986 ▾

Source Box

Attach to Family Tree

◀ Image 1283 of 1522 ▶

esday at 9
us. Contribu-
tions for
fully ap-
rs SUNDAY
AY 24 AND

announced. Interment at 12, St. Peter
Cemetery. Friends may call at 11 a.m. In-
BERRYS FUNERAL HOME, 7200
DETROIT AVE., 3-5 AND 7-9 P.M.
MONDAY.

RAGUSA

See Monastra notice.

od, beloved
mother of
ard H. Ned,
of nine, sis-
Marshall and
and the late
er and How-
's Lutheran
y, Nov. 12 at
all at THE
OME, Be-
D 7-9 P.M.
ontributions
Church.

of the late
of Mrs.
of Philadelphia

B ▾ Image Index

Name	Event Type	Event Date	Event Place	Gender	Age	Relationship to D...	Birth Date	Birth
Ted C Stevenhagen	Obituary	11 Nov 1973	Cleveland, Cuyahoga, Ohio, United States	Male				
Edward J Hovorka Sr	Obituary	11 Nov 1973	Cleveland, Cuyahoga, Ohio, United States	Male				

This image courtesy of [GenealogyBank](#)

Interment Holy Cross Cemetery.
Friends may call at CHAMBERS
FUNERAL HOME, 4420 ROCKY
RIVER DR. AT PURITAS. SUN-
DAY 3 TO 5 AND 7 TO 9 P.M.

STEVENHAGEN

Ted C. Stevenhagen, beloved hus-
band of Dorothy (nee Calhorn).
brother of Catherine Otto of Del-
tona, Fla., two nieces and two
nieghbours. Nov. 7. Contributions to
the memorial fund of Lake Shore
Christian Church would be appre-
ciated. Memorial services Sunday,
Nov. 11 at the Lake Shore Christian
Church (2899 Lake Shore
St., Euclid, O.) at 3 p.m.

STOA

Anne M. Stoa (nee Verciglio-
tti) age 61, wife of Vernon A.,
mother of Karen (Mrs. William)
Cordes of Elkhart, Ind., and Eric
Jon of South Russell, grandmoth-
er of Billy, Kathy and Laurel.
Sister of Phyllis (Mrs. Willis)
Moore of Cleveland. Dorothy

bond or boarder, sister of John
Gordon, brother of Luther R. and
Mrs. Jimmie L. Smith. Family re-
ceiving friends at STROWDER'S
FUNERAL CHAPEL, 818 EAST
105TH ST., MONDAY 8:30-9:30
P.M. Services and burial Wednes-
day in Malvern, O. THE MAT-
THEWS FUNERAL HOME OF
CANTON, O., IN CHARGE.

WOLOSZ

Frank Wolosz, beloved husband of
Mary Ann (nee Sitarz), dearest
brother of Edward Wallace and the
late Charlotte Mantel, dearest un-
cle. Funeral Monday, Nov. 12, at
9:30 a.m. from MOSINSKI FUNER-
AL HOME, 3675 E. 65TH ST., serv-
ices 10 a.m., at Our Lady of
Lourdes Church. Interment Calvary
Cemetery. VISITING HOURS SUN-
DAY 2-5 AND 7-9 P.M.

YARMUS

Sophye R. Yarmus, dearly beloved
wife of Joseph, devoted mother of
Jane C. Dominguez, James C.,
dearest sister of Edith Weinberg.

Today

2001

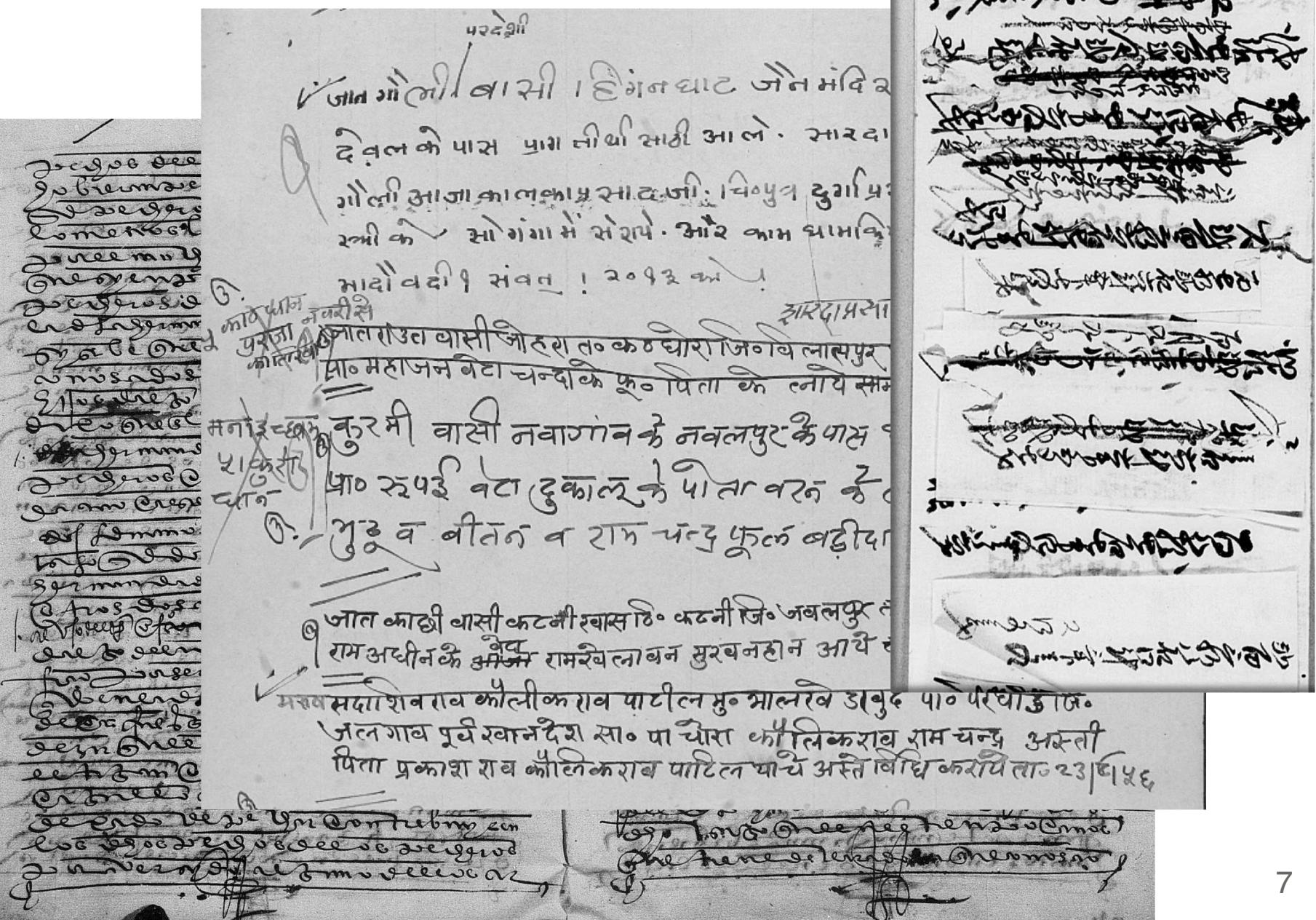
Leygally, Alzaber jch
fondet mit Guerard.

Mengalt. L'isoly fawij branc
die klos lob Maria
galt. nach der branc
aus der Heil Reit.
Als aber all die Fawij
nach dem ordre omstey
werden.

Und den Disposicione abende fayre
forni don. alles d'or ein gescht.
Conspicuit. bis Memorial des fornij
an Liège. die Memoral des fornij
fayre. die Conspicuit. und vnd gescht
frogetheus. und vnd gescht
und vnd geschtet, werden.

Et auant le fest de la mort de Jésus
vngescht. die Dioces. et Temples.

Not Yet



(Surprisingly) Not Yet

P E R S O N ID	LITERACY: ANSWER FOR PERSONS 11 YEARS AND OLDER		FULL TIME EDUCATION: ANSWER FOR PERSONS 3 YEARS AND OLDER							P12c HIGHEST EDUCATIONAL GRADE COMPLETED FOR THOSE CURRENTLY ATTENDING SCHOOL OR ATTENDED IN THE PAST What is the highest grade (form/class/level) [NAME] has completed at that level of schooling?		
	P11 In what language can [NAME] read and write?		P12a Has [NAME] ever attended school or is [NAME] attending school now?			P12b What is the highest level of schooling [NAME] is attending now/attended in the past?						
	1 None (not literate) 2 English only 3 Ghanaian language only 4 English and Ghanaian language 5 English and French 6 English, French and Ghanaian language 7 Other (Specify)					IF NEVER, GO TO P13a						
			N E V E R	E O N O W	P A S T	01 Nursery 02 Kindergarten 03 Primary 04 JSS/JHS 05 Middle 06 SSS/SHS 07 Secondary 08 Voc/technical/commercial 09 Post middle/secondary certificate (teacher training/college of education, agric, nursing, etc) 10 Post secondary Diploma (University diploma, HND, teacher training/college of education diploma, etc) 11 Bachelor degree 12 Post graduate (Cert., Diploma, Masters, PHD, etc)						
	CODE	OTHER (SPECIFY)										
	0 1	4	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		0 4	3
	0 2	1	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		0 4	1
	0 3	4	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		0 3	5
	0 4		<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		0 3	0
	0 5		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>			
0 6		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
0 7		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
0 8		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
0 9		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
1 0		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				

Hopefully Someday

trinotata f. C. podomignea. Il suo coto (am
po) è di circa 10 cm. e la sua lunghezza
è di circa 15 cm. È un pesce molto diffuso
e si trova nei fiumi, laghi e stagni d'acqua
dolce. La carne è bianca e succulenta.
È molto gustoso e si può cucinare in
varie maniere. È particolarmente buono
fritto o alla griglia. È un pesce molto
popolare in tutta Italia.

Infinity Project

- Volunteers are more likely to index records containing their own family
- “Just in time” quality vs. up-front perfection
- Removing other bottlenecks

[The Year 1812]

Page 253

N^o 527 } James Morrey of this Parish Cordwainer
and Mary Topham of this parish Spinster were
Married in this Church by Banns
this fourth Day of September in the Year One Thousand eight Hundred
and twelve By me John Latham Curate
This Marriage was solemnized between Us James Morrey
In the Presence of Charlotte Robinson
Henry Morrey

Event Type: Marriage

Event Date: 4 September 1812

Event Place: St. Mary's, Nantwich, Cheshire, England, UK

Bride: Mary Topham

Groom: James Morrey

Groom's Occupation: Cordwainer

Bride's Occupation: Spinster

Marriage Officiator: John Latham

Marriage Officiator Title: Curate

Marriage Validation: Banns

Witness 1: Charlotte Robinson

Witness 2: Henry Morrey

Page Number: 253

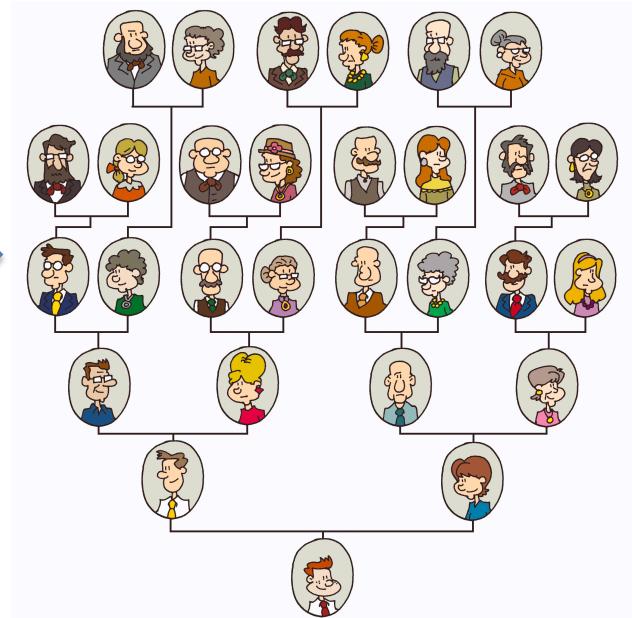
Record Number: 527

So we're done, right???

End goals of FH records

- A family tree containing one—and only one—record of every person who has lived on the Earth
- Tell a coherent (and hopefully entertaining) story
- Render value to all people worldwide

AI to the rescue?



How might we structure the data differently to achieve these goals?

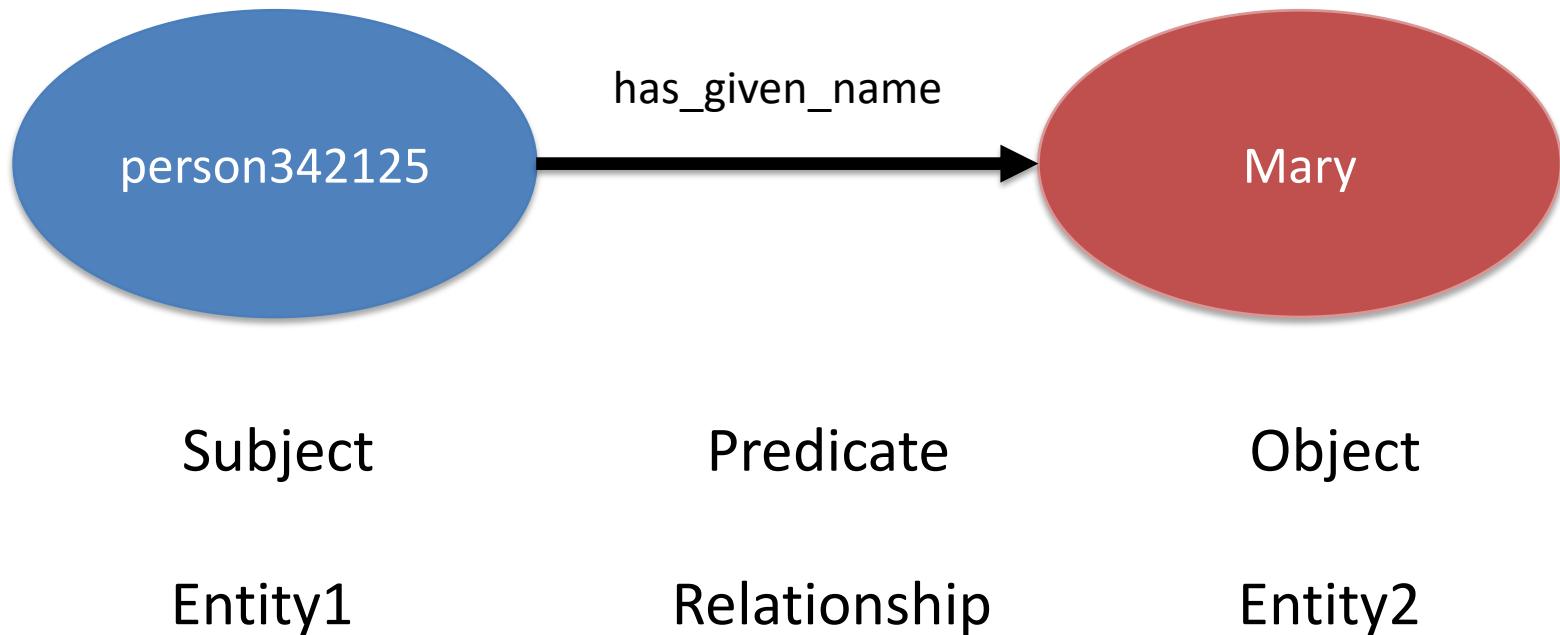
- Break open the “silos”
- Take advantage of *everything* we know
- Make *everything* queryable



Knowledge Graphs

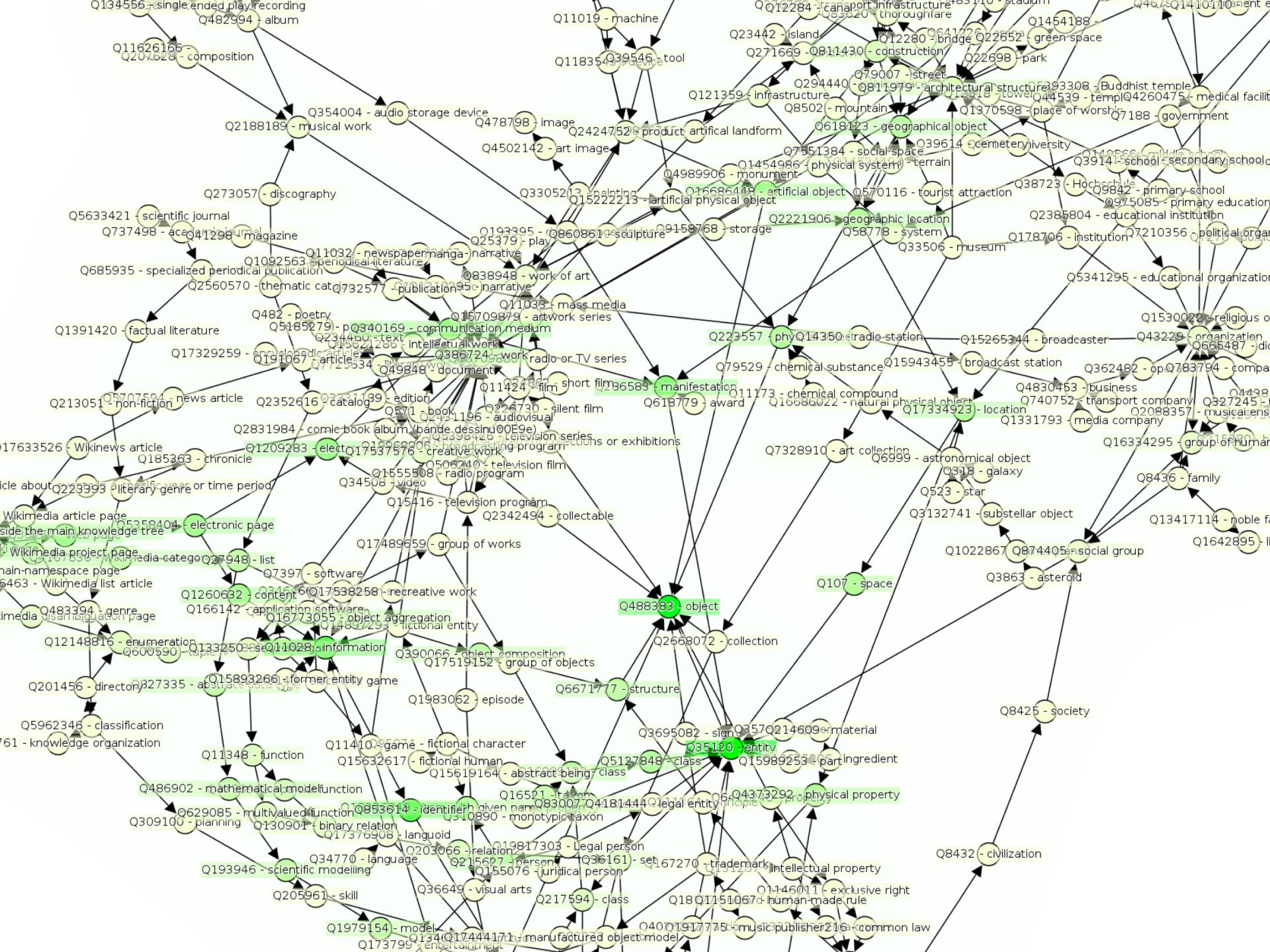
(Rebooted)

Triple Stores





- Launched 2012
- Backs Wikipedia and sister projects
- Funded by the Allen Institute for AI, Gordon & Betty Moore Foundation, and Google
- Currently contains almost 80M entities



Ancient intellectual network

This dataviz maps the relationships between master and student from Socrates to the end of the hellenistic period using data I'm currently adding to [Wikidata](#). As of today, this work is not completed (and, consequently, this graph is merely an early beta)

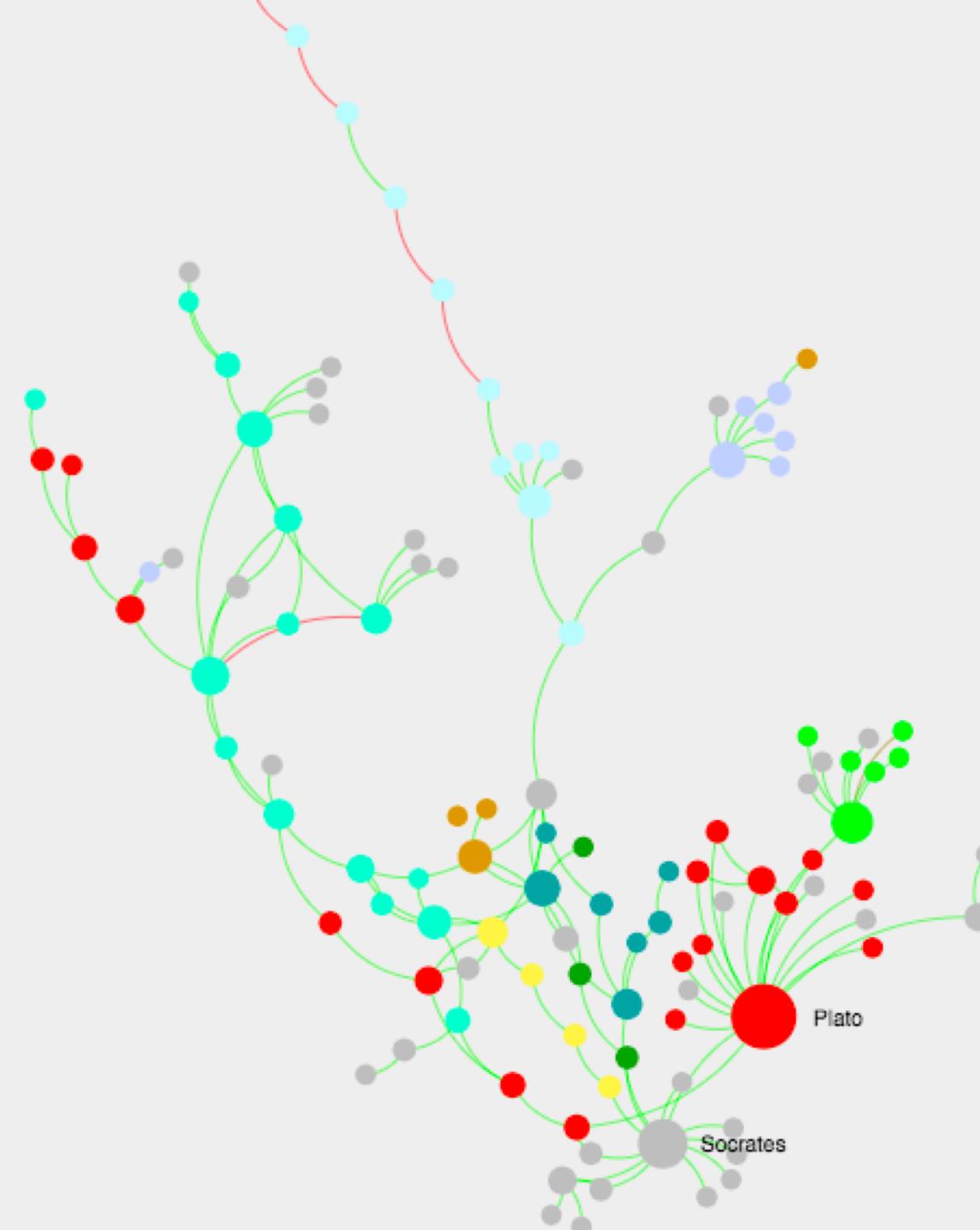
The philosophical communities are represented with the following color chart:

- █ Platonism
- █ Peripatetician (Aristotelism)
- █ Megarian school
- █ Cyrenaic school
- █ Epicureanism
- █ Stoicism
- █ Cynic school
- █ Eretrian school
- █ Pyrrhonism
- █ No known affiliation

Since the original data are quite fragmentary, the student-master links between each philosopher have been colored according to their accuracy :

- Well-established
- Likely
- Hypothetical

i More about this visualisation



Ancestors of George Washington

Wikidata Query Service Examples Help More tools 文 English

1 #George Washington and his ancestors...
2 SELECT DISTINCT ?item ?itemLabel ?dateofbirth ?image
3 WHERE
4 {
5 wd:Q23 (wdt:P22|wdt:P25)* ?item .
6 OPTIONAL { ?item wdt:P569 ?dateofbirth } .
7 OPTIONAL { ?item wdt:P18 ?image. }
8 SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 } ORDER BY DESC(?dateofbirth)

4643 results in 3669 ms </> Code Download Link

Search

item	itemLabel	dateofbirth	image
wd:Q23	George Washington	22 February 1732	 commons:Gilbert Stuart Williamstown Portrait of George Washington.jpg
wd:Q458119	Mary Ball Washington	30 November 1708	 commons:Mary Ball Washington(Pine).jpg
wd:Q768342	Augustine Washington	12 November 1694	 commons:Augustine Washington.jpg
wd:Q3314035	Mildred Gale	1 January 1671	
wd:Q3219928	Lawrence Washington	1 September 1659	 commons:Portrait of Lawrence Washington B&W.jpg
wd:Q16738459	Joseph Ball	1 January 1649	
wd:Q76112240	Baldwin de Redvers, 1st Earl of Devon	11 May 1645	

Ancestors of George Washington

Wikidata Visualization

Charts

Query Examples

About



```
1 #George Washington and his ancestors...
2 SELECT DISTINCT ?item ?gender ?itemLabel
3 WHERE
4 {
5   wd:Q23 (wdt:P22|wdt:P25)* ?item .
6   ?item wdt:P19 ?placeofbirth .
7   ?item wdt:P21 ?gender .
8   ?placeofbirth wdt:P625 ?geoloc
9   SERVICE wikibase:label { bd:serviceParam
10 }
```

Table

Basic

Trees

Maps

More



Submit

717 results found



Data options

Coordinate

geoloc

Color

gender

Radius

none

Label

none

Link from coordinate

none

Link to coordinate

none

Display options

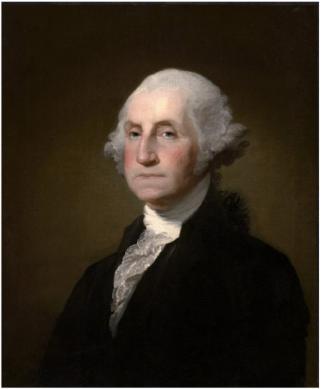
Canvas options

Ancestors of George Washington

Wikidata Query Service Examples Help More tools English

```
1 #defaultView:ImageGrid
2 SELECT DISTINCT ?item ?itemLabel ?dateofbirth ?image
3 WHERE
4 {
5   wd:Q23 (wdt:P22|wdt:P25)* ?item .
6   OPTIONAL { ?item wdt:P569 ?dateofbirth } .
7   OPTIONAL { ?item wdt:P18 ?image. }
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 } ORDER BY DESC(?dateofbirth)
```

Image grid ? 4643 results in 4996 ms </> Code Download Link



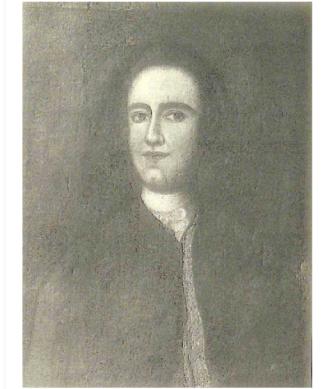
commons:Gilbert Stuart Williamstown ...
22 February 1732
George Washington



commons:Mary Ball Washington(Pine).jpg
30 November 1708
Mary Ball Washington



commons:Augustine Washington.jpg
12 November 1694
Augustine Washington



commons:Portrait of Lawrence Wa...
1 September 1659
Lawrence Washington



commons:Augustine Warner Jr.gif
1 January 1642
Augustine Warner, Jr.



George Washington



Mary Ball

97%

Descendants of Genghis Khan

```
1 # Descendants of Genghis Khan
2
3 SELECT
4 ?personItem ?englishName ?chineseName ?
5 ?childItem ?childEnglishName ?childChir
6 ?CBDB ?depth
7 WHERE
8 {
9   SERVICE gas:service {
10     gas:program gas:gasClass "com.bigdata.sparql.lang.gas.GAS"
11     gas:in wd:Q720 ;
12     gas:traversalDirection
13     gas:out ?personItem ;
14     gas:out1 ?depth ;
15     gas:maxIterations 5 ;
16     gas:linkType wdt:P40 .
17 }
```

Submit

434 results found

< ^ >

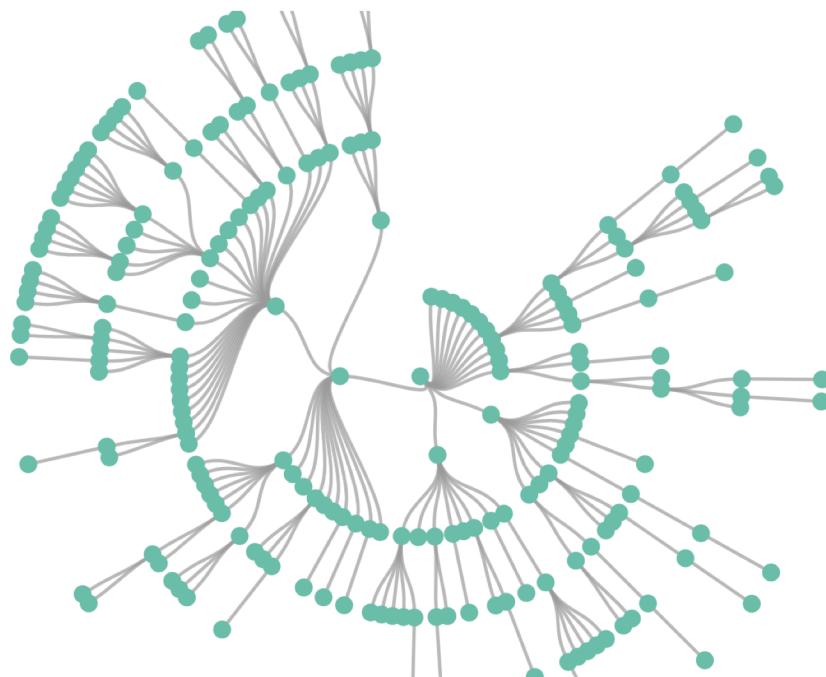
Data options

Link from	Link to
personItem	childItem
Label	Color
none	none

Display options

Canvas options

Table Basic Trees Maps More



C D L

Wikidata People with a Family Tree ID

Wikidata Query Service Examples Help More tools English

```
1 SELECT ?item ?itemLabel ?fsID
2 WHERE
3 {
4   ?item wdt:P31 wd:Q5 .
5   ?item wdt:P2889 ?fsID .
6   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
7 }
```

Play

2732 results in 1487 ms </> Code Download Link

item	itemLabel	fsID
------	-----------	------

wd:Q26718317	Richard Alsager	LZGX-KLN
wd:Q26928891	Feliks Szczęsny	L81B-DBS
wd:Q26972887	James Paul Boland	LT85-X5Q
wd:Q30349633	Valentine Charles Outwater	LD2M-XN6
wd:Q33026914	Michael Joseph Mulvihill	LTF8-NPC
wd:Q33082977	Josephine Corliss Preston	K8F5-CMB
wd:Q33103568	Margaret Elizabeth Tout	KWZW-KKP
wd:Q33110665	Virgil Lee Highland	L4MJ-BQK
wd:Q34119528	Donald George Wilhelm, Jr.	LTVG-Z4B
wd:Q36423395	Lars Öknaberg	LDMZ-BC7

Descendants of John Lathrop who played sports

Wikidata Query Service Examples Help More tools English

```
1 #Descendants of John Lathrop who played sports
2 SELECT DISTINCT ?item ?itemLabel ?dateofbirth ?teamLabel
3 WHERE
4 {
5   wd:Q6245304 (wdt:P40)* ?item .
6   OPTIONAL { ?item wdt:P569 ?dateofbirth } .
7   ?item wdt:P54 ?team .
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
9 }
```

2 results in 482 ms </> Code Download Link

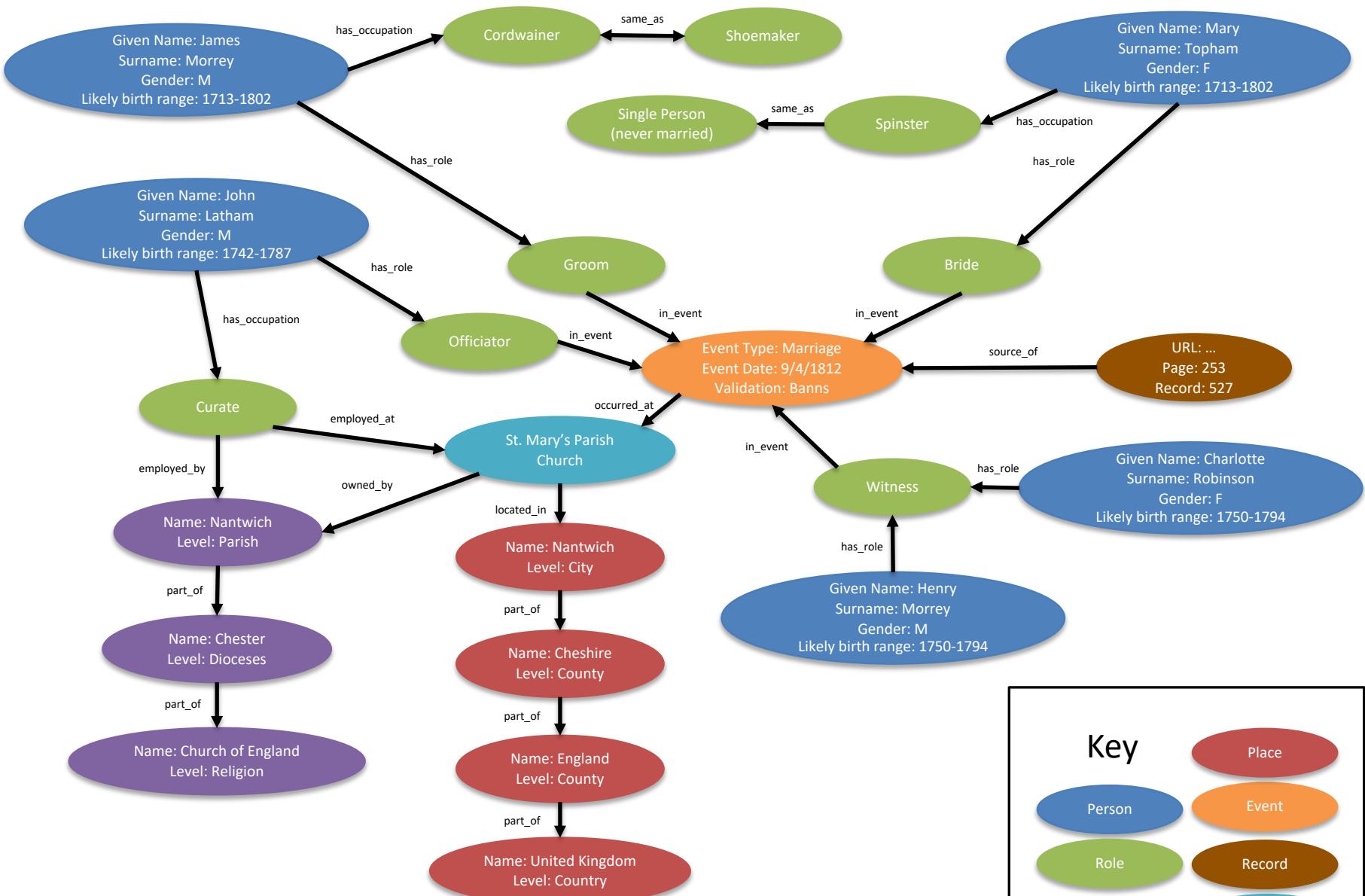
item	itemLabel	dateofbirth	teamLabel
wd:Q324742	Prescott Bush	15 May 1895	Yale Bulldogs football
wd:Q23505	George H. W. Bush	12 June 1924	Yale Bulldogs

Which facts are “genealogical”?

- Name, gender, birth date, relationships, etc.
- Occupations?
- Religion?
- Business associates?
- Whereabouts on a given day?
- Sports team?

I want to know...

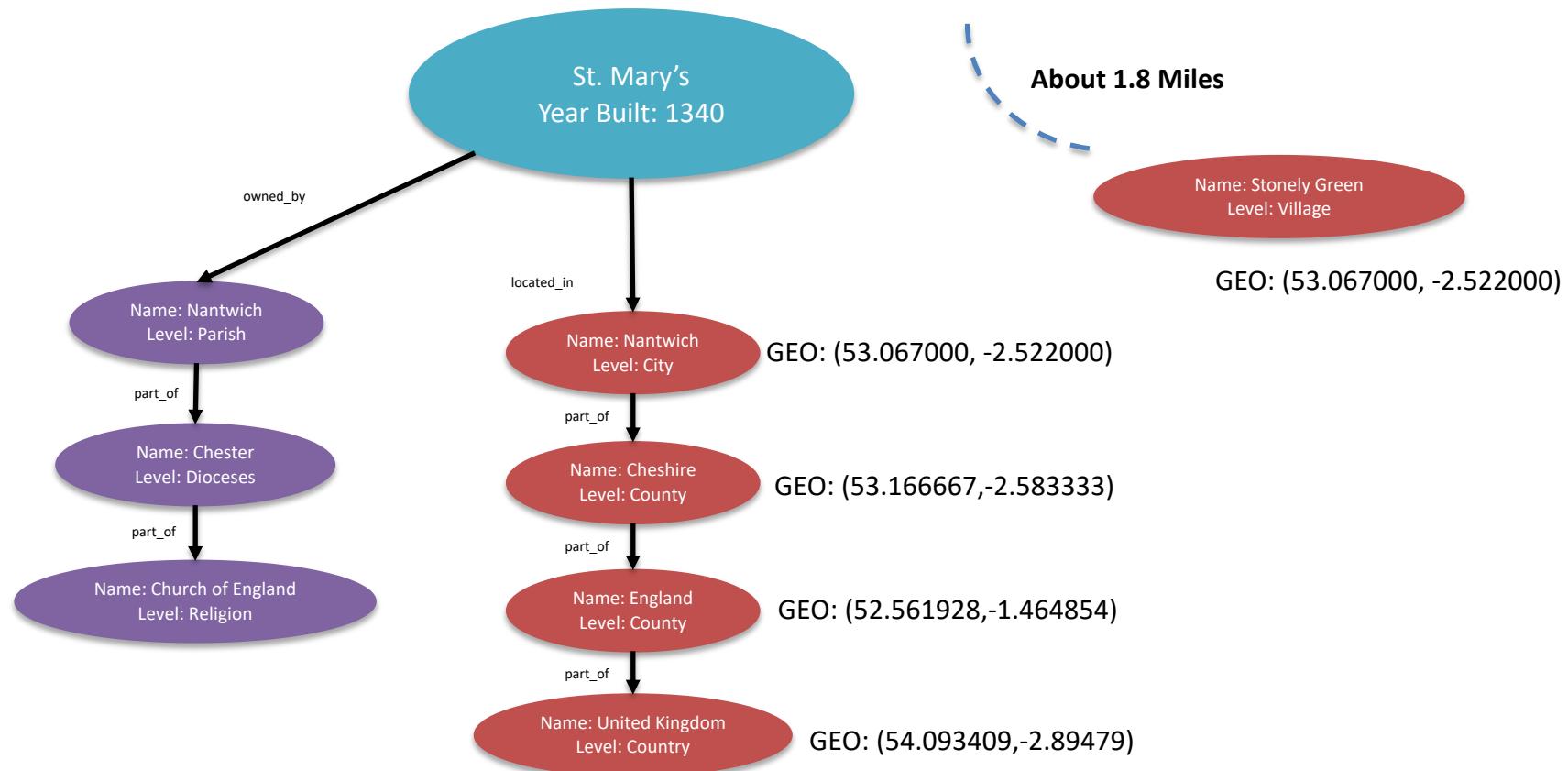
- Who might have taught James Morrey the shoemaker trade?
- Were the witnesses related? How?
- Who *might* have been Henry Morrey's wife (given age, proximity, religion, social status, etc.)?
- How many *different* James Morreys are there within a 50 mile radius?
- Does this possible duplicate have an “alibi”?



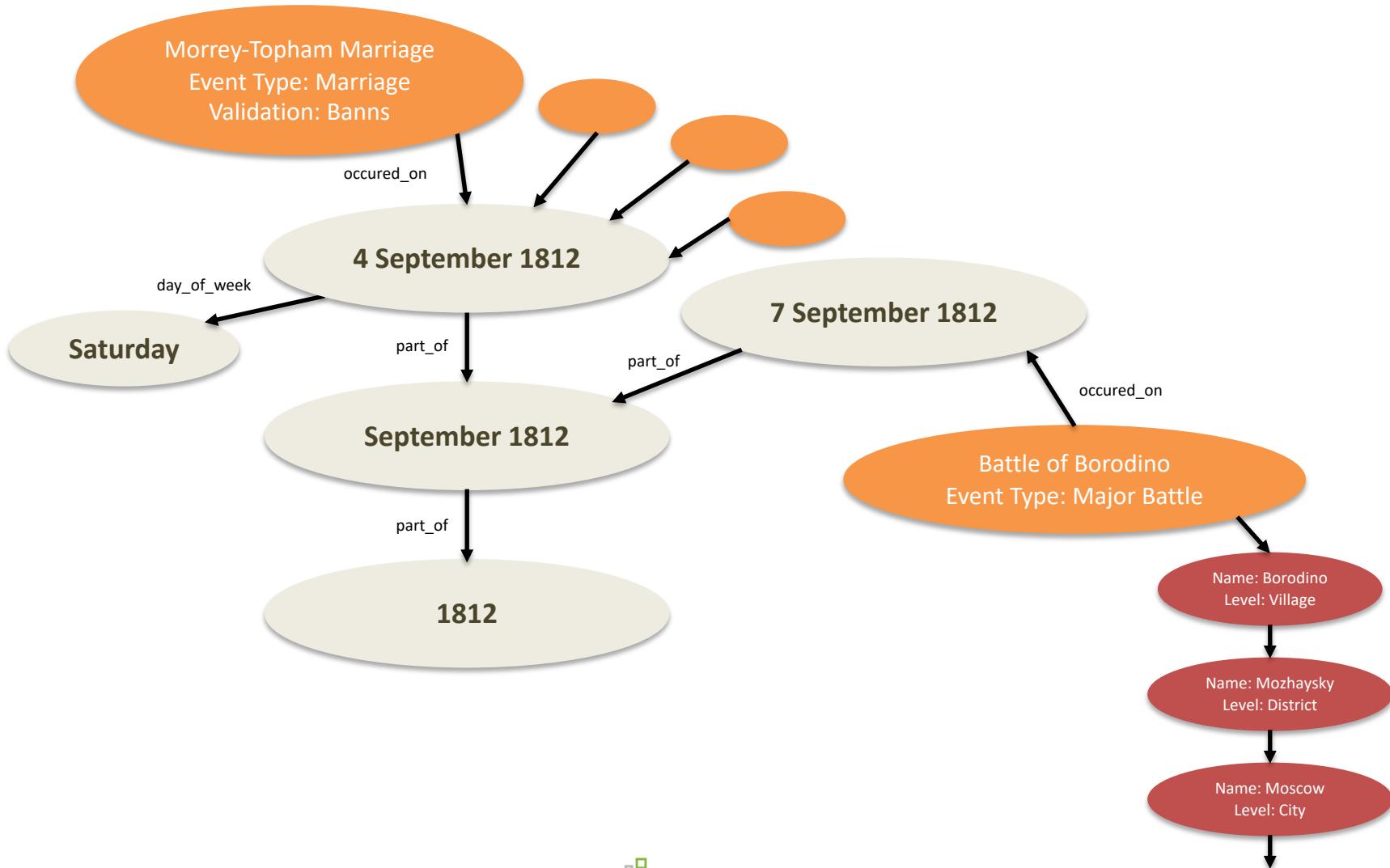
Other World Knowledge

GEOPOLY: [(53.0672,-2.5210), (53.0672,-2.5202), (53.0668, -2.5202), ..]

GEO: (53.067112, -2.520631)



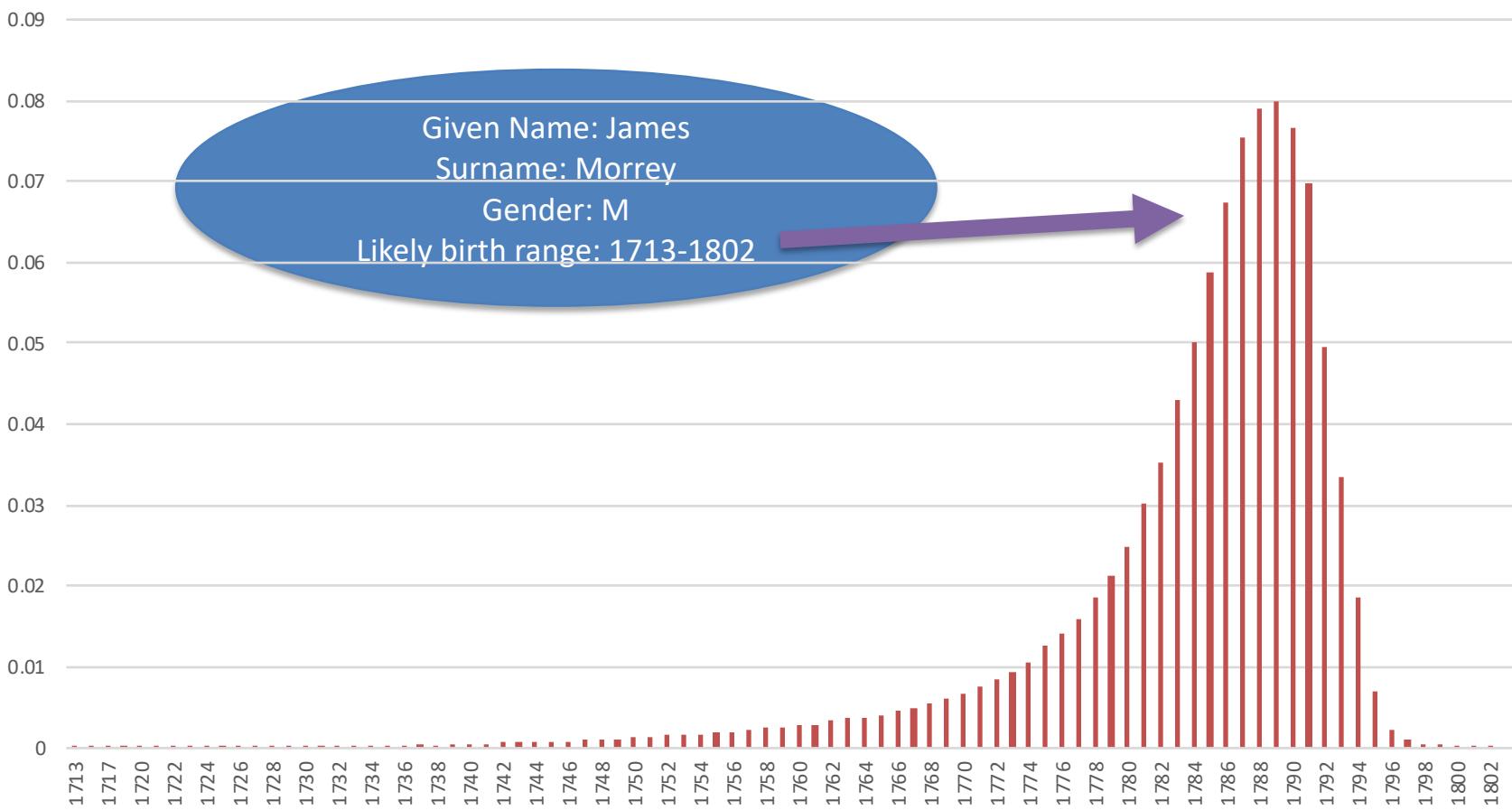
Other World Knowledge (Cont.)



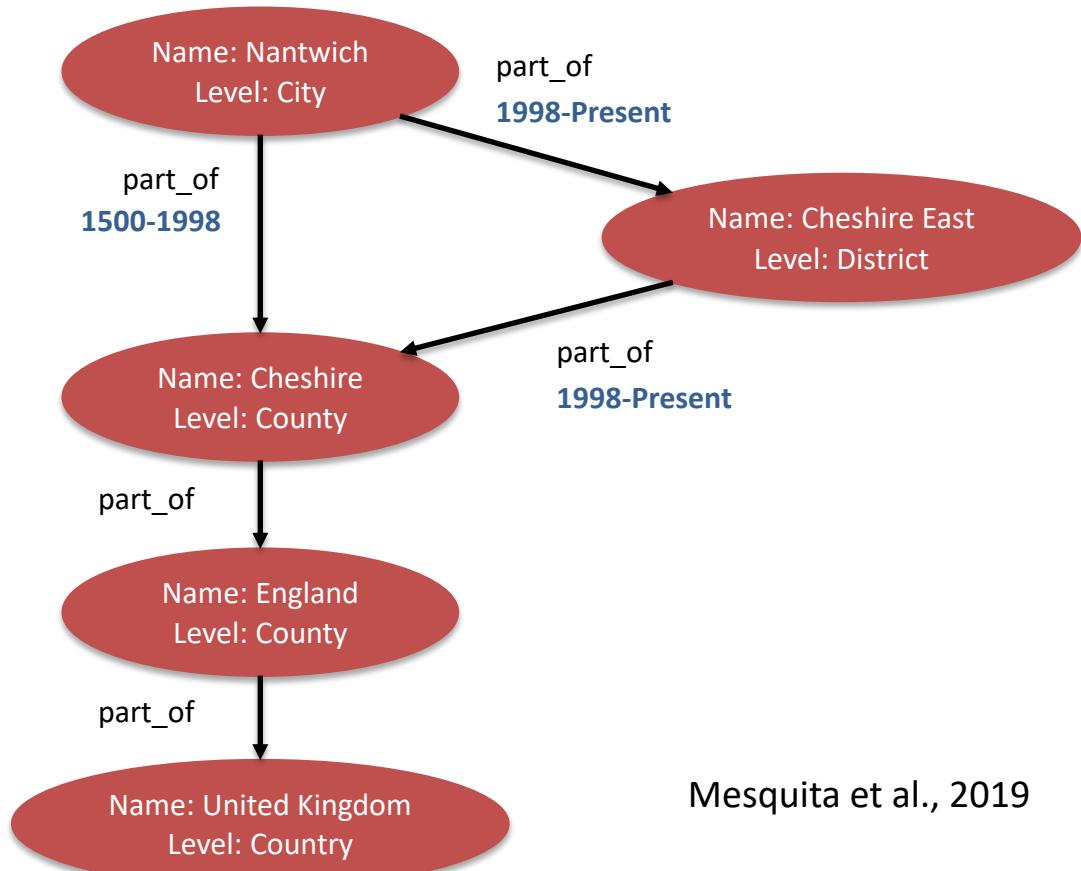
“Facts” as Probabilities

Given Name: Charlotte
Surname: Robinson
Gender: F (0.94)

Ranges as Distributions



Temporal Qualifiers



Other Digital Attributes

- Autographs
- Thumbprints
- DNA data
- Photos

Sources of knowledge

- Historical Records
- Family Tree
- Standards/Authorities
- Catalogs
- Books
- Wikipedia and Help Wikis
- Family History Guide
- Public genealogy sites
- Other knowledge graphs
- ...

How Big?

- ~20 entities per image * 4B images \approx 80B entities
- 2-3 times as many triples \approx 160B
- Not just records—Family Tree, etc... 1T triples?

Collapsing the Graph

- The knowledge graph is typically not the “system of record”
- Knowledge Fusion
- Regular rebuilds

Is this even possible?

AI

Diffbot launches AI-powered knowledge graph of 1 trillion facts about people, places, and things

KYLE WIGGERS @KYLE_L_WIGGERS AUGUST 30, 2018 8:00 AM



diffbot

Image Credit: Diffbot

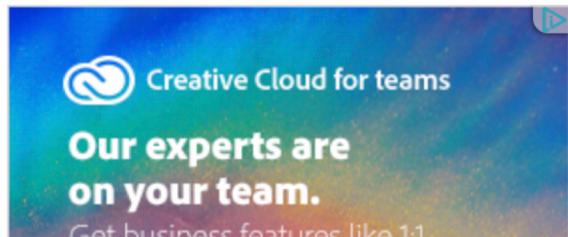
If you've ever performed a Google search for a celebrity, famous landmark, or

VB TRANSFORM

The AI event for business leaders

San Francisco
July 15 - 16

[Learn More](#)



Diffbot

- Fully automated
- 1T triples (10B entities)
- Rebuilt from scratch every 4 days

...but good ontologies are *hard*.

- Some of the ontology already exists
(Wikidata)
- Deal with it! It'll be worth it ☺

Can AI “learn on its own” from a Knowledge Graph?

- Short answer: not really (yet)
- We'll never know if we don't have one!



Questions/Comments?

Resources: Wikidata Query Builder

<https://query.wikidata.org/>



Resources: Diffbot TWIML&AI talk

<https://twimlai.com/twiml-talk-319>



References

- Filipe Mesquita, Matteo Cannaviccio, Jordan Schmidek, Paramita Mirza, Denilson Barbosa. 2019. KnowledgeNet: A Benchmark Dataset for Knowledge Base Population. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 749–758, Hong Kong, China, November 3–7, 2019. Specifically, see “Beyond Binary Relationships” in appendix.