

**Historical Gazetteer Integration:
CHGIS, Regnum Francorum & GeoNames**

Working Digitally with Historical Maps

AAG 2012

Merrick Lex Berman & Johan Åhlfeldt

Harvard University

Regnum Francorum

Geographic Info Retrieval [GIR]

Search for Place Names, features, and / or locations

Queries to existing gazetteer services **work** for:

names, feature types, footprints, administrative districts...

GeoTemporal Info Retrieval [GTIR]

Search for Place Names, events, and / or locations with dates

Queries to existing gazetteers **don't work** because:

dates are optional or entirely missing

Extend GeoNames Schema By Adding Dates

geoname

geonameid int,
name varchar(200),
asciiname varchar(200),
alternatenames varchar(6000),
latitude float,
longitude float,
fclass char(1),
fcode varchar(10),
country varchar(2),
cc2 varchar(60),
admin1 varchar(20),
admin2 varchar(80),
admin3 varchar(20),
admin4 varchar(20),
population bigint,
elevation int,
gtopo30 int,
timezone varchar(40),
moddate date

alternate name

alternatenameid int,
geonameid int,
isoLanguage varchar(7),
alternateName varchar(300),
isPreferredName boolean,
isShortName boolean,
isHistoricalName boolean,
existDate date,
endDate date,
srcCiteName varchar(300),
srcCiteLink varchar(300)

Testbed Methodology

- Automate process for matching placenames via open API
- Test methods for geonomial and geospatial matching
- Discover problems in current datasets, schemas
- Evaluate matching results
- Propose a schema for exchange of historical placename data
- Recommend the types of attributes needed to refine matching

Existing Gazetteers as Web Services

Google code

Google Maps API Web Services

The Google Geocoding API

[What is Geocoding?](#)
[Audience](#)
[Usage Limits](#)
[Geocoding Requests](#)
[Geocoding Responses](#)
 [JSON Output Formats](#)
 [XML Output Formats](#)
 [Status Codes](#)
 [Results](#)
 [Address Component Types](#)
[Reverse Geocoding](#)

Maps API Web Services
[Overview](#)
[Directions API](#)
[Distance Matrix API](#)
[Elevation API](#)
[Geocoding API](#)

GeoNames WebServices overview

WebService	XML	JSON	RDF	CSV	TXT	RSS	KML
children	XML	JSON					
cities	XML	JSON					
countryCode	XML	JSON			TXT		
countryInfo	XML	JSON		CSV			
findNearbyPlaceName	XML	JSON					
gtopo30	XML	JSON			TXT		
hierarchy	XML	JSON					
neighbourhood 	XML	JSON					
search	XML	JSON	RDF				



Placemaker

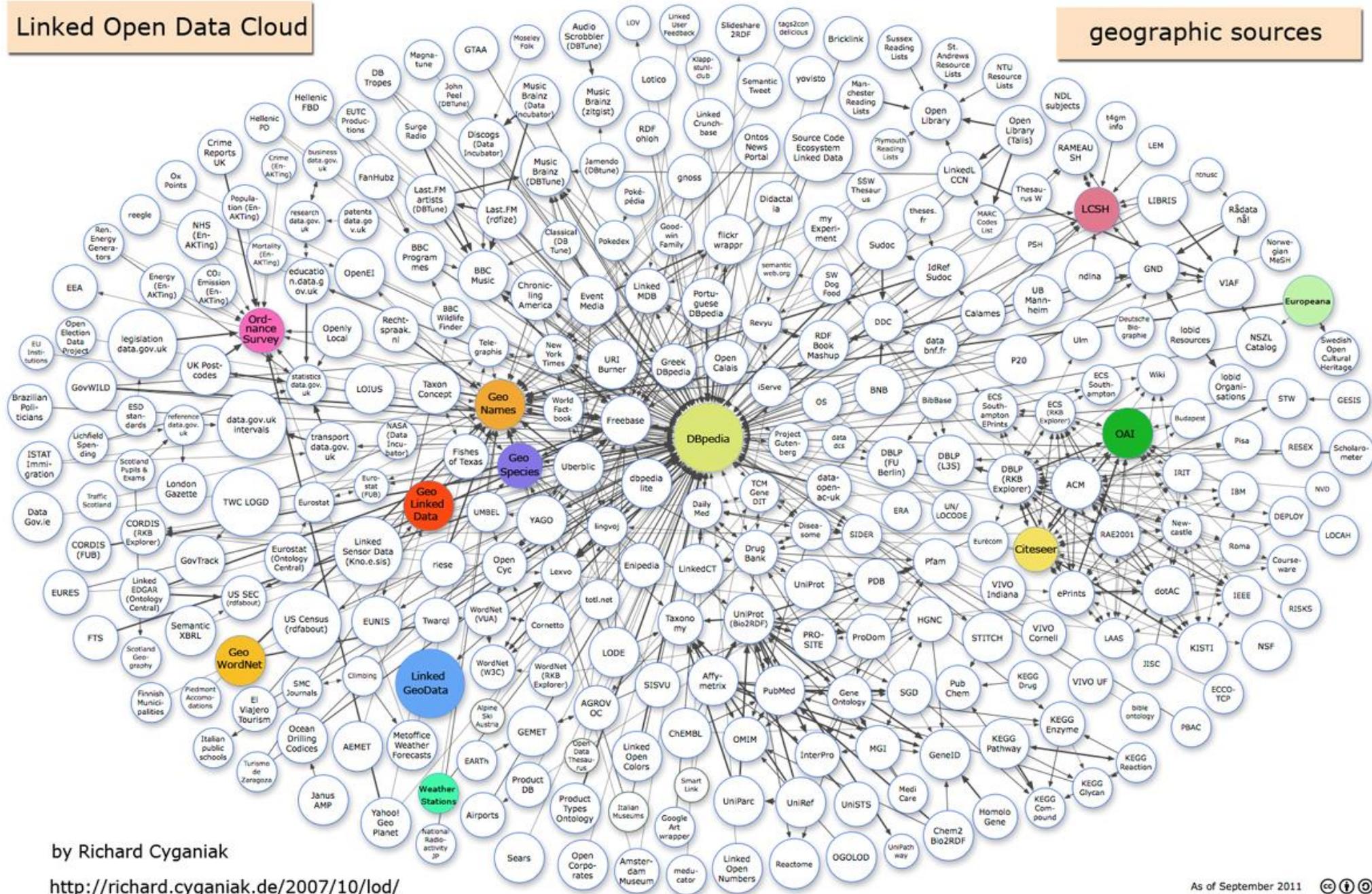
PLACES V. PLACE NAMES

Yahoo! Placemaker aims to capture all forms of how a place is called, and disambiguate the place-name to its canonical form. The platform identifies and disambiguates every place-name to a specific place concept, referenced by its unique identifier, the Where-on-Earth ID (WOEID). WOEIDs always reference a place, not a place name. For example, "New York", "New York City", "NYC", and "the Big Apple" are all variant names for WOEID 2459115. If Placemaker find these variants in the text, it will understand them to be multiple appellations of the same place.

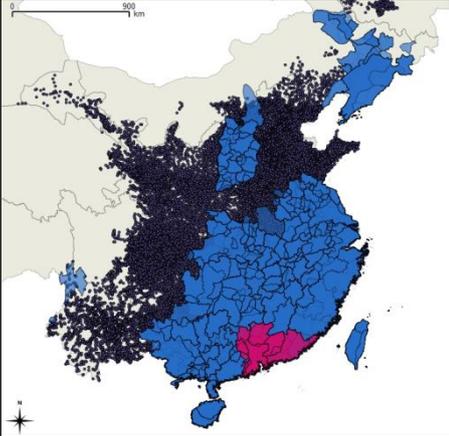
Linked Geographic Data on the Semantic Web

Linked Open Data Cloud

geographic sources



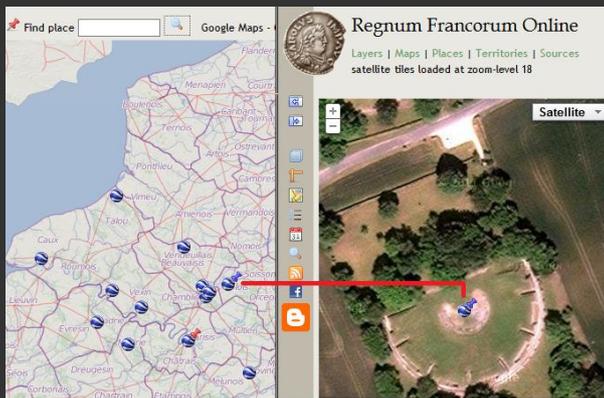
Existing Historical Gazetteers



Gazetteer of 50,000+ historical placenames

Administrative Hierarchy

Points have description of “present location”



Gazetteer of 10,000+ historical placenames

Administrative Hierarchy

Includes explicit links to GeoNames ID



Gazetteer of 120,000+ placenames

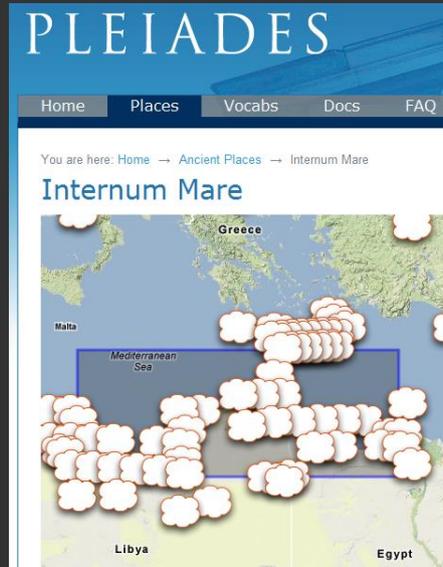
Administrative Unit Ontology

Query API or tabular datasets ?

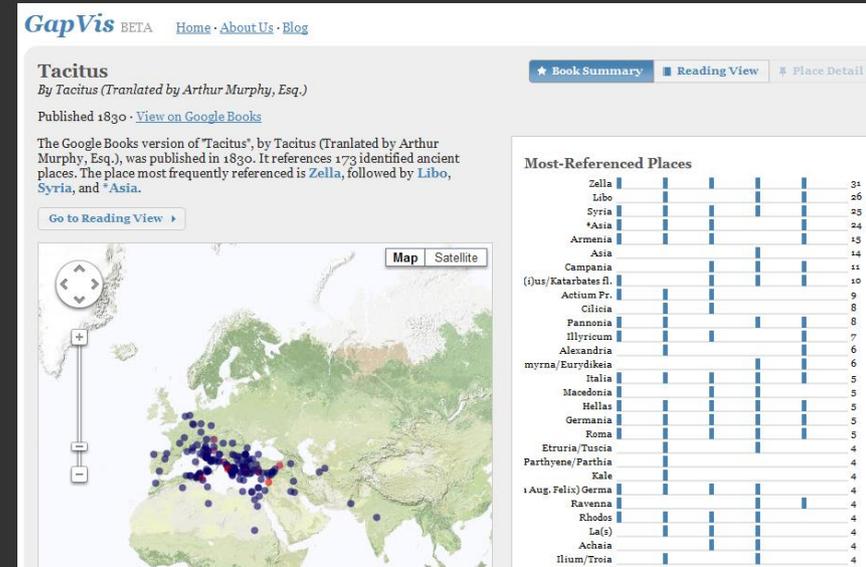
State of the Art - Locations in Classical Texts and Atlases



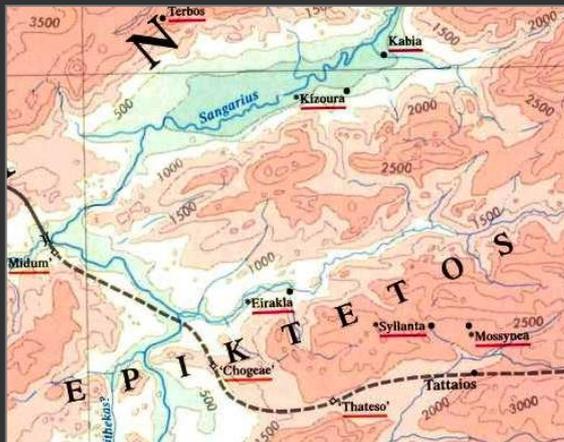
Perseus Project



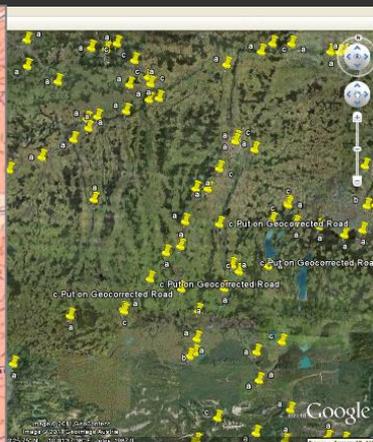
Pleiades



Google Ancient Places



Barrington Atlas



- Hadrianopolis ALB, 54 B2
- Hadrianopolis GRE, 51 C2;
- Hadrianopolis LBY, 38 B1
- Hadrianopolis SYR, 68 F4
- Hadrianopolis TKY, 56 F3
- Hadrianopolis TKY, 65 D3
- Hadrianopolis TKY, 65 D5
- Hadrianopolis TKY, 66 F3
- Hadrianopolis TKY, 86 C3
- Hadrianou Hormos ITL, 45

Digital Atlas of Roman and Medieval Civilization



State of the Art - Web Services for historical placenames



Unlock: geoparsing (place name text mining and mapping)

Chalice: creating linked data historic gazetteer through text mining

DEEP: extend Chalice by digitizing the 86 volumes of the English Place Name Survey

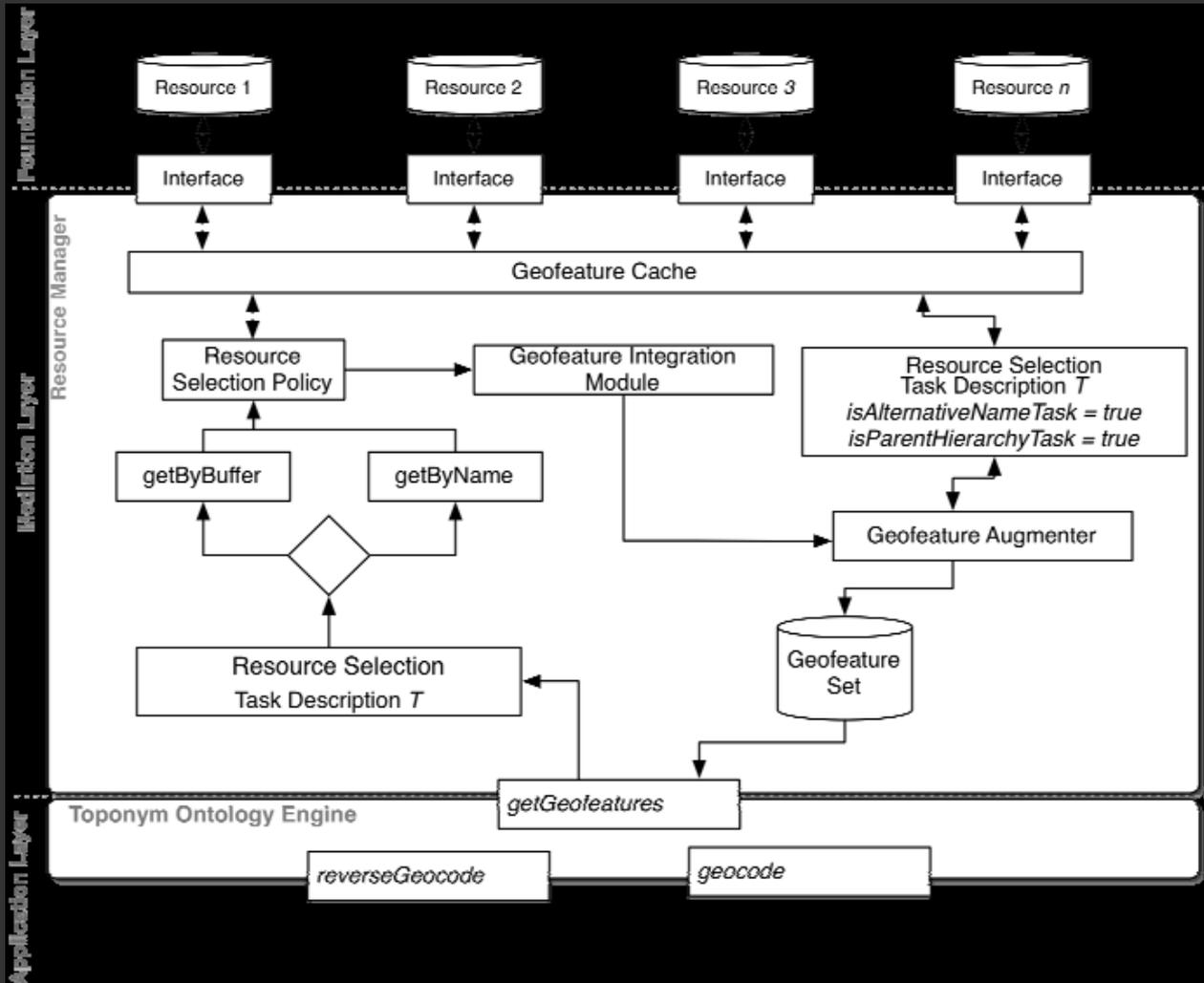


Unlock web service only - UK only

Chalice Schema not published

DEEP plans to develop schema for historical place name queries

Gazetteer Augmentation



Queries directed to different resources

Results augment existing records

Geofeature Set

Sample data for testbed: CHGIS

Historical Place Details:

- ▶ **name:** Tengyue Ting
- ▶ **vernacular:** 腾越厅
- ▶ **alternate:** 腾越廳
- ▶ **feature type:** independent sub-prefecture 直隶厅 Zhiliting
- ▶ **date range:** 1820 to 1820
- ▶ **present location:** 云南腾冲县

▶ admin hierarchy SUPERIOR units:

from 1820 to 1820 Tengyue Ting was part of: [云南 Yunnan](#)

▶ admin hierarchy SUBORDINATE units

- ▶ **coordinate type:** point location
- ▶ **latitude:** 25.024349 N
- ▶ **longitude:** 98.49498 E

[Google 地图](#) - [get Google Map](#) - [get CHGIS Map](#) -

(left ◀ 91.99498 E) (top ▲ 31.52435 N)



- ⊗ Historical Place Location
- ★ Prefecture Seats 1820
- ∩ Qing Provinces
- Qing Prefectures

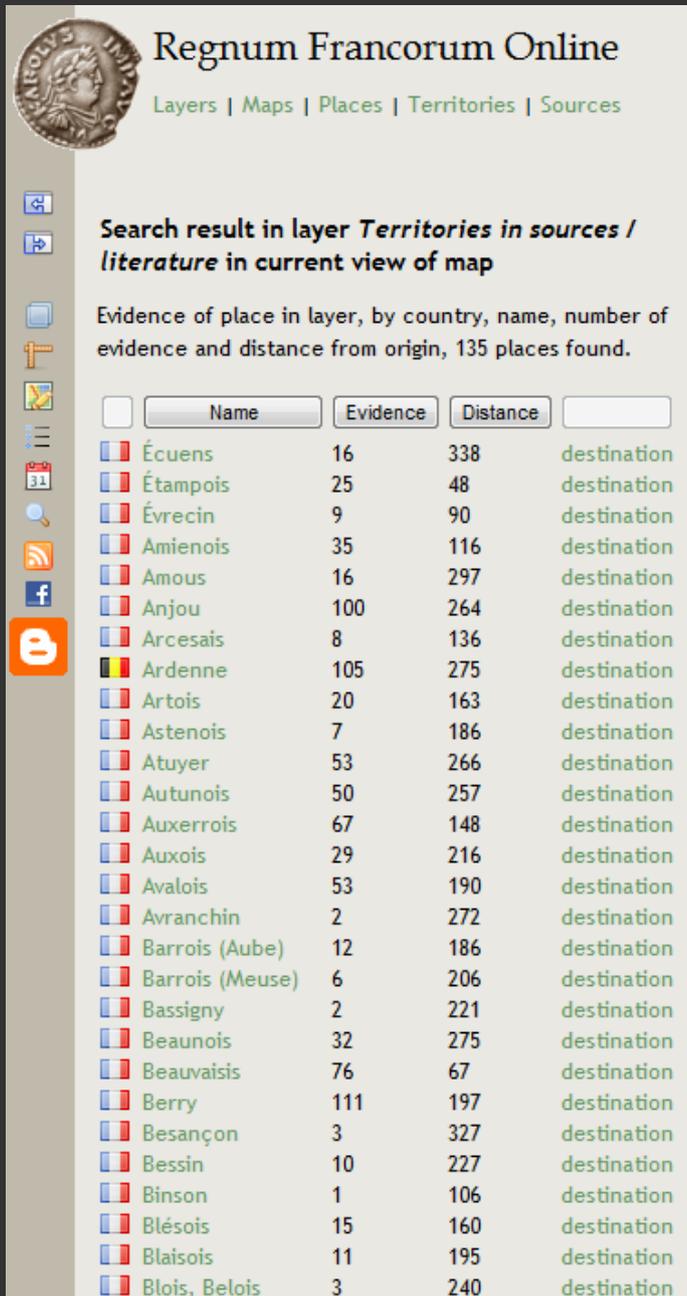
Historical placename:
Tengyue Ting

Present placename:
云南腾冲县

Romanized present placename:
yunnantengchongxian

Time series county seats:
403 placename records

Sample data for testbed: Regnum Francorum, France



Regnum Francorum Online
Layers | Maps | Places | Territories | Sources

Search result in layer *Territories in sources / literature* in current view of map

Evidence of place in layer, by country, name, number of evidence and distance from origin, 135 places found.

Name	Evidence	Distance	
Écuens	16	338	destination
Étampuis	25	48	destination
Évrecin	9	90	destination
Amienois	35	116	destination
Amous	16	297	destination
Anjou	100	264	destination
Arcesais	8	136	destination
Ardenne	105	275	destination
Artois	20	163	destination
Astenois	7	186	destination
Atuyer	53	266	destination
Autunois	50	257	destination
Auxerrois	67	148	destination
Auxois	29	216	destination
Avalois	53	190	destination
Avranchin	2	272	destination
Barrois (Aube)	12	186	destination
Barrois (Meuse)	6	206	destination
Bassigny	2	221	destination
Beaunois	32	275	destination
Beauvaisis	76	67	destination
Berry	111	197	destination
Besançon	3	327	destination
Bessin	10	227	destination
Binson	1	106	destination
Blésois	15	160	destination
Blaisois	11	195	destination
Blois, Belois	3	240	destination

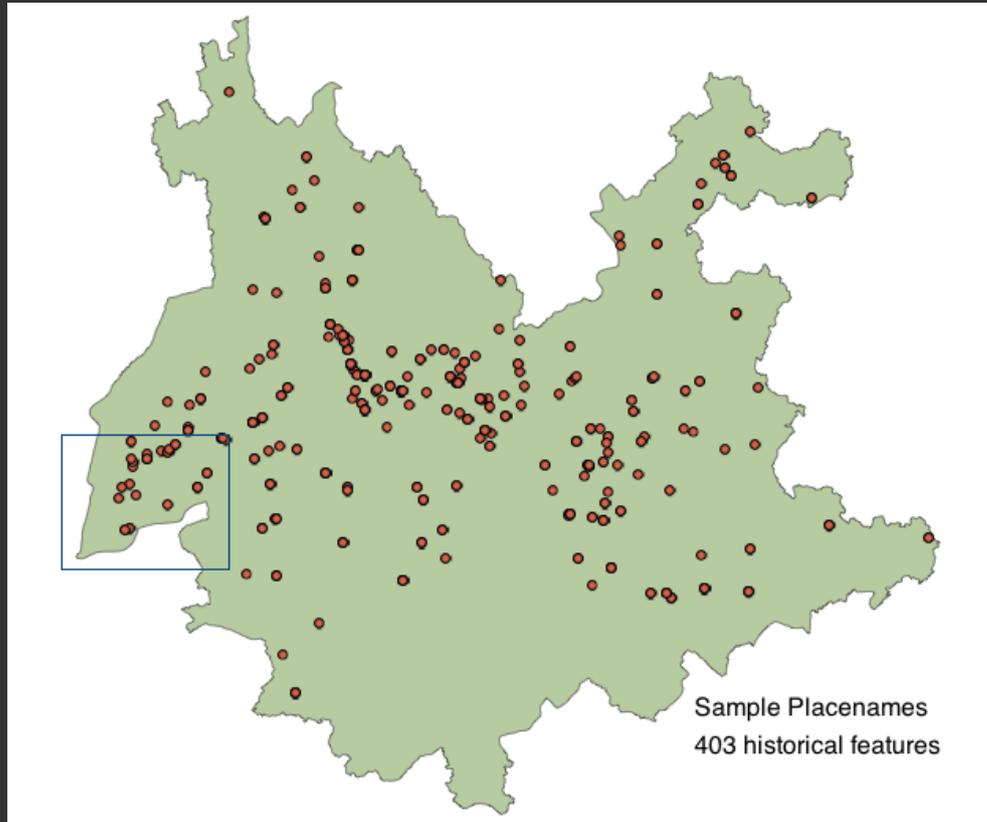
Historical placename:
Éclance

GeoNames ID:
3020485

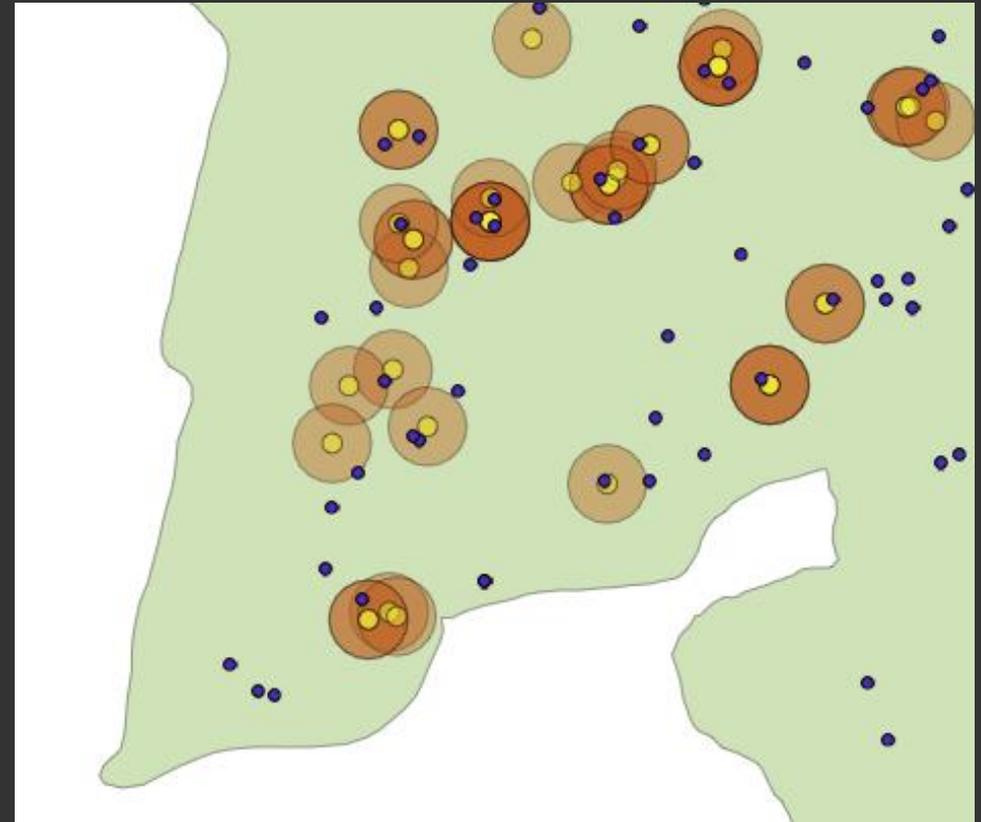
France places in RFO:
4164 placename records

num	oid	oname	lng	lat	insee	geon	
1	11051	Ébreuil		3.0882	46.1161	3107	3020569
2	7856	Échemines		3.8315	48.3886	10134	3020517
3	8958	Éclance		4.6354	48.3058	10135	3020485
4	8507	Écoivres		2.6781305	50.3445124	62283	3020470
5	3664	Écommoy		0.2757986	47.8300763	72124	3020462
6	2351	Écouen		2.3791247	49.0184959	95205	3020440

GIS sampling of CHGIS names and GeoNames

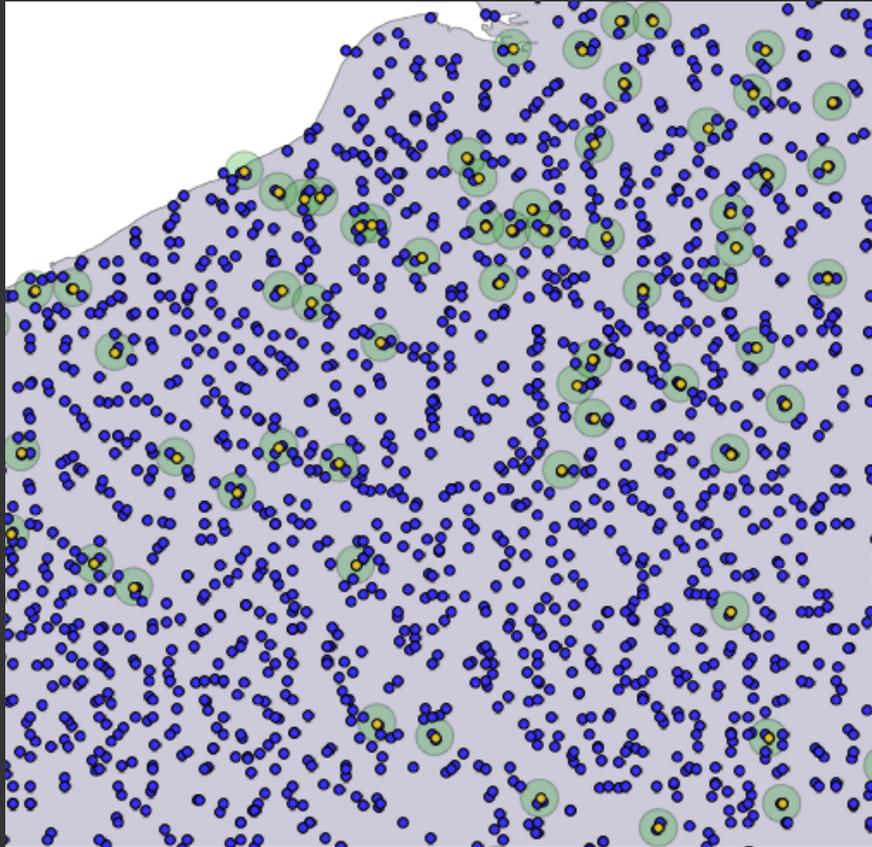


Distribution of study sample

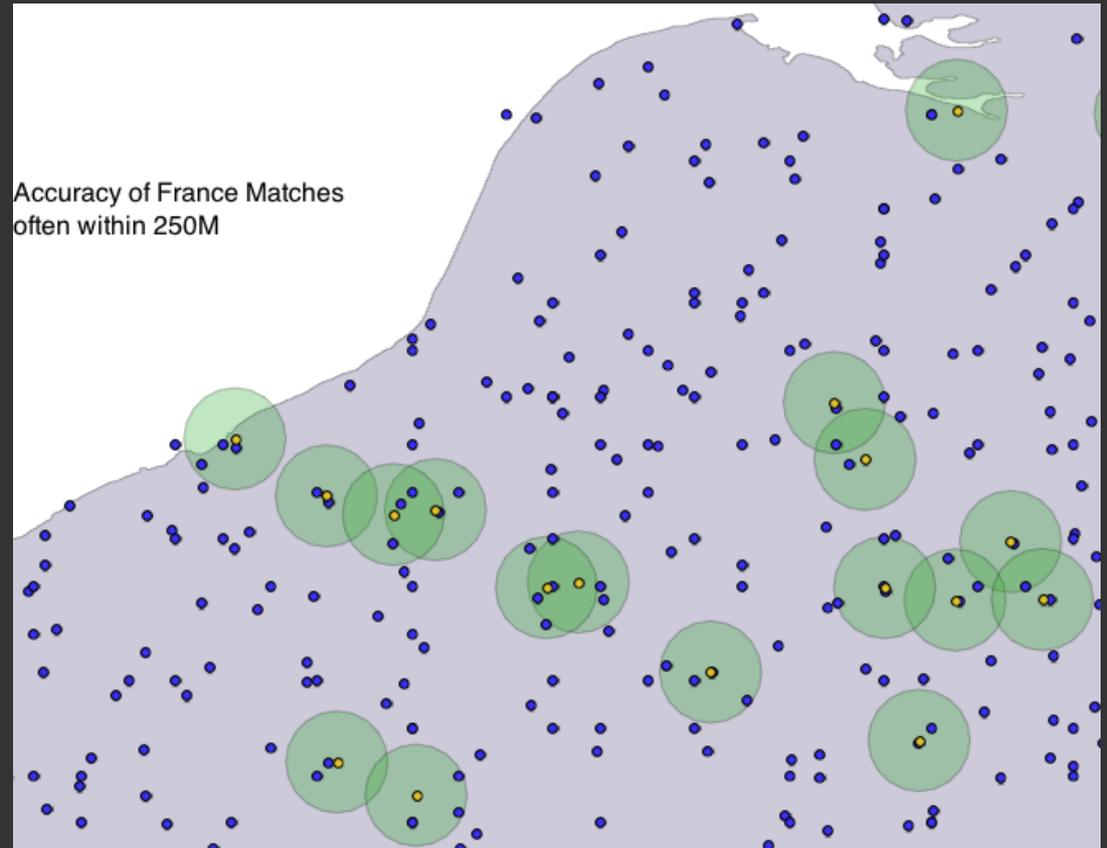


2km and 8km buffers + GeoNames

GIS sampling of France names and GeoNames

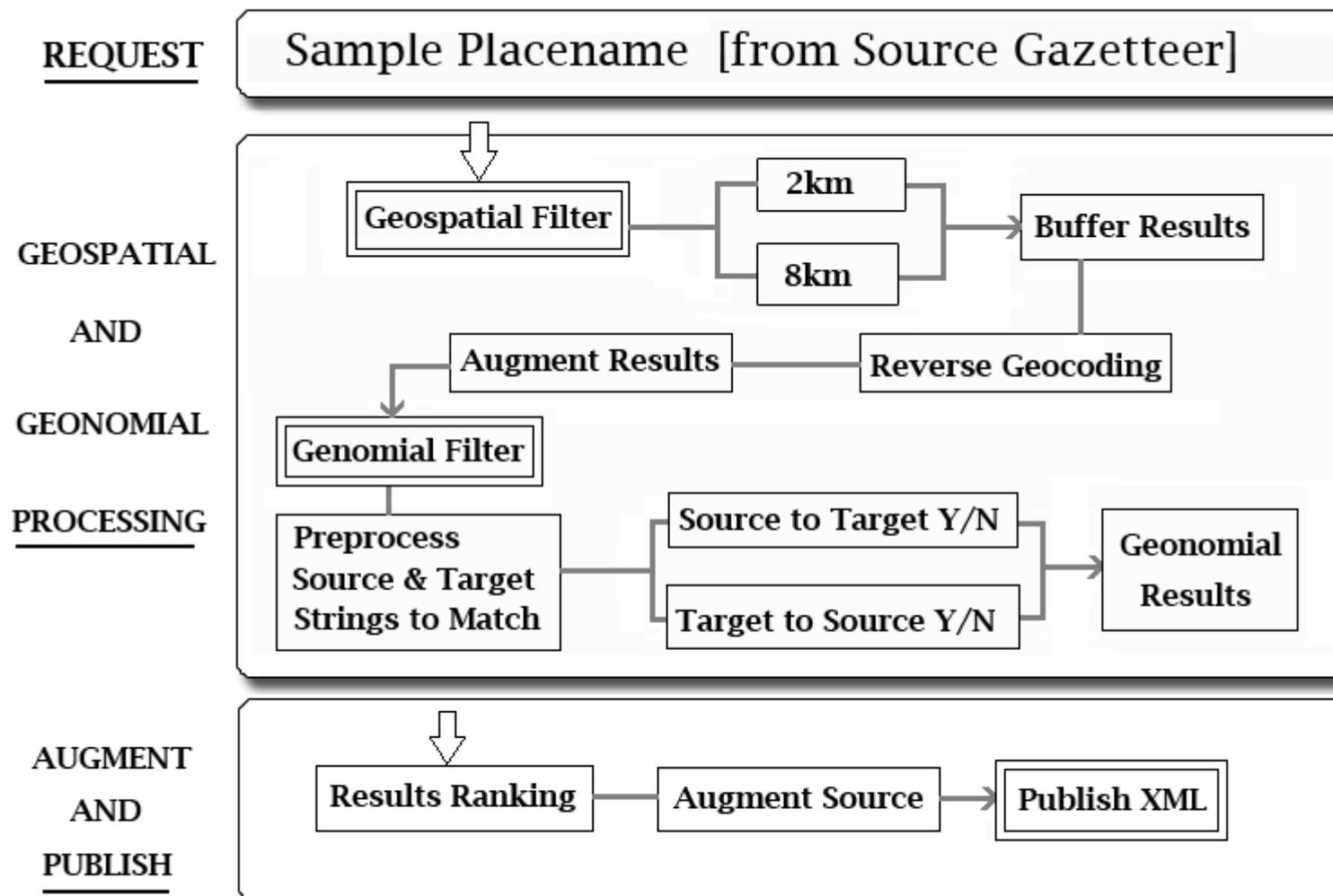


Distribution of study sample



2km buffers + GeoNames

Matching process



Why trim to five characters? Why match in both directions?

"Henansheng Puyang Shi Puyang Xian Qinghetou Xiang" to "Puyang"

After special chars trim:

"HenanshengPuyangShiPuyangXianQinghetouXiang" to "Puyang"

After first five chars trim:

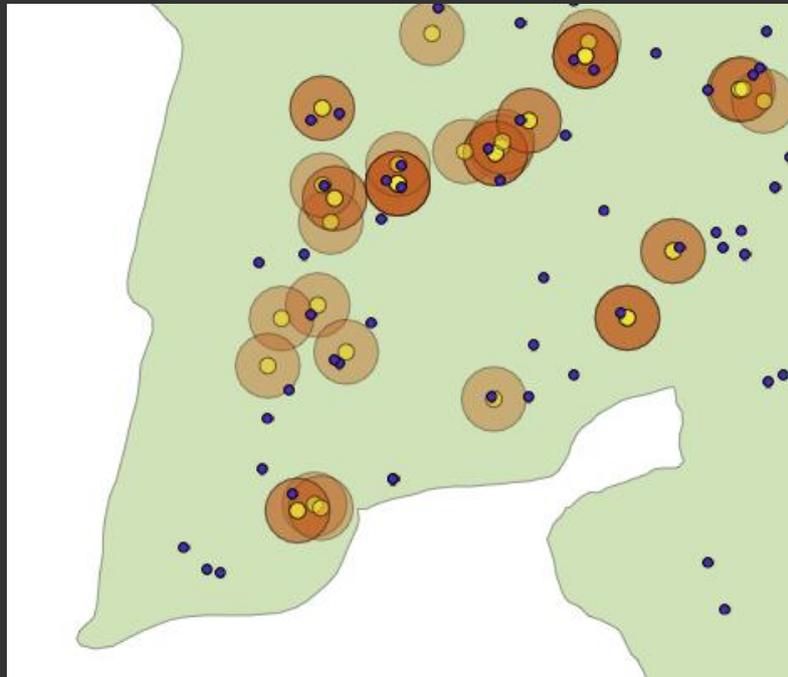
Source to Target:

"Henan" to "Puyang" **no match**

"Puyan" to "HenanshengPuyangShiPuyangXianQinghetouXiang" **match**

Results based on distance filtering

Sample	Total Names	2km Hits	Avg < 2km	8km Hits	Avg <8km	Avg All
Yunnan	403	297	1.15 km	1994	4.67 km	6.07 km
France	4164	17705	0.84 km	n.a.	n. a.	1.02 km



Results based on string matching within buffer distances

Sample	Total Names	One-way match	Two-way match	Source to Target	Target to Source
France	4164	97%	86%	97%	192%
Yunnan	403	74%	18%	27%	87%

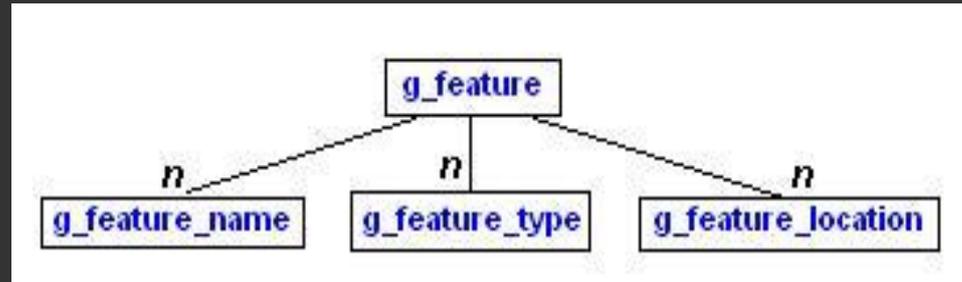
Main discovery:

Historical placenames not often found in GeoNames alternates

Much more likely for current GeoName to be found in historical gazetteer record when “present location” is attested

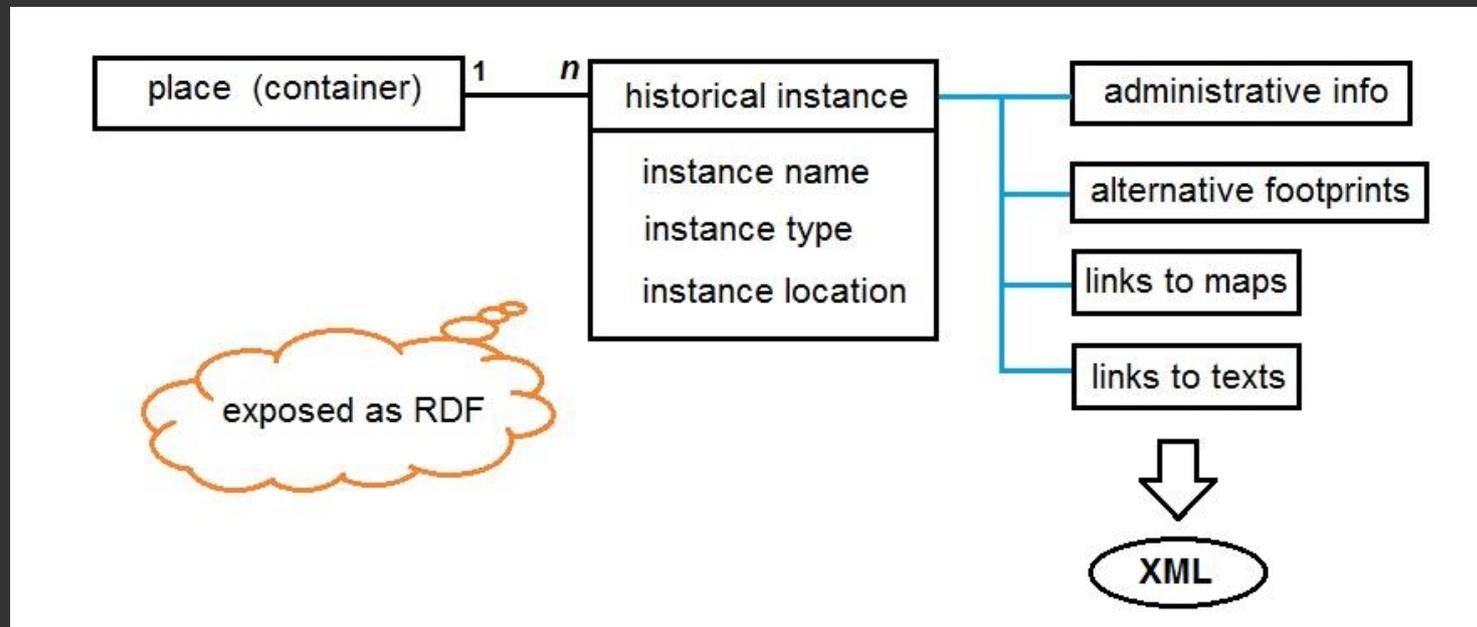
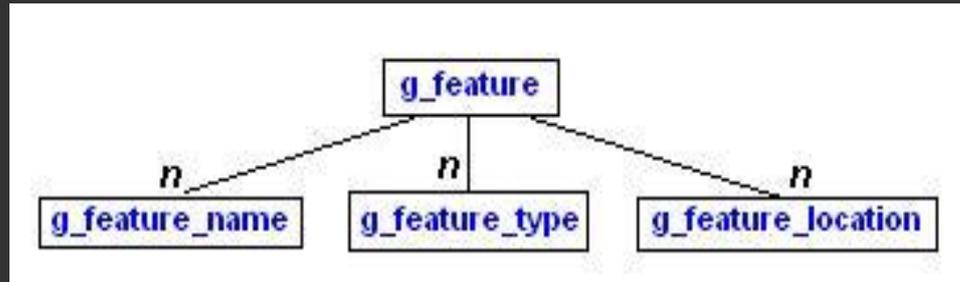
Schema for the augmented gazetteer

Linda Hill - ADL



Schema for the augmented gazetteer

Linda Hill - ADL



Augmented gazetteer for faceted geo-temporal queries

- **Placenames**
- **Dates of existence (or related time periods)**
- **Administrative jurisdictions (past and present)**
- **Alternative footprints**

Existing XML schema for historical placenames (for CHGIS)

```
<query>
```

```
  <results>
```

```
    <item>
```

```
      <placename>
```

```
      <feature_type>
```

```
      <temporal>
```

```
      <spatial>
```

```
      <part_of>
```

```
      <evidenced_by>
```

```
      <preceded_by>
```

```
      <links>
```

Old schema - reorganized for multiple placename attestations

```
<placename>  
  <name_romanized>Fu Zhou  
  <name_vernacular>抚州  
  <name_alterate>撫州  
<temporal>  
  <begin_year>1161  
  <begin_year_rule>4  
  <end_year>1276  
  <end_year_rule>4  
<part_of>  
  <parent_name>江南西路  
  <part_of_begin>1134  
  <part_of_end>1276
```

New schema - need to allow for multiple placename attestations

```
<placename>  
  <name>撫州  
<lang>zh  
<lang_class>traditional Chinese  
  <variant>N  
  <begin>1161  
  <begin_attestation>jurisdiction area increased  
  <end>1276  
  <end_attestation>placename changed  
  
  <parent_address>江南西路  
  <parent_begin>1134  
  <parent_end>1276  
  
  <source>CHGIS  
  <verified>Y  
  <name_rank>default  
  <counterclaim>N
```

New schema - need to allow for multiple placename attestations

```
<placename>  
  <name>Linchuan  
<lang>zh  
<lang_class>transliteration  
  <variant>N  
  <begin>20120212  
  <begin_attestation>exists  
  <end>20120212  
  <end_attestation>exists  
  
  <parent_address>Linchuan, Fuzhou, Jiangxi, China  
  <parent_begin>20120212  
  <parent_end>20120212  
  
  <source>Google  
  <verified>N  
  <name_rank>default  
  <counterclaim>N
```

Publications & Resources

<http://fas.harvard.edu/~chgis/gazetteer>