



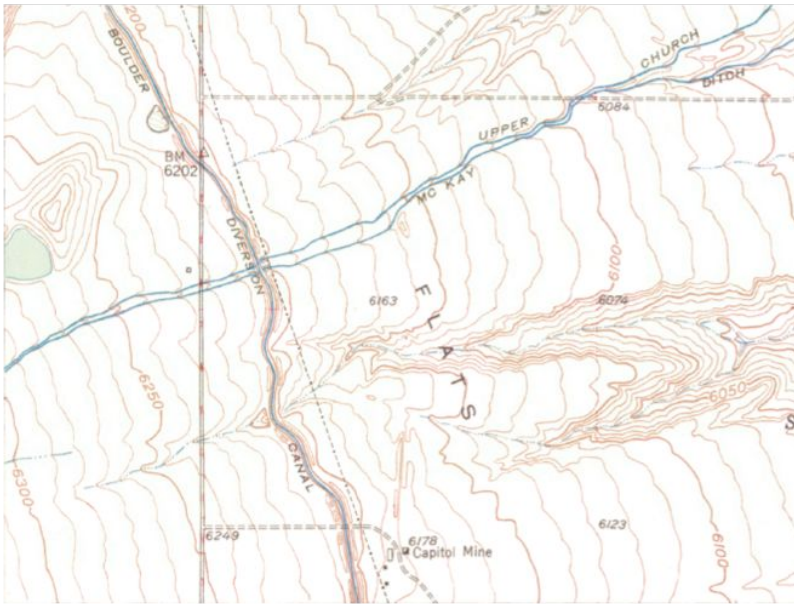
Building Linked Data from Historical Maps

*Chun Lin, Hang Su,
Craig A. Knoblock,
Yao-Yi Chiang, Weiwei Duan,
Stefan Leyk and Johannes H. Uhl*

Center on Knowledge Graphs, USC/ISI



PROBLEM: RAILROADS APPEARANCE



1950

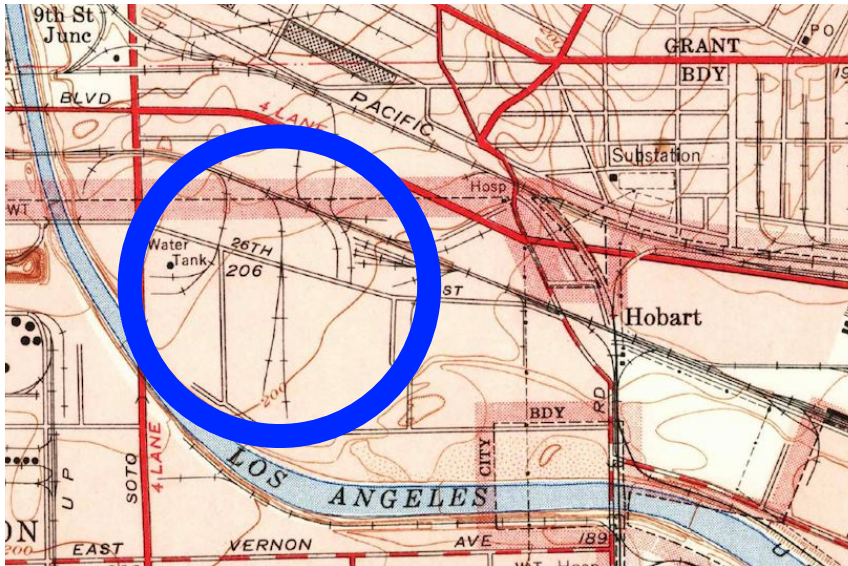


1965

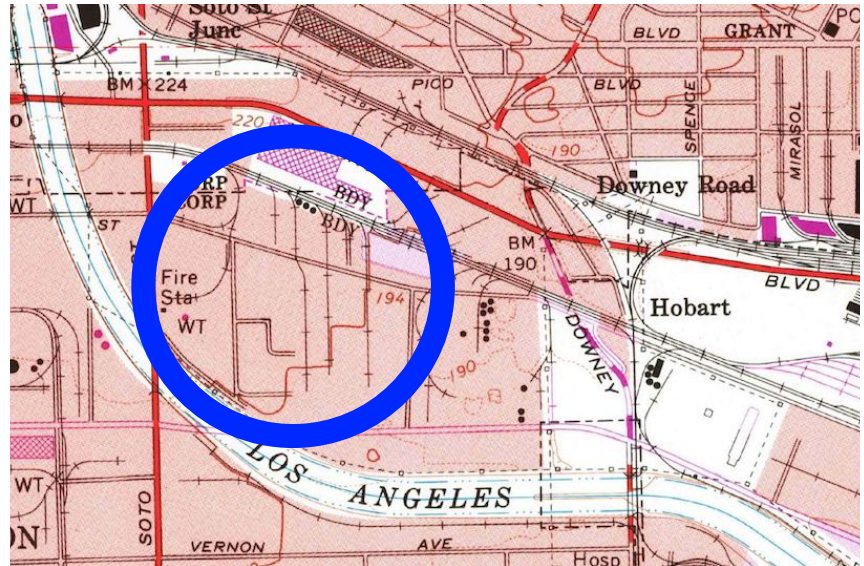
Railroads appearance in Louisville, Colorado



PROBLEM: RAILROADS CHANGE



1953



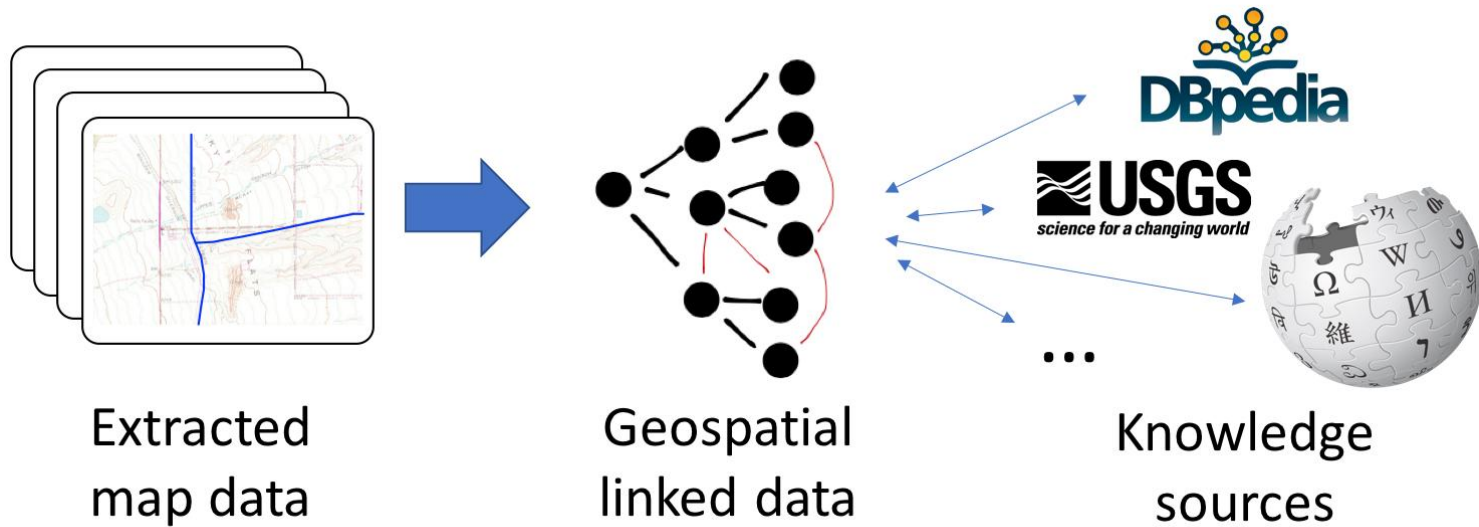
1966

Railroads change in Los Angeles, California



GOALS

- Geospatial change analysis
- Utilize other knowledge sources
- Enable semantic rich queries





TECHNICAL CHALLENGES

- How to determine same and different parts across maps?
- How to incrementally integrate new data?
- How to organize the data for change analysis queries?
- How to represent geospatial data in RDF?
 - Geometry representation
 - Assign unique URI



PIPELINE

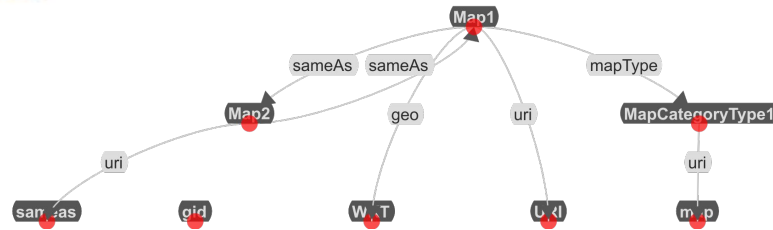
Linking

Creating Linked Data

Querying

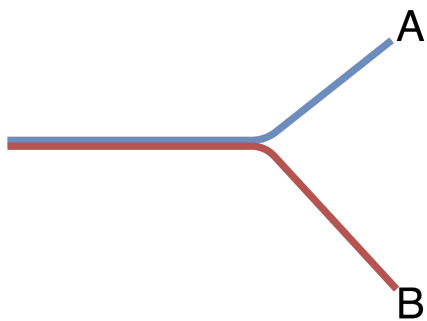


Karma

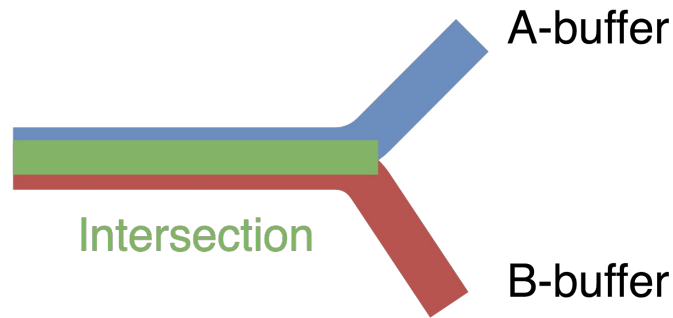




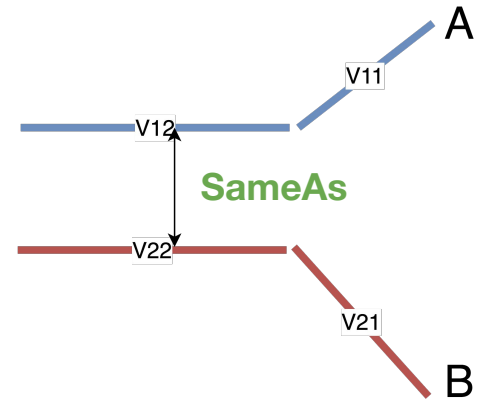
LINKING: LINE SEGMENTATION



Map **A** and **B** have common and distinct segments



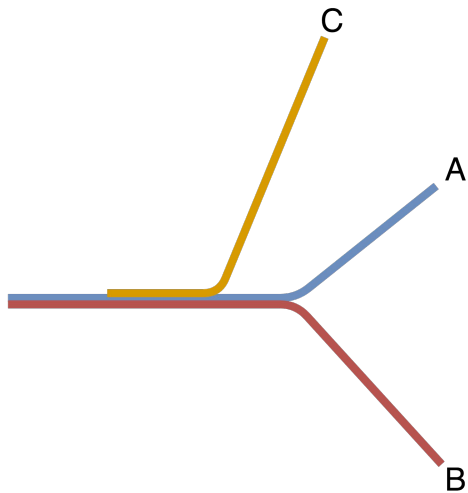
Buffer out and find the **common parts**



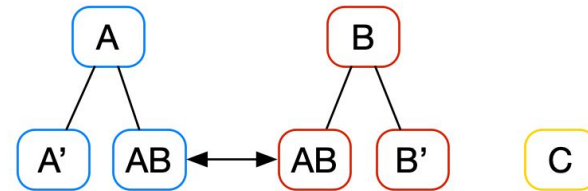
Partition line segments and mark **"SameAs"**



LINKING: “CONTAINS” RELATIONSHIP



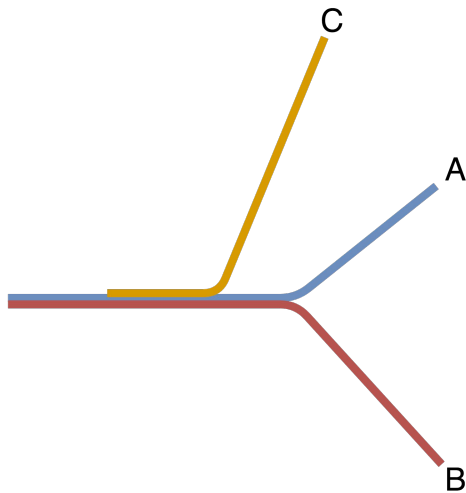
3rd map comes in...



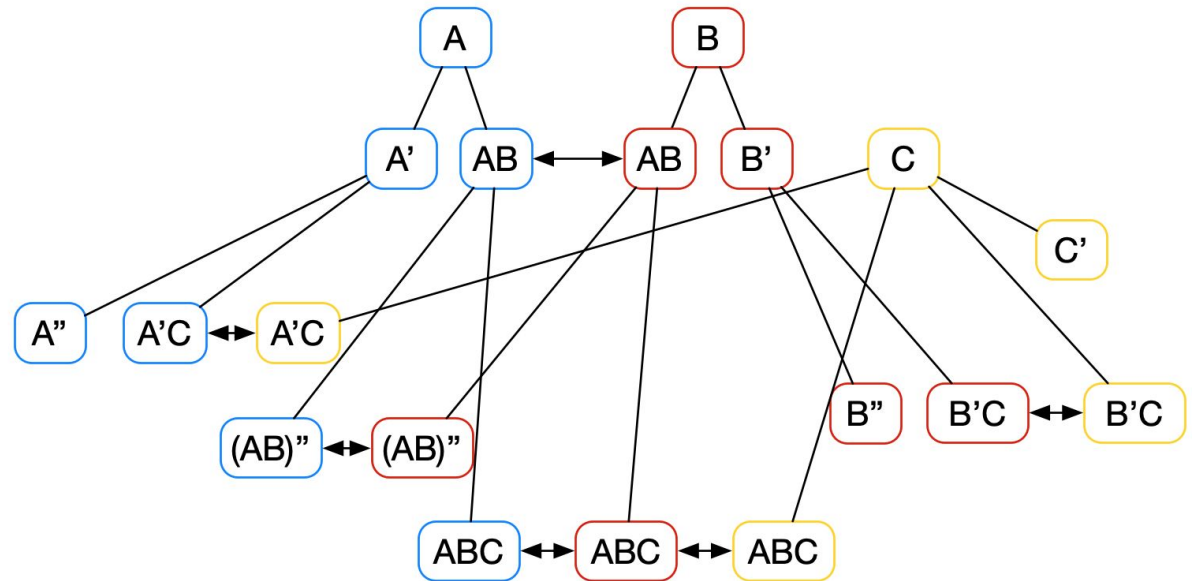
“Contains” relationship tree



LINKING: “CONTAINS” RELATIONSHIP



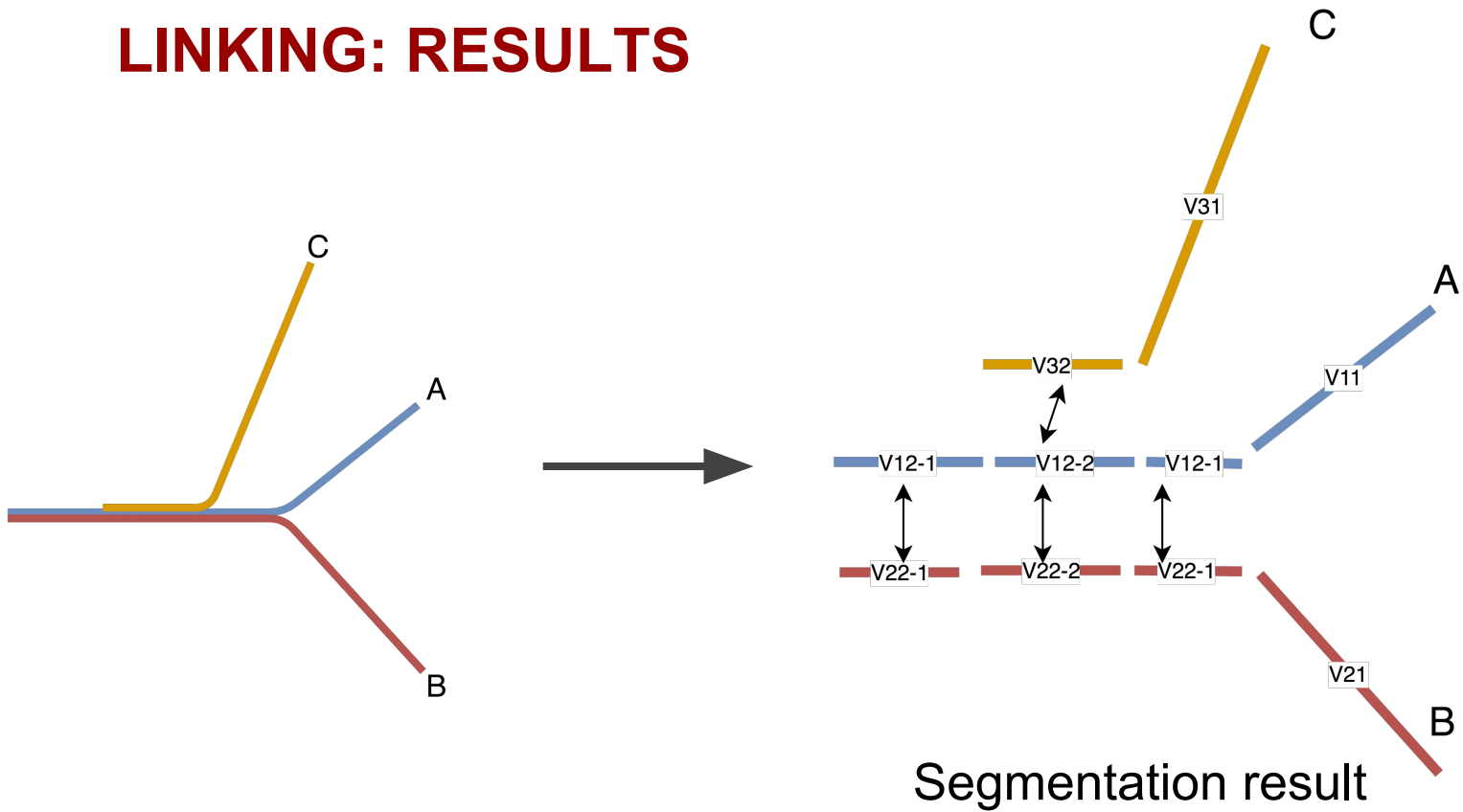
3rd map comes in...



“Contains” relationship tree



LINKING: RESULTS





CREATING THE LINKED DATA: PREPARATION

WKT

WKB
0105000020AD100000510200000102 0000000E000000DD1FA6140F8A5D C0080000000000414048D576....



WKT
MULTILINESTRING((-118.157170450431 34.00000000000001,-118.15711628537 34.0001117586834,-118.157055418703 34.0002624753498,-118.15701194162 34.0004160920163,.....))

URI

WKT	MULTILINESTRING((-118.157170450431 34.00000000000001,-118.15711628537 34.0001117586834,.....))
Map Source	USGS vector data for Los Angeles, CA



“45bf12bd”



RESULTS IN TABLES

Map vectors

URI	WKT	Map
V12-2	LineString...	A
V22-2	LineString...	B
V32	LineString...	C
...

“SameAs”

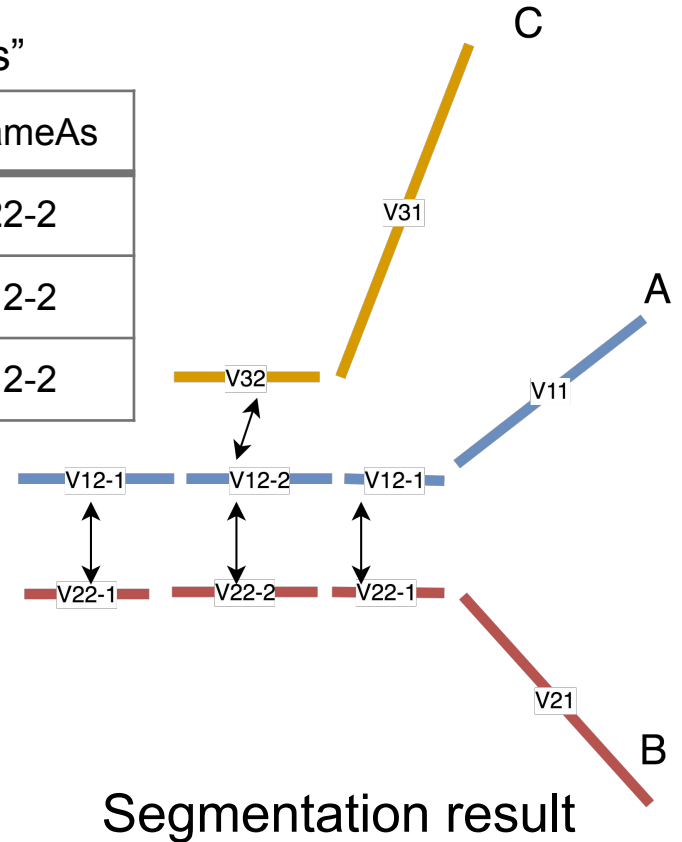
URI	SameAs
V12-2	V22-2
V22-2	V12-2
V32	V12-2

“Contains”

URI	Contains
A	V12
B	V22
V12	V12-1
...	...

Map metadata

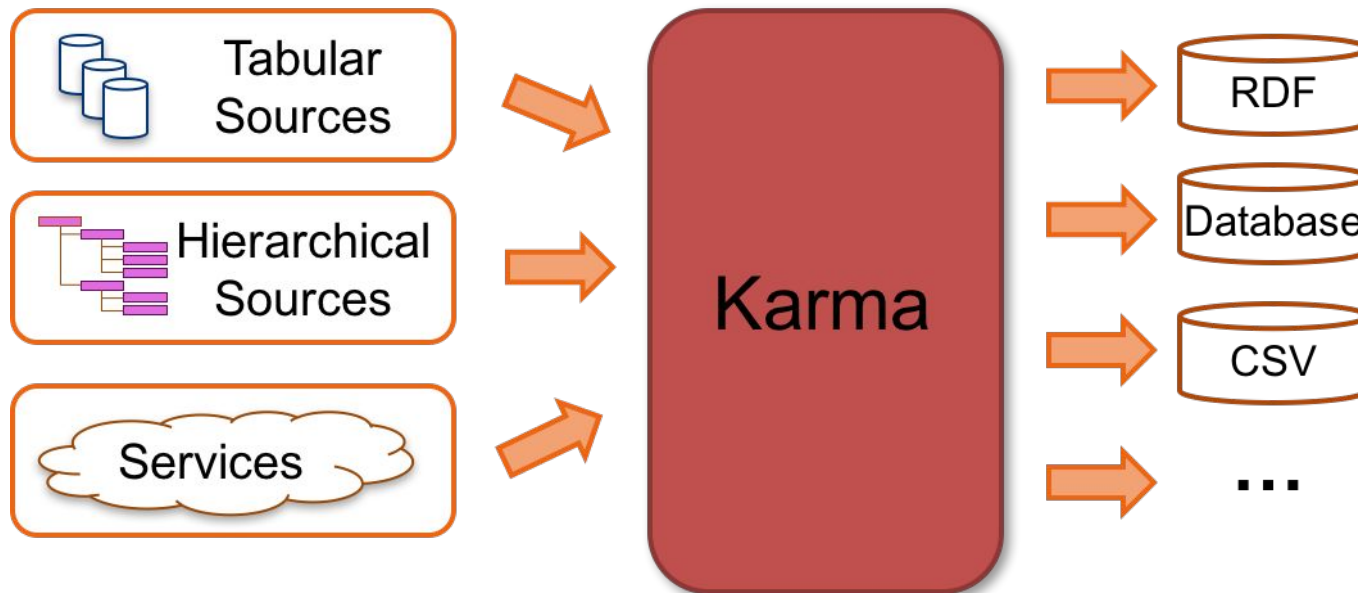
Map	Year
A	2000
B	2003
C	2005





KARMA: A DATA INTEGRATION TOOL

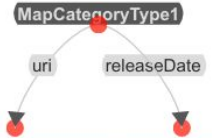
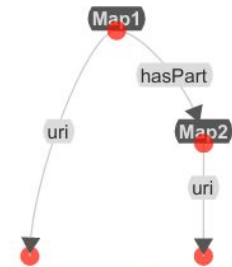
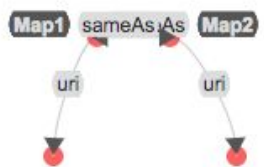
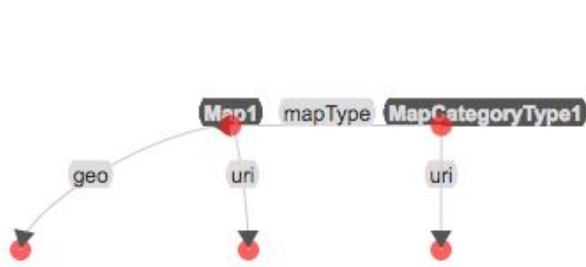
Interactive tool for rapidly extracting, cleaning, transforming, integrating and publishing data



Slide from Pedro Szekely & Craig A. Knoblock



RDF MAPPING



WKT	URI	map
MULTILINESTRING ((-118.125 34.007013813651... 34.0070680878393),	6ef3ccc0	usgs_trans
MULTILINESTRING ((-118.157170450... 34.0,-118.157116... 34.000111758683...	45bf12bd	usgs_trans

Map vectors

URI	sameAs
6ef3ccc0	264be5c9
2577ce5a	794c5e1e
x19d3c9c7	x798f2cd6

“SameAs”

id	contains
6ef3ccc0	2577ce5a
6ef3ccc0	x5c406ca0
45bf12bd	x19d3c9c7
45bf12bd	7c8bedd4

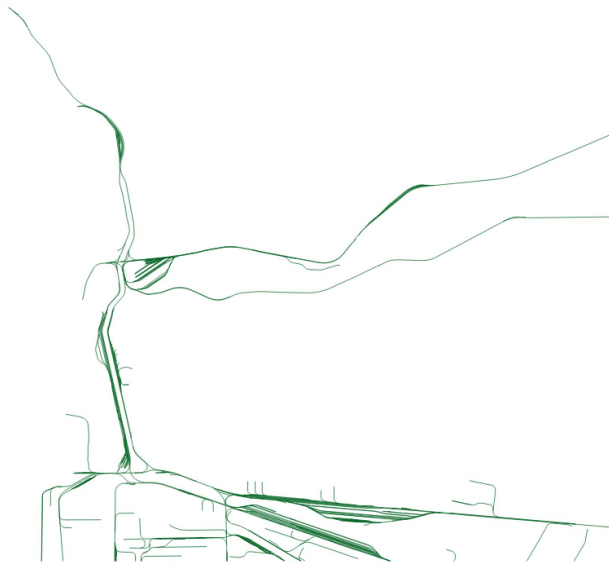
“Contains”

mapid	year
clip_rail13	2003
clip_usa	2005
usgs_trans	2000

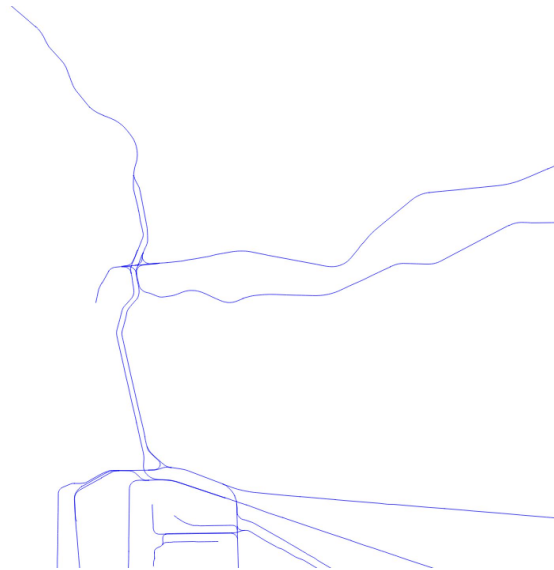
Map metadata



ALTERNATIVE RAILROAD MAPS FOR LA



USGS vector data for
Los Angeles, California
(373 KB)



California Rail Network
(183 KB)



National Atlas of the
United States
(45 KB)



QUERYING

What railroads are in 2000 but not in 2005?

```
PREFIX schema: <http://schema.org/>
select distinct ?a ?mapa
where {?a schema:geo ?geo.
      ?a schema:mapType ?mapa.
      ?mapa schema:releaseDate "2000".
      filter not exists{
        ?a schema:sameAs ?b.
        ?b schema:mapType ?mapb.
        ?mapb schema:releaseDate "2005".}
      minus{?a schema:contains ?x}}
```



■ Common ■ Difference



RELATED WORK

- Linking process
 - Map vector data conflation [Ruiz et al. 2011]
 - Feature matching between maps
 - No segmentation on feature vectors
 - Similarity measures of vector data [Sherif et al. 2015]
 - Detailed measurement for “SameAs” with point set
 - Trade-off: computationally expensive
- Creating the Linked Data
 - Integrating geospatial information using Linked Data [Usery et al. 2012, Sehgal et al. 2006, Yu et al. 2018]
 - Focuses on points of interest data, not vector



DISCUSSION

- Proposing general pipeline from linking to querying
- Integrating map vector data using Linked Data
- Making the data widely available to researchers
- Enabling the ability to answer complex queries



FUTURE WORK

- Scaling up to accommodate large volume of map data
- Optimizing Linking / Querying process
- Pipeline automation