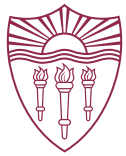


Building Linked Spatio-Temporal Data from Vectorized Historical Maps

April 21st, 2020

Basel Shbita[1], Craig A. Knoblock[1], Weiwei Duan[2], Yao-Yi Chiang[2], Johannes H. Uhl[3], and Stefan Leyk[3]

[1] Information Sciences Institute and Department of Computer Science, USC
[2] Spatial Sciences Institute and Department of Computer Science, USC
[3] Department of Geography, University of Colorado Boulder

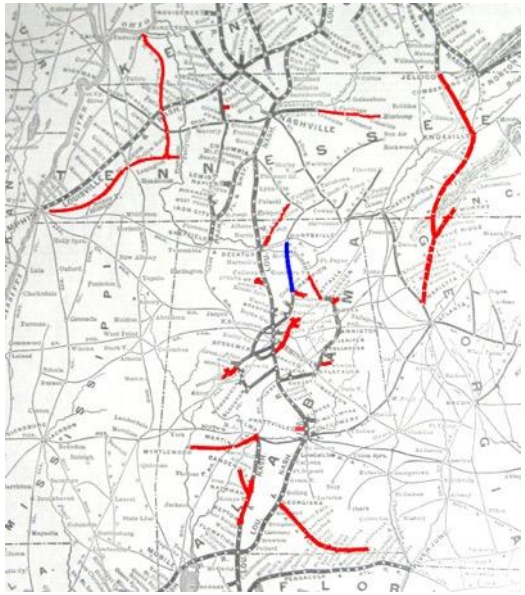
Problem

Digitized Historical Maps
=
Rich sources of information

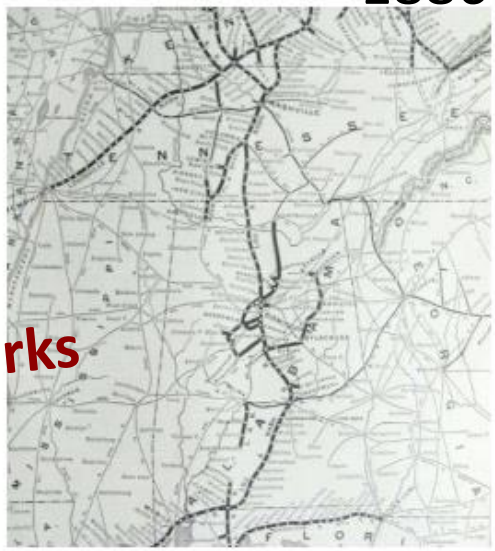
Natural- & Human-made features

Wetlands

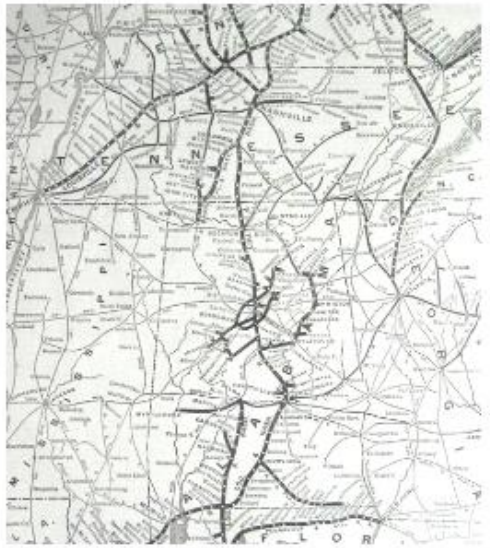
Railroad Networks



1886



1904



Labor-intensive to analyze across time & space

Additional discovery

Idea

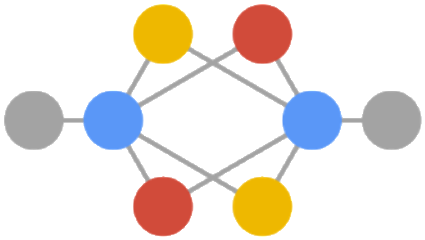
Decompose

to building blocks

then use

Linked Data & the Semantic Web

OpenStreetMap



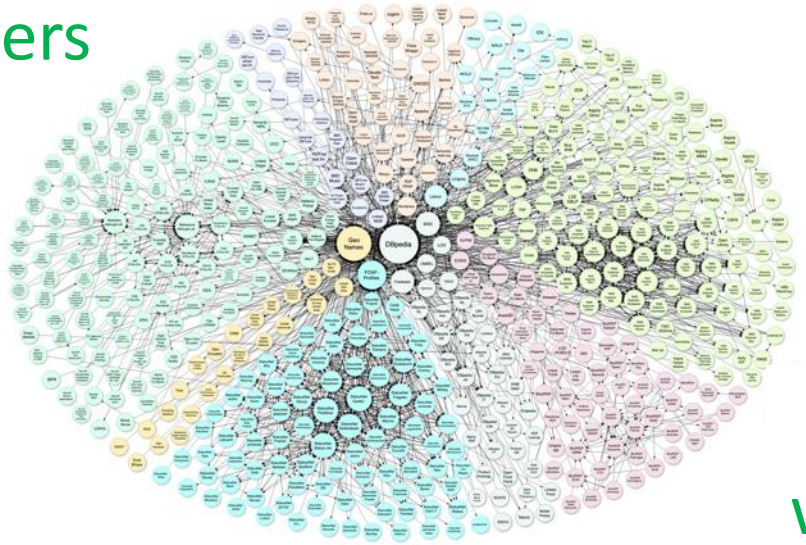
to build a

Knowledge Graph

Why Linked Data?

Break down data barriers

Across-domains



Structured

interpreted by **humans** & **machines**

Semantic

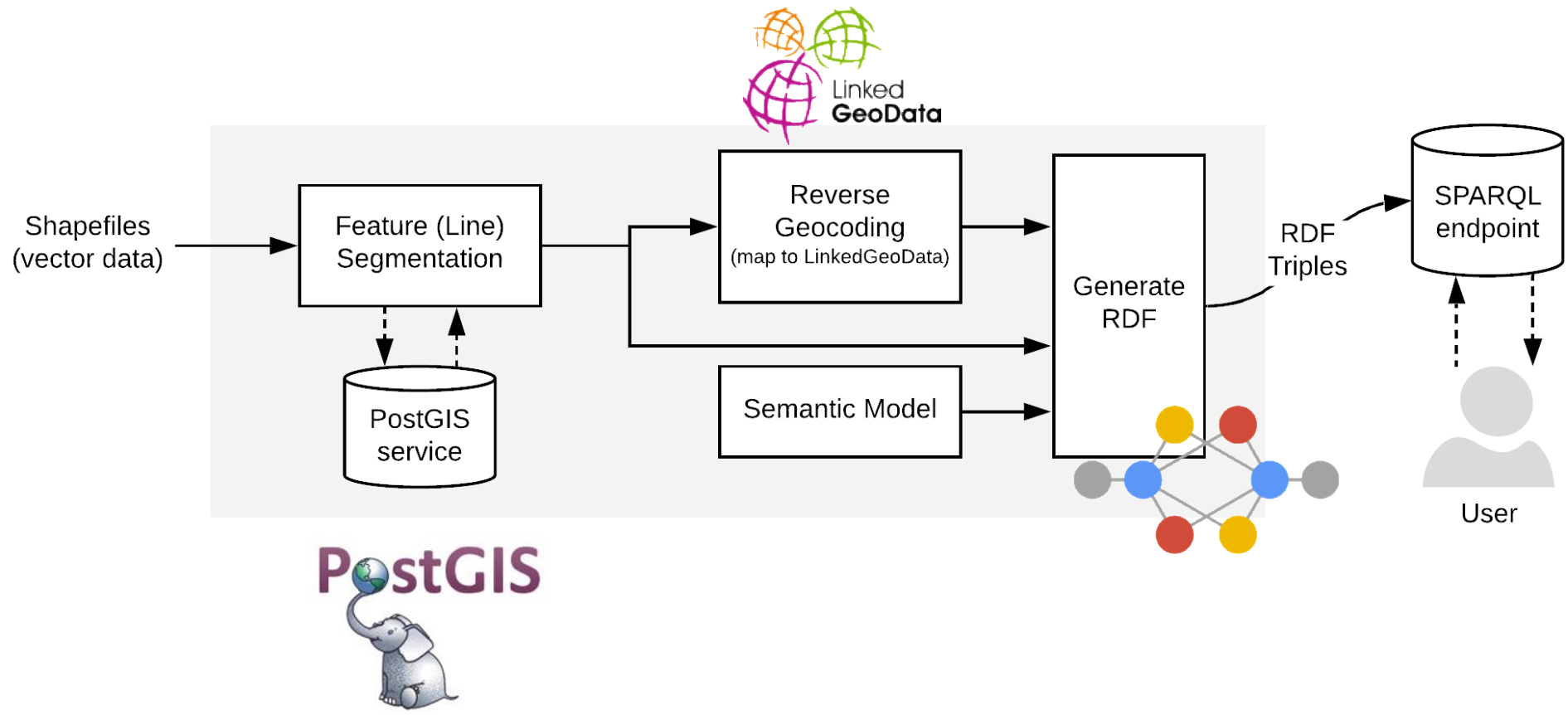
relationships, properties, metadata

Make data widely available

Fuel **Discovery**

Query & **visualize**

Our Approach



Automatic Feature Segmentation

- Goal: create partitions of geo features (segments)
  - Entity matching/linking & entity "partitioning" task
  - "Building Blocks"

- Use a spatially-enabled database service (PostGIS)
  - PostgreSQL extension
    - Manipulate & transform spatial data

- Allow incremental additions over time

Feature Segmentation – cont'd

segments of different map edition

current "building blocks"

foreach $i \in \mathcal{M}$ **do**

  foreach $k \in \mathcal{L}$ **do**

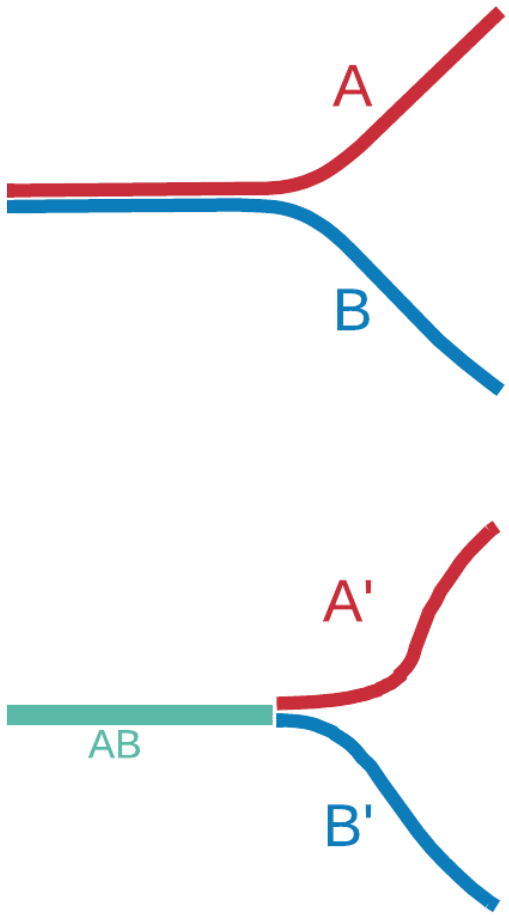
    $\mathcal{F}_\alpha = \mathcal{F}_i \bigcap \mathcal{F}_k;$

    $\mathcal{F}_\gamma = \mathcal{F}_k \setminus \mathcal{F}_\alpha;$

  end

  $\mathcal{F}_\delta = \mathcal{F}_i \setminus (\bigcup_{j \in \mathcal{L}} F_j);$

end



Feature Segmentation – cont'd

segments of different map edition

current "building blocks"

foreach $i \in \mathcal{M}$ **do**

  foreach $k \in \mathcal{L}$ **do**

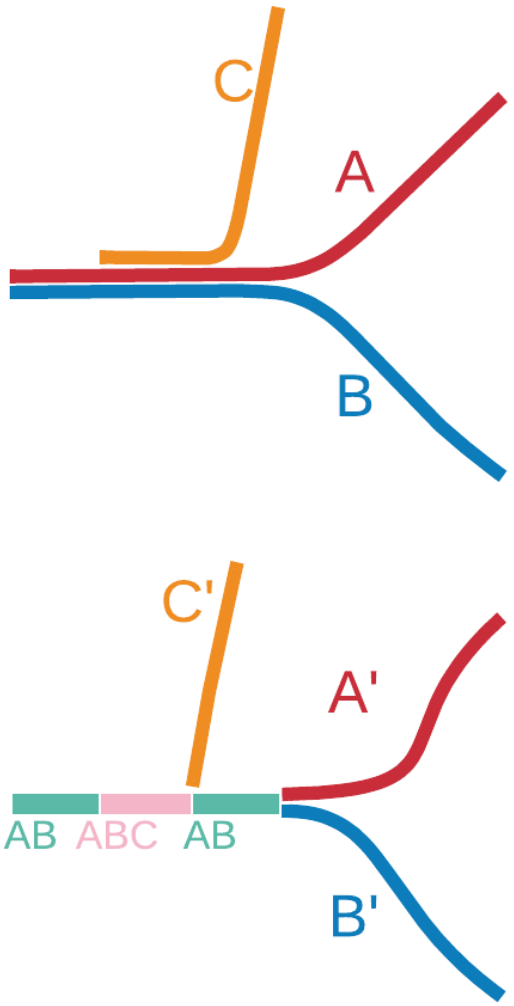
    $\mathcal{F}_\alpha = \mathcal{F}_i \bigcap \mathcal{F}_k;$

    $\mathcal{F}_\gamma = \mathcal{F}_k \setminus \mathcal{F}_\alpha;$

  end

  $\mathcal{F}_\delta = \mathcal{F}_i \setminus (\bigcup_{j \in \mathcal{L}} F_j);$

end



Geo-linking

- Goal: map segments to Linked Open Vocabularies
  - Entity matching
  - Enrich data to fuel discovery

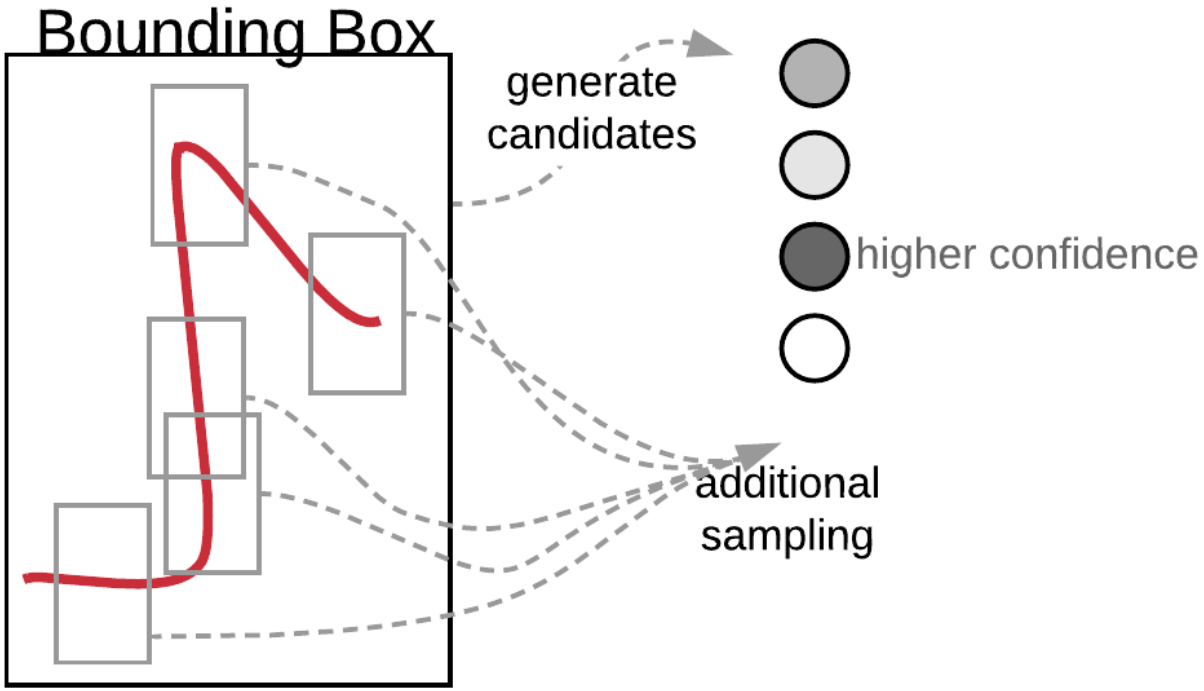
- Use a reverse geocoding service (OpenStreetMap)
  - LinkedGeoData instances

OpenStreetMap

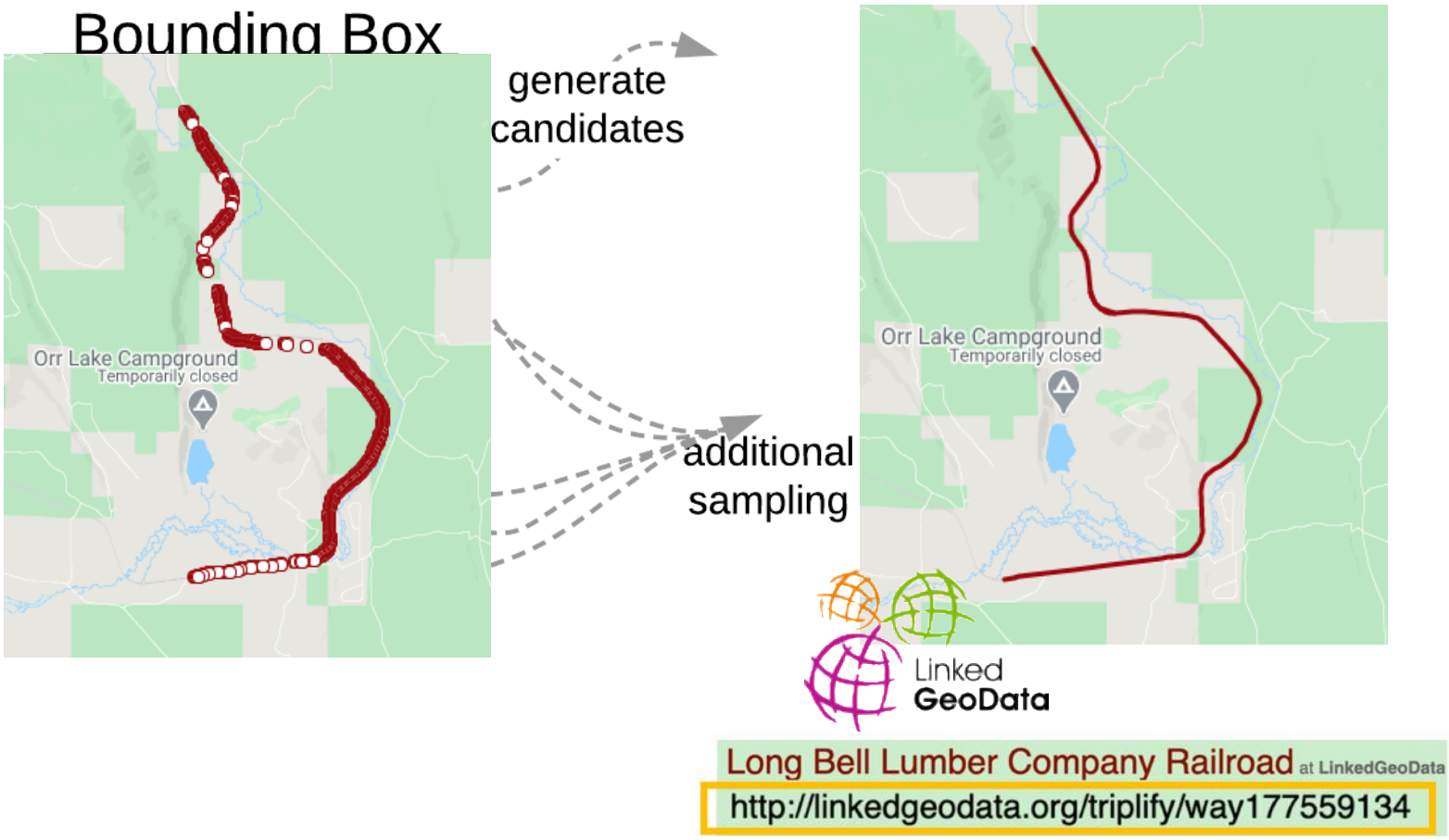
Geo-linking – cont'd

B_s = bounding box wrapping s;
$\mathcal{L}$ = reverse-geocoding(B_s, T);
for 1...N **do**
    e = randomly sample a `Point` in segment s;
    E = reverse-geocoding(e, T);
    $\mathcal{L}$.add(E);
end
filter out instances with a single appearance in $\mathcal{L}$;
return $\mathcal{L}$;

Geo-linking – cont'd



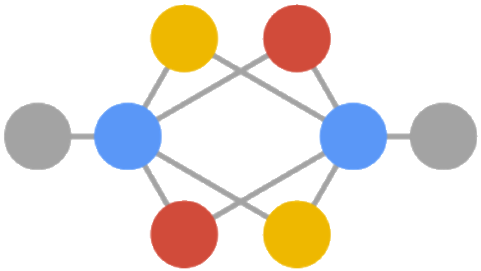
Geo-linking – cont'd



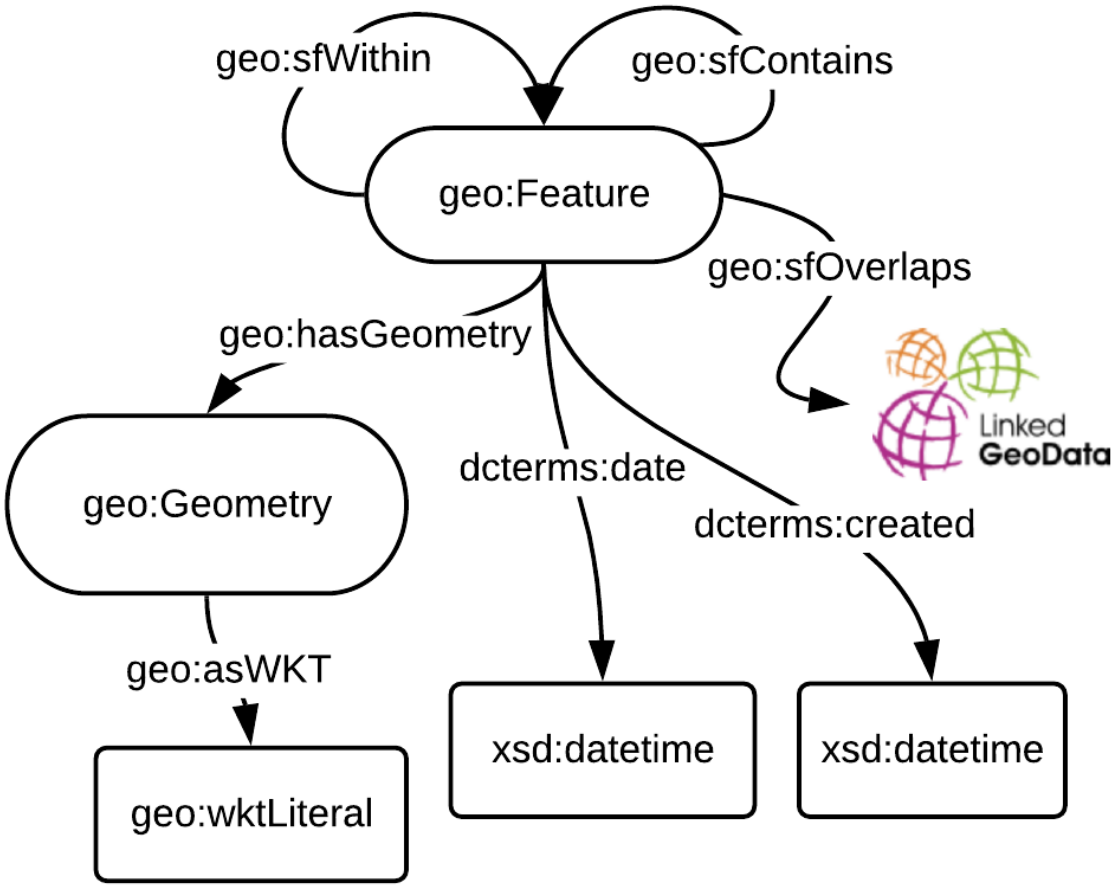
Generate RDF

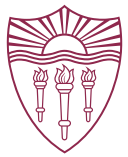
- Goal: construct a KG from the data we collected
  - Useful semantic representation
  - Support downstream **spatial reasoners**

- Construct a meaningful semantic model
  - OGC GeoSPARQL standard
  - Universal conventions
  - Easily queried



Semantic Model

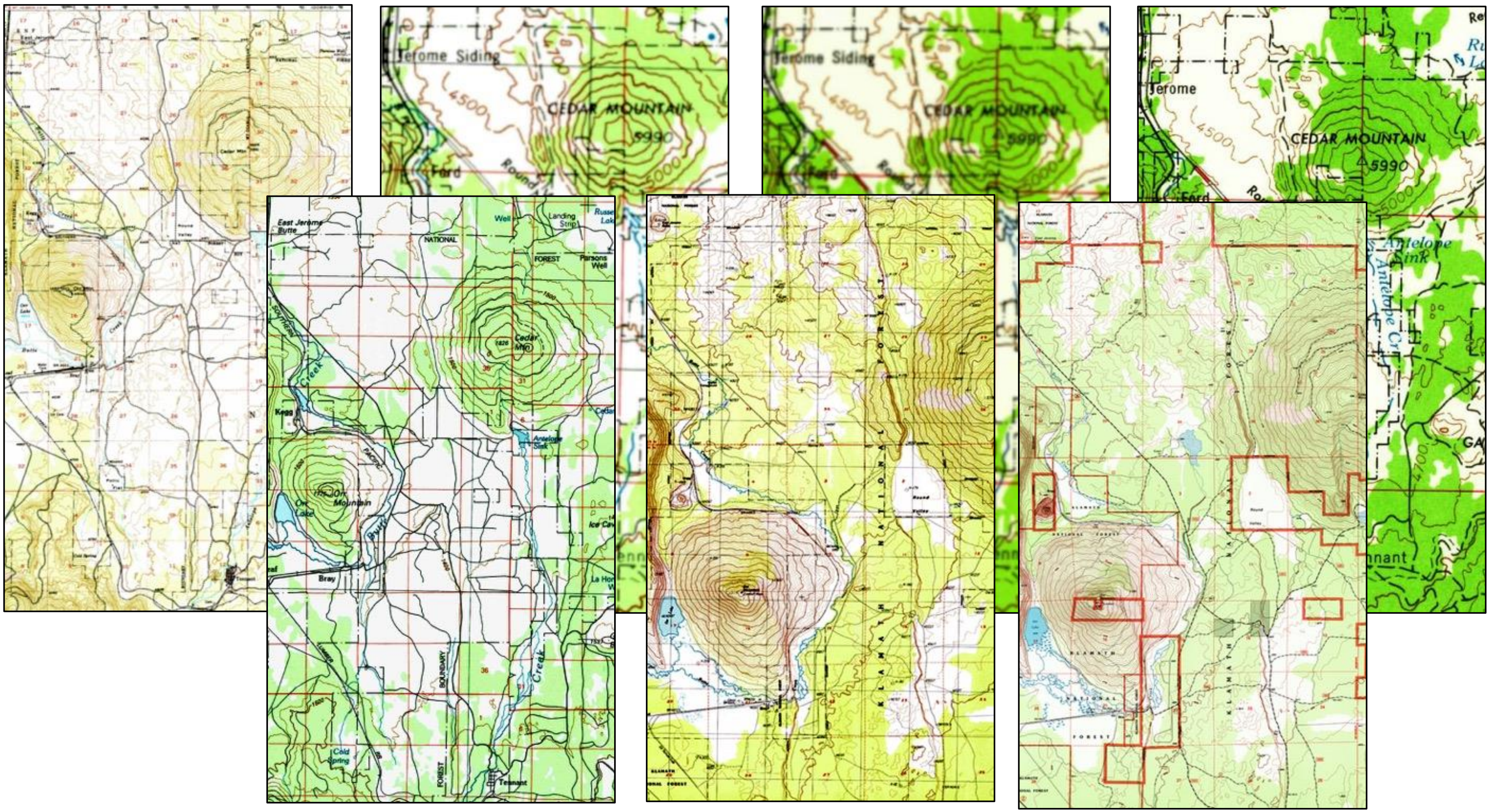




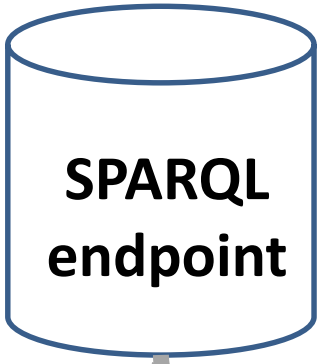
We constructed a KG

Now what?

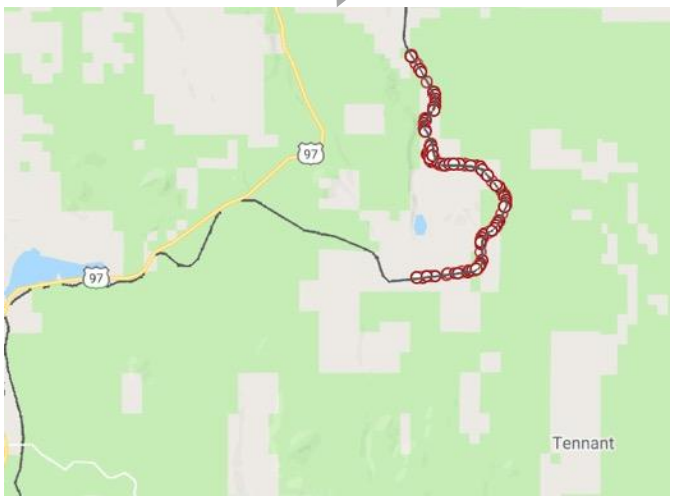
Use Case


Jerome Siding
CEDAR MOUNTAIN
5990
4500
5000
Jerome
Antelope Sink
Antelope Cr
Cedar Mtn
Creek
Bray
Tennant
KLAMATH NATIONAL FOREST

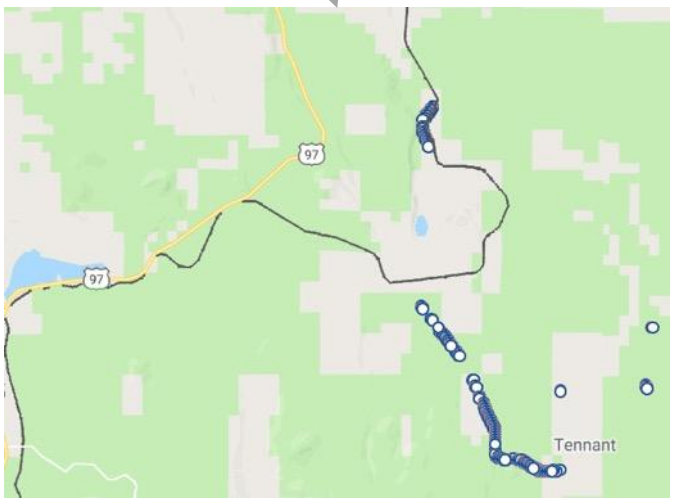
Use Case



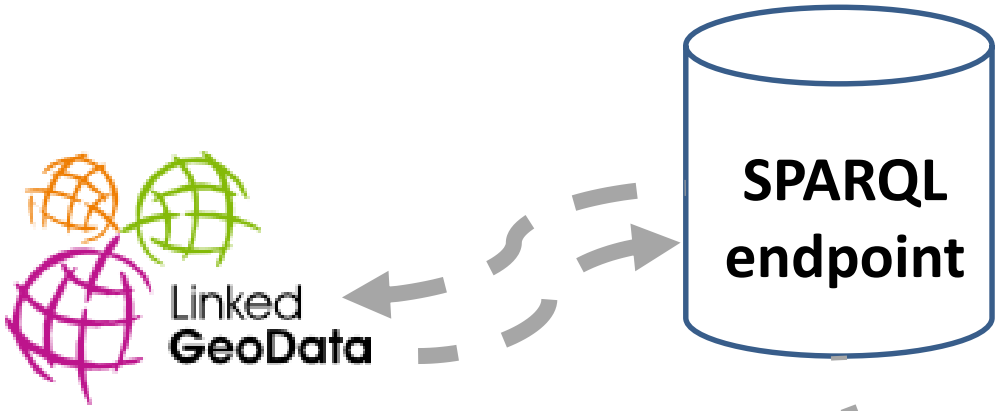
Railroad segments that are similar in 1962 and 2001



Railroad segments that are present in 1962 but are not present in 2001



Use Case



Can you show me a subset of what's abandoned?

Long Bell Lumber Company Railroad at LinkedGeoData

http://linkedgeodata.org/triplify/way177559134

Property	Value
lgdo:changeset	12836533 (xsd:int)
dcterms:contributor	lgdo:user194231
geom:geometry	lgdg:way177559134
rdfs:isDefinedBy	lgd:meta/way177559134
rdfs:label	Long Bell Lumber Company Railroad
dcterms:modified	2012-08-23T21:09:32 (xsd:dateTime)
lgdo:tiger%3Acfcc	B11
lgdo:tiger%3Acounty	Siskiyou, CA
lgdo:tiger%3Aname_base	Long Bell Lumber Company RR
lgdo:tiger%3Areviewed	no
lgdo:tiger%3Asource	tiger_import_dch_v0.6_20070809
lgdo:tiger%3Atlid	113280414:113280416:113280418:113280420:113280421
rdf:type	spatial:Feature lgdm:Way lgdo:AbandonedRailway lgdo:RailwayThing
lgdo:version	1 (xsd:int)

lgdo:AbandonedRailway

Evaluation

Railroad data from a collection of historical maps:

- Bray, California (7)
- Louisville, Colorado (4)

- Segmentation
  - Runtime
  - Number of nodes
- Geo-linking
  - Runtime
  - Correctness (Precision, Recall & F1)
- RDF
  - Query time
  - Query complexity
  - Query robustness

Results

- Segmentation

Table 1. Segmentation Statistics for Bray

Year	# vecs	Runtime (s)	# nodes
1954	2382	<1	1
1962	2322	36	5
1988	11134	1047	11
1984	11868	581	24
1950	11076	1332	43
2001	497	145	57
1958	1860	222	85

Table 2. Segmentation Statistics for Louisville

Year	# vecs	Runtime (s)	# nodes
1965	838	<1	1
1950	418	8	5
1942	513	5	8
1957	353	4	10

Results – cont'd

- Geo-linking

Table 3. "Geo-linking" Results

	Precision	Recall	F1
BRA-baseline	0.193	1.000	0.323
BRA	0.800	0.750	**0.774**
LOU-baseline	0.455	1.000	0.625
LOU	0.833	1.000	**0.909**

Results – cont'd

- RDF

```
SELECT ?f ?wkt WHERE {
  ?f a geo:Feature ;
     geo:hasGeometry [ geo:asWKT ?wkt ] ;
     dcterms:date "1962-01-01T00:00:00"^^xsd:dateTime ;
     dcterms:date "2001-01-01T00:00:00"^^xsd:dateTime .
  FILTER NOT EXISTS { ?f geo:sfContains _:_ } }
```

Table 4. Query Time Statistics (in milliseconds)

	avg	min	max
SIM-BRA	12	10	18
SIM-LOU	11	9	20
DIFF-BRA	10	8	20
DIFF-LOU	10	9	14
UNIQ-BRA	14	8	28
UNIQ-LOU	15	9	17

Discussion

- Complexity of changes in original topographic maps
- Quality & level of detail
- Crowdsourcing
  - LinkedGeoData

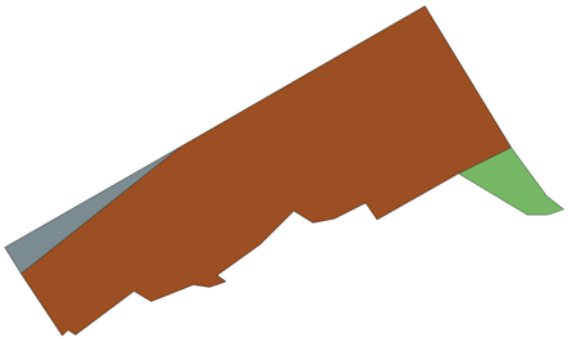
- How can we do better?
  - Segmentation:
    - Optimize buffer size hyperparameter (heuristics/learning)
    - Normalize & denoise the original data
    - Parallel processing
  - Geo-linking:
    - Expand to additional KBs (Wikidata)

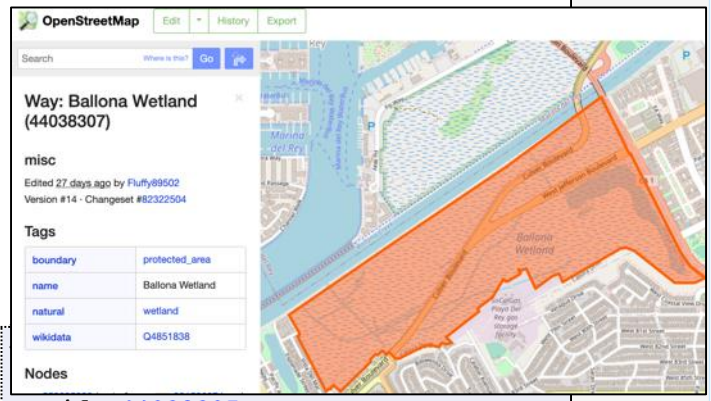
Related Work

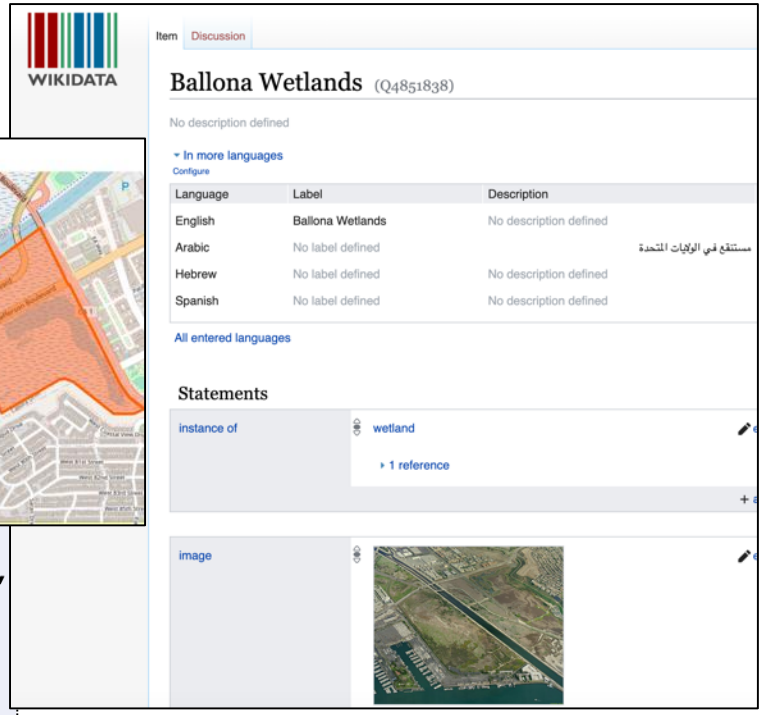
- Transforming geospatial vector data into RDF
  - Kyzirakos et al. [1], Usery et al. [2]
  - Do not address:
    - Geospatial entity intra-linking or distant linking
    - Semantics
- Geographical data conflation
  - Li et al. [3], Ruiz et al. [4]
  - Do not address:
    - Linked Data or Semantics
- Geospatial data integration in the web
  - Prudhomme et al. [5]
  - Do not address:
    - Geospatial entity intra-linking or distant linking

Future (present) Work

- Extend
  - **Wetlands**
  - Forests
  - Highways

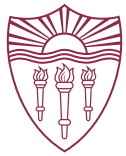







```
    id: 44038307,
    timestamp: "2020-03-18T01:26:12Z",
    version: 14,
    changeset: 82322504,
    user: "Fluffy89502",
    uid: 10165657,
  - nodes: [
    ],
  - tags: {
        boundary: "protected_area",
        name: "Ballona Wetland",
        natural: "wetland",
        wikidata: "Q4851838"
    }
},
```

Conclusions



- Unsupervised approach to integrate, relate, & interlink geospatial data from digitized resources
- Publishable structured semantic-rich linked spatio-temporal data
- Enables users to easily understand & analyze geographic information across time & space
- Fuel discovery

- Source code available at:

https://github.com/usc-isi-i2/linked-maps

References

[1] Kyzirakos, K., Vlachopoulos, I., Savva, D., Manegold, S., Koubarakis, M.: *Geotriples: a tool for publishing geospatial data as rdf graphs using r2rml map- pings* (2014)

[2] Usery, E.L., Varanka, D.: *Design and development of linked data from the national map* (2012)

[3] Li, L., Goodchild, M.F.: *An optimisation model for linear feature matching in geographical data conflation* (2011)

[4] Ruiz, J.J., Ariza, F.J., Urena, M.A., Bla´ zquez, E.B.: *Digital map conflation: a review of the process and a proposal for classification* (2011)

[5] Prudhomme, C., Homburg, T., Ponciano, J.J., Boochs, F., Cruz, C., Roxin, A.M.: *Interpretation and automatic integration of geospatial data into the semantic web* (2019)