



# Constructing a Knowledge Graph of Historical Mining Data

**Basel Shbita, Namrata Sharma, Binh Vu,  
Fandel Lin & Craig A. Knoblock**  
*USC Information Sciences Institute*

*6th International Workshop on Geospatial Linked Data (GeoLD)  
co-located with the 21st Extended Semantic Web Conference (ESWC 2024)*  
05/26/2024

# Agenda

- Intro ←
- Problem
- Approach
- Evaluation & Discussion
- Spatio-Temporal Analysis via GeoSPARQL
- Related Work
- Future Directions
- Conclusions

# Historical & Geo Data

- Rich sources of information
  - understanding human & environmental systems
  - describing human & natural activities
- Labor-intensive to analyze across time & space
  - e.g., economic viability, physical changer, geo-related characteristics
- Often require grounding & additional contextual information
  - e.g., demographics, geology, stratigraphy, etc...

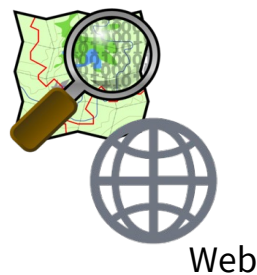
**USGS**  
Mineral Resources / Online Spatial Data

**USMIN Mineral Deposit Database**  
A developing national-scale geospatial database that will be information on the most important mines, mineral deposits, and United States.

**Mineral Resources Data System (MRDS)**  
MRDS is a collection of reports describing metallic and nonmetallic mineral resources throughout the world. Included are deposit name, location, commodity, deposit description, geologic characteristics, production, reserves, resources, and references. It is assumed the original MRDS and MAGNUS. MRDS is a large, complex, and somewhat problematic. This service provides a subset of the database comprised of those data fields deemed most useful and which most frequently contain some information, but full reports of most records are available as well.

**View**  
Show in a web browser window:  
Show in Google Earth or other KML viewer:  
deposit.kml (uncompressed 2.5M bytes)  
deposit.kmz (compressed 690K bytes)

Geospatial databases (e.g., mining sites)



Web

Commodity	Country	Project	Status	Lat	Long	Mine Type	Company	Economic Mineralogy	Principal Process	MRDS ID	MRDS
NIL	Albania	Kales-Korine-Lure	Deposit				unknown				
NIL	Albania	Livshaki-Pogonice	Deposit				unknown				
NIL	Albania	Drevoit Group	Deposit				unknown				
NIS	Argentina	Luz Aguilar	Deposit				Maj Hill Mines		F		
Cu-Co-Ni	Australia	Stanton	Deposit				Northern-COMET		F		
NIS	Australia	Aurubay	CGM			UG	Dundas Mining		F	3.8	
Cu-Ni-Co	Australia	Mount Fyfe (Ni-Co-Cu)	Deposit			OC	China Minerals		HM	0	
NIS	Australia	Bamboo Creek	Deposit			OC	Sakra Resources, Malak Australia		F	12.5	0.11
NIS	Australia	Murrumbidgee	Deposit			UG	Mer Mining 50%, Destra 20%, Hoama Mining 20%		F	0	
NIL	Australia	Murrumbidgee	Operating				Glencore		HPAL	108.6	
NIL	Australia	KNP East-Bulung Turcus	Deposit			OC	Arden Resources		HPAL	0	
NIL	Australia	KNP East-Bulung Bulung East	Deposit			OC	Arden Resources		HPAL	0	
NIL	Australia	Lake Yinbergoods	Deposit			OC	Fireant Resources		HPAL	0	
NIL	Australia	Wilkes-Hoop	Deposit			OC	GME Resources		HL	1.6	
NIL	Australia	NiWest Mount Kichony	Deposit			OC	GME Resources		HL	8.8	
NIL	Australia	NiWest Escaygulus	Deposit			OC	GME Resources		HL	0	
NIL	Australia	NiWest Warranana	Deposit			OC	GME Resources		HL	0	
NIL	Australia	NiWest Murrin North	Deposit			OC	GME Resources		HL	3.4	
NIL	Australia	NiWest White Nauri	Deposit			OC	GME Resources		HL	1.5	
NIL	Australia	NiWest Merredin	Deposit			OC	GME Resources		HL	0	
NIL	Australia	Coronation Dam-Duck Hill	Deposit			OC	White Cliff Minerals		HPAL	0	
NIL	Australia	NiWest Magpie Hill	Deposit			OC	unknown		HPAL	0	
NIL	Australia	Greenfield	Deposit			OC	White Cliff Minerals		HPAL	0	
NIL	Australia	Pyke Hill	Deposit			OC	Cougar Metals, Admiralty Resources		HPAL	4.2	
NIL	Australia	Pelican	Deposit			OC	GSM Mining Company		HPAL	0	
NIL	Australia	Larkins-East End	Deposit			OC	Golden Life-Northern (PVA)		HPAL	0	
NIL	Australia	Mount Clifford-Marricla	Deposit			OC	Norwell Minerals		HPAL	0	

Tables with geo-data

**TECHNICAL REPORT**  
On the  
**KALZAS TUNGTEN PROJECT**  
MAYO MINING DISTRICT  
YUKON, CANADA

NI 43-101 Technical Report  
Resource Estimate  
of the  
Fox Property, Ridley Creek

Happy Creek Minerals Ltd.  
NI 43-101 Resource Update for the  
BN Zone and Maiden Resource  
of the BK Zone of the Fox Property,  
British Columbia

Prepared for  
Prospector Resources Corp.  
6300 Cedarhurst Street  
Vancouver, B.C.  
V6N 1A1

By  
R. Allan DeJong, P. Geo.  
Aurum Geological Consultants Inc.  
1181 Granite Road  
Whitehorse, Yukon, T1A 2V9

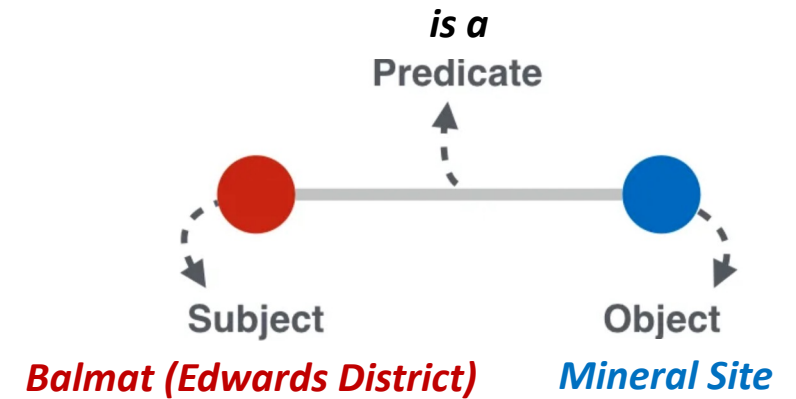
October 12, 2016

Report Date: April 15, 2016  
Data cut-off date: December 22, 2015  
Effective date of the resource estimate: March 15, 2016

Mining reports

# KGs

- Knowledge Graph (KG)?
  - **Graphs** are natural way to **encode** data
  - Using **semantic concepts & relationships**
    - Semantic Network = **Knowledge Graph**
- Why use KG?
  - Combine **expressivity, interoperability, & standardization** in the **Semantic Web** stack
- Semantic Web?
  - Extension of WWW, enabling the Web of Data (aka “Linked Data”)
  - Encoding of **semantics** with the data
  - Linked Open Data **principles** // FAIR



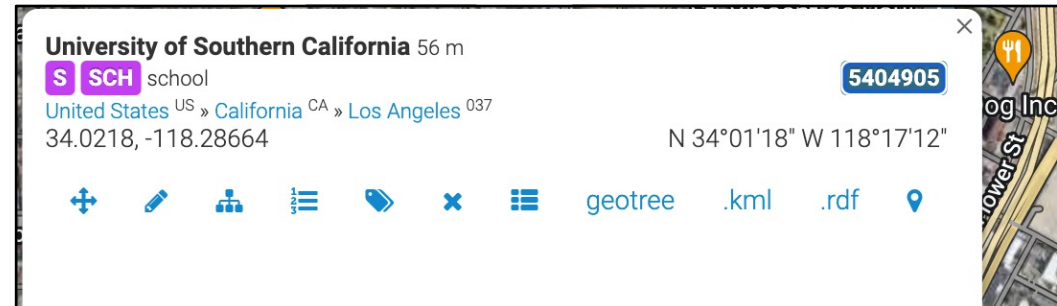
# Geo & Spatio-Temporal KGs

- Spatio-Temporal KGs

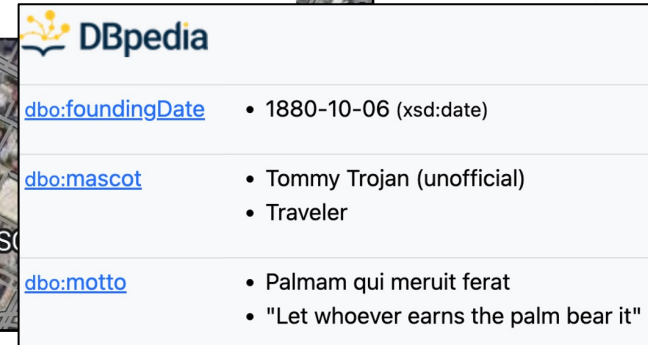
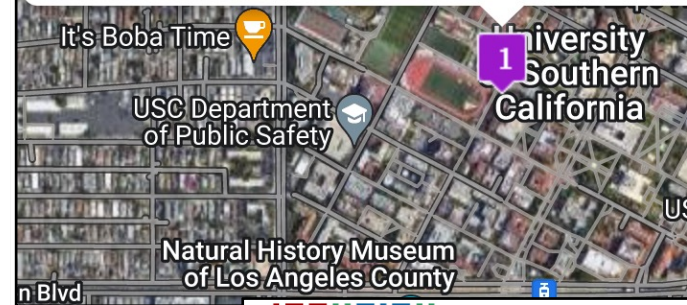
- Contextual (**what**)
- **Spatial** (**where**)
- **Temporal** (**when**)

- Geo-**semantics**

- Representation, annotation, & reasoning
- Modeling & ontology development
- Integration & interoperability

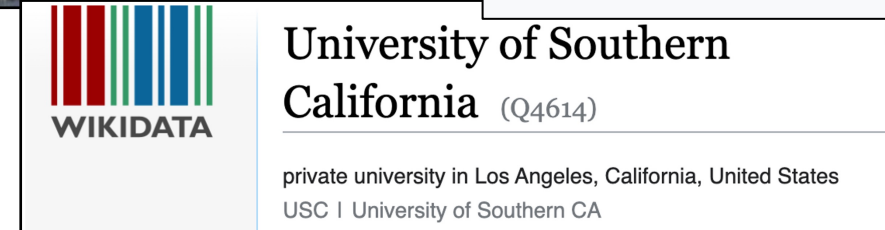


University of Southern California 56 m  
SCH school 5404905  
United States US » California CA » Los Angeles 037  
34.0218, -118.28664 N 34°01'18" W 118°17'12"



DBpedia

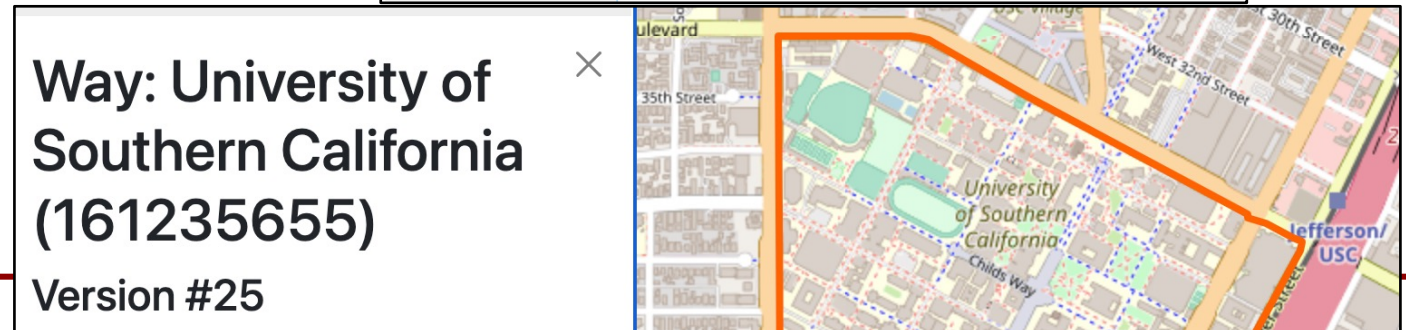
<a href="#">dbo:foundingDate</a>	• 1880-10-06 (xsd:date)
<a href="#">dbo:mascot</a>	• Tommy Trojan (unofficial) • Traveler
<a href="#">dbo:motto</a>	• Palmm qui meruit ferat • "Let whoever earns the palm bear it"



WIKIDATA

University of Southern California (Q4614)

private university in Los Angeles, California, United States  
USC | University of Southern CA



Way: University of Southern California (161235655)  
Version #25



# Intro: Spatio-Temporal KGs

- So, what's so **special** about them?
  - Spatial analysis
  - Temporal analysis
  - Spatio-temporal aggregations
  - Geographic QA
  - Environmental & social science
  - Urban planning
  - Transportation
  - etc...

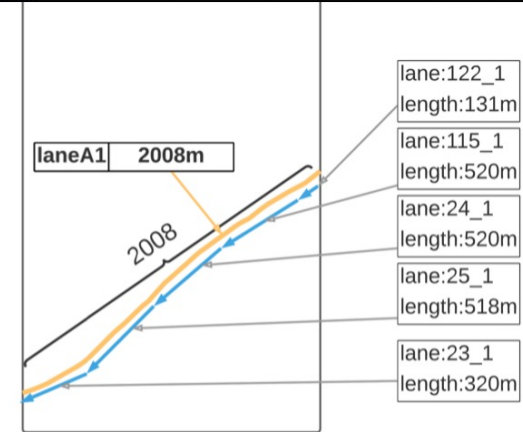
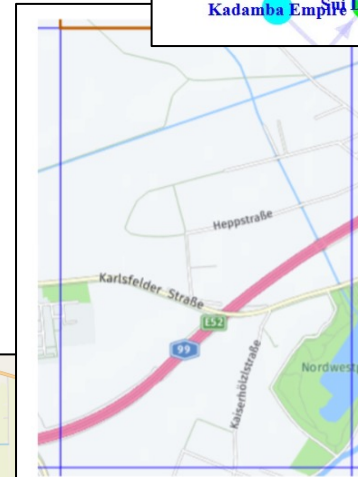
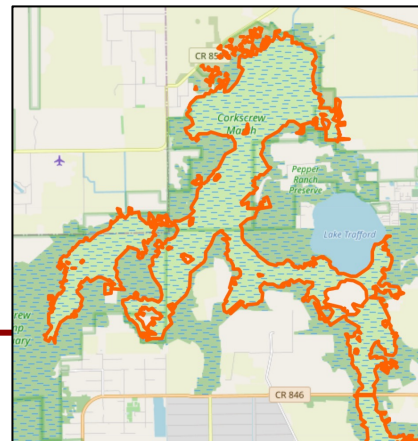
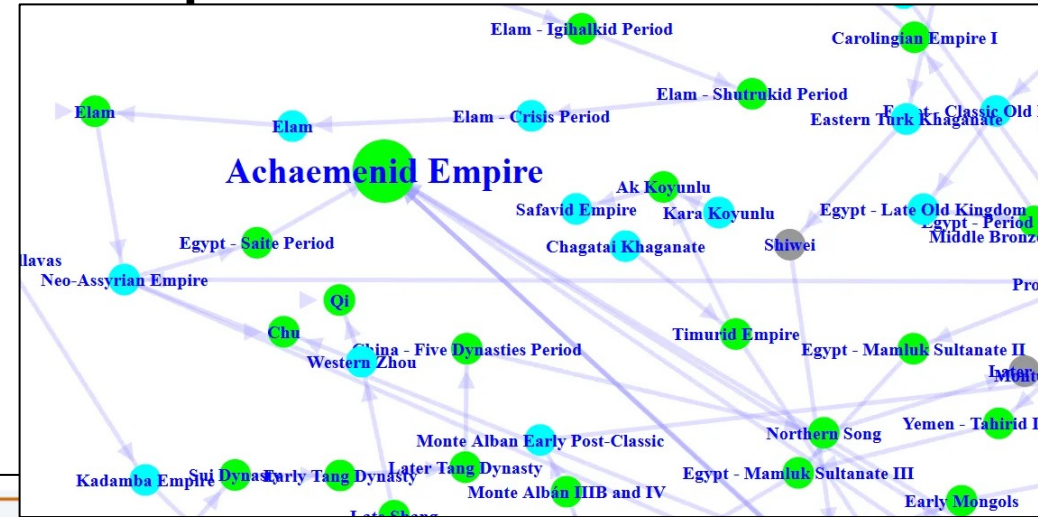



figure from Szwoch, G. (2019). Combining road network data from openstreetmap with an authoritative database. *Journal of Transportation Engineering, Part A: Systems*, 145(2), 04018085.

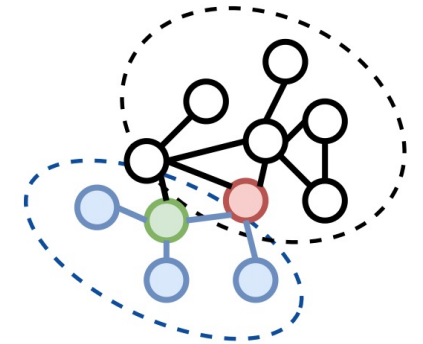
figure from <https://terminusdb.com/blog/human-history-knowledge-graph/>

# Agenda

- Intro
- Problem 
- Approach
- Evaluation & Discussion
- Spatio-Temporal Analysis via GeoSPARQL
- Related Work
- Future Directions
- Conclusions

# Research Problem

- How can we **transform & link unstructured digitized & historical** geo-data into **structured, semantic, & queryable spatio-temporal KGs**?
- Objectives:
  - **Auto KG construction & entity resolution (ER)** from various historical geo-data sources
  - **Semantic enrichment by linking (EL)** to additional sources on the web
  - Adherence to **Semantic Web principles**
    - shared, **accessible**, visualized, standardized **across-domains**, & **scalable** for easy use by downstream tasks for easy analysis & expressive **integration**

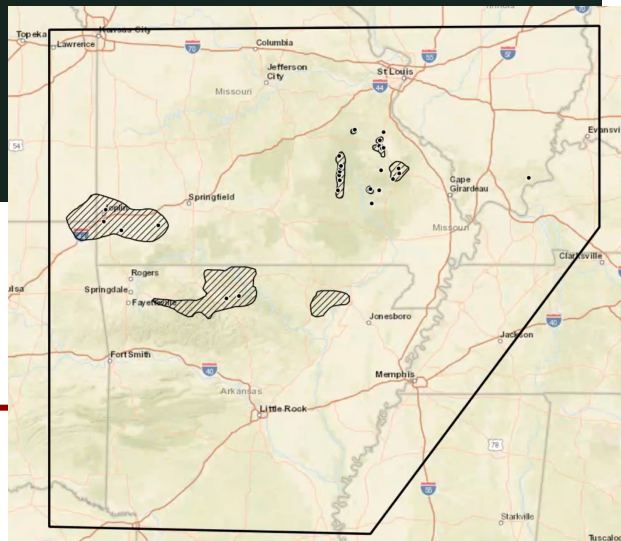




# Goals

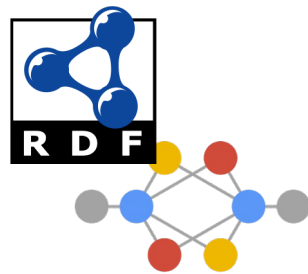
- Integrating **geo-referenced textual & historical data** with quantitative information into a comprehensive, dynamic, & **spatio-temporal KG**
  - capture data & entity semantics, entity resolution, & accurate data modeling
- Demonstrate via a KG of **historical mining data**
  - Historically takes months of work; geologists describe it as a “soul crushing exercise”

```
{  
  "MineralSite": [  
    {  
      "source_id": "https://w3id.org/usgs/z/4530692/2P29BJHV",  
      "record_id": 1,  
      "name": "NI 43-101 Technical Report for the Lantinen Koillismaa Project in Europe, Finland dated March 2017",  
      "location_info": {  
        "location": "POINT(28.17472 65.94611)",  
        "crs": "WGS84",  
        "country": "Finland",  
        "state_or_province": "Central Finland"  
      },  
    }  
  ]  
}
```



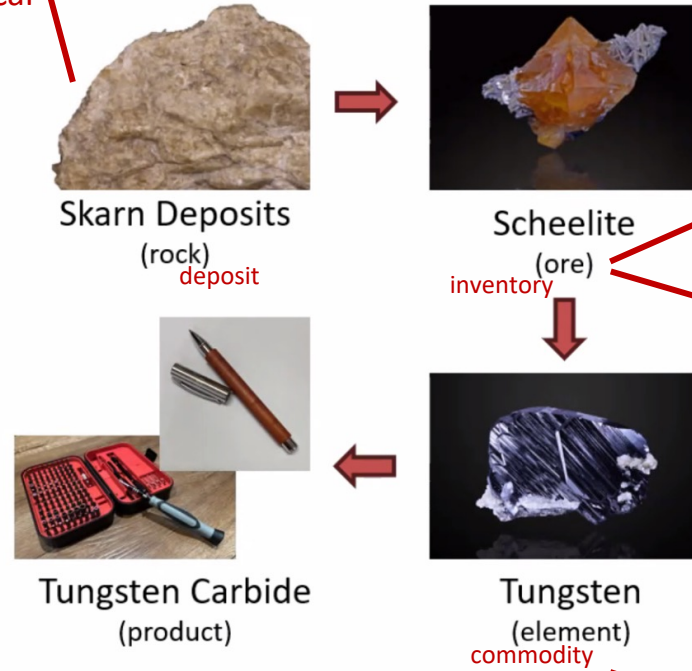
Extracted  
Textual &  
Vector Data

**Data understanding &  
KG construction**

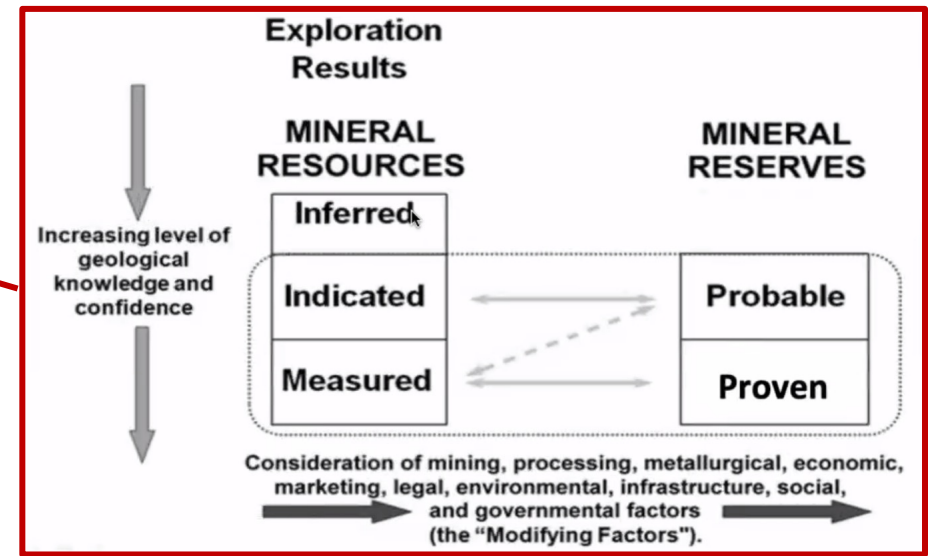


# Transformation & Aggregation

natural occurrence of minerals (geological formations)



naturally occurring solid material usually mixtures of minerals and other materials (bauxite = ore of aluminum; hematite = ore of iron)



material or primary agricultural product

COMMODITY	AS-REPORTED FORM	AS-REPORTED ELEMENT	CONVERSION FACTOR	CONVERTED FORM
Aluminum	Al		1.8895	Al <sub>2</sub> O <sub>3</sub>
Antimony	Sb		1.1971	Sb <sub>2</sub> O <sub>3</sub>
Barium	Ba		1.6995	BaSO <sub>4</sub>
Borates	B	Boron	3.2198	B <sub>2</sub> O <sub>3</sub>
Borates	H <sub>3</sub> BO <sub>3</sub>	Boric Acid	0.5629	B <sub>2</sub> O <sub>3</sub>
Cesium	Cs <sub>2</sub> O	Cesium Oxide, C	0.9432	Cs
Chromium	Cr		1.4615	Cr <sub>2</sub> O <sub>3</sub>
Cobalt	CoOH	Cobalt Hydroxide	0.776	Co
Iron Ore	Fe <sub>2</sub> O <sub>3</sub>	Hematite, Haem	0.6994	Fe

# One desired output: Grade-Tonnage model

- For a given commodity/deposit type/location/time-range:
  - Construct grade and tonnage models from the data on existing mines
  - Compile rich mineral site data

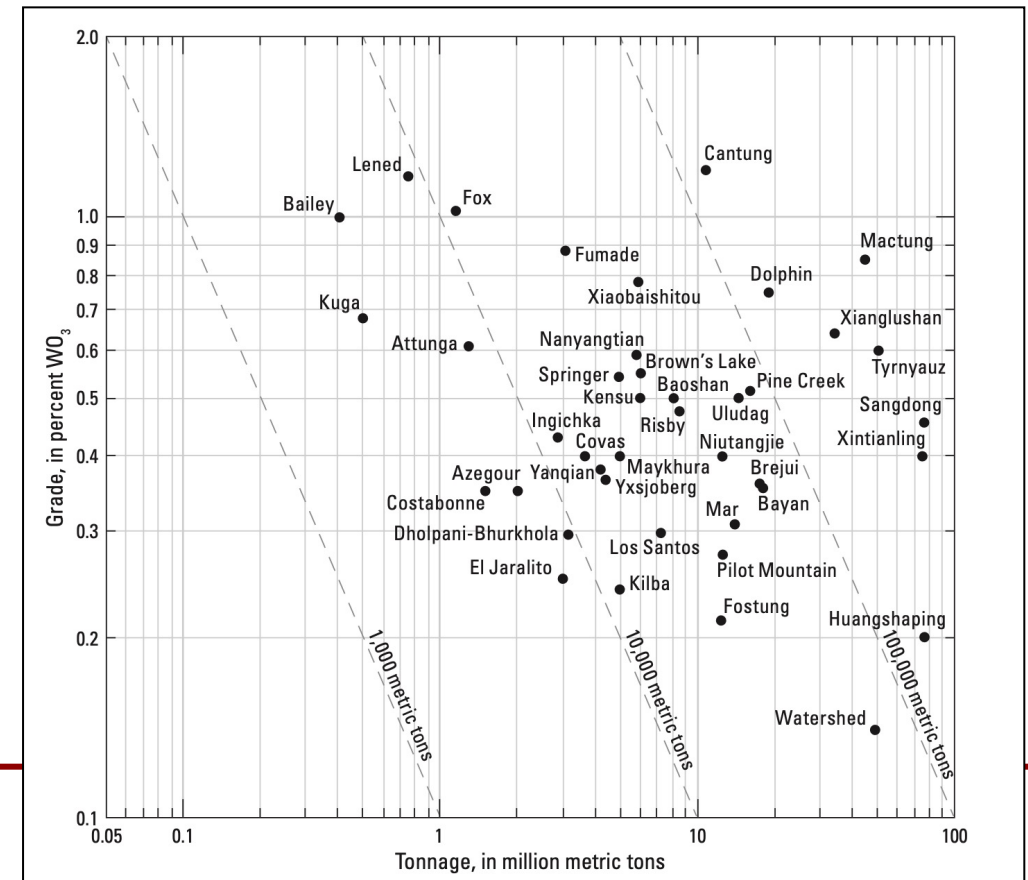
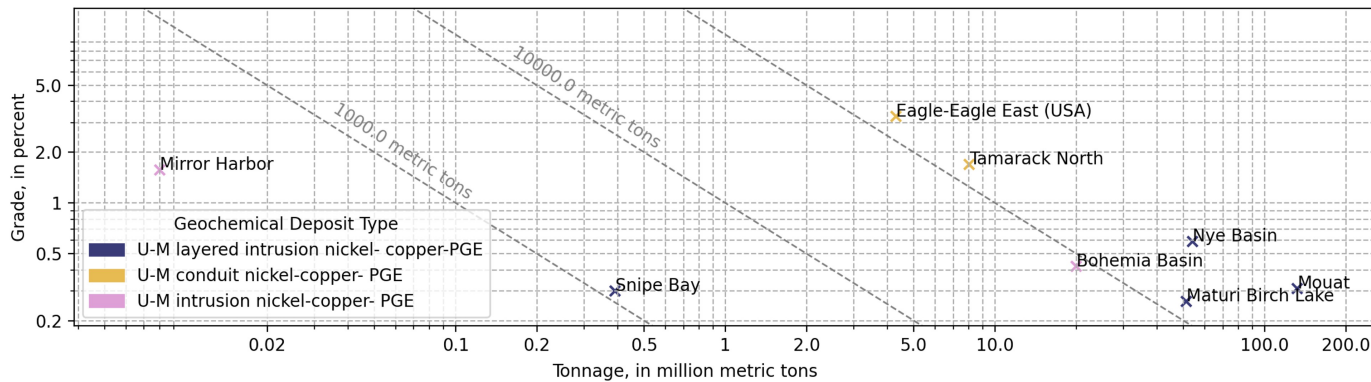



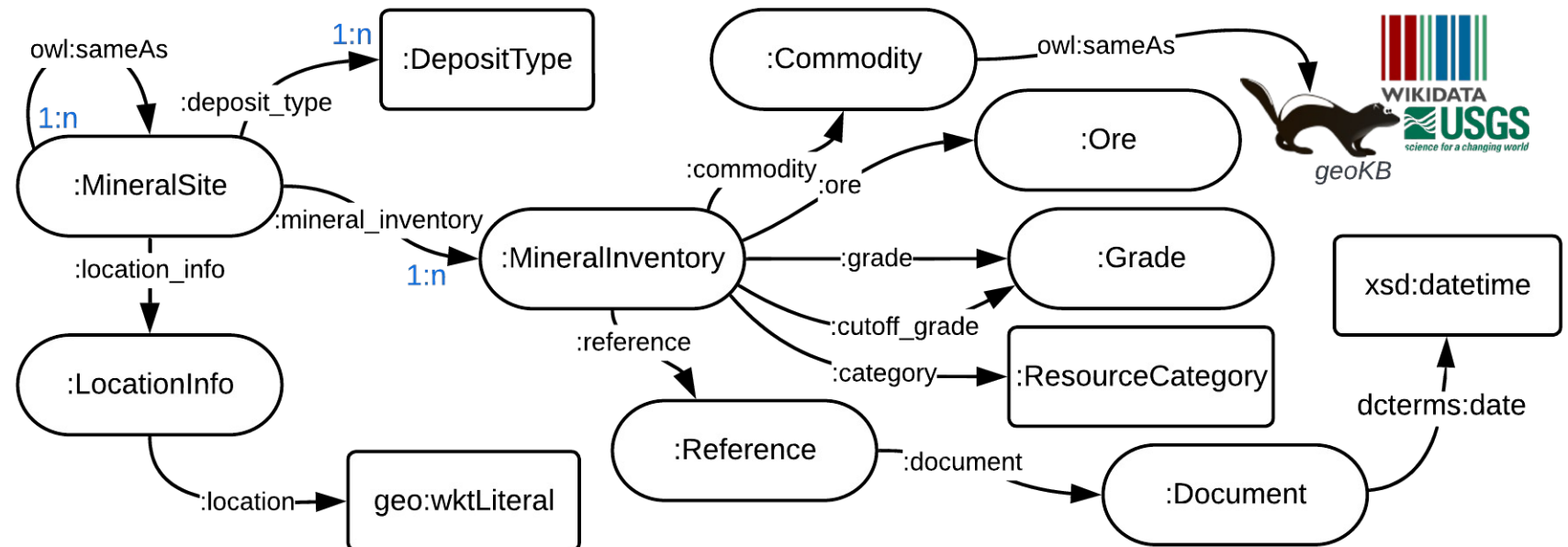
Figure from Green, C. J., Lederer, G. W., Parks, H. L., & Zientek, M. L. (2020). Grade and tonnage model for tungsten skarn deposits—2020 update (No. 2020-5085). US Geological Survey

# Agenda

- Intro
- Problem
- Approach 
- Evaluation & Discussion
- Spatio-Temporal Analysis via GeoSPARQL
- Related Work
- Future Directions
- Conclusions

# Approach

- Step 1. Semantic Modeling & URI assignment (Data Representation)
  - Transform & materialize the data (construct KG)
    - **Generate entities** (URIs) based on unique identifiers
    - Provide a **useful semantic representation** supporting downstream tasks
  - Construct a **meaningful semantic model**
    - Follows W3C & OGC standards (**GeoSPARQL**)



# GeoKB as a Target KB for EL

GEOKB

## Nickel (Q162561)



mineral species in the Iron Group sourced from Mindat and the Geoscience Ontology

Niccolum | nikle | Nickel | نیکل | Nikelo | ニッケル | Nikal | Nichel | Nikiel | Nikl | IMA1966-039 | 니켈 | Niķelis | Nikkeli | Նիկել | Níquel | Nikelj | निकेल | ნიკელი | Нікел | Nikèl | Никель | Nîkel | Iztāctepoztli | निकेल | നിക്കൽ | نیکل | Niquèl | Никл | לניקל | Nickyl | நிக்கல் | Нікель | Nikeli | Nikil | Nikelis | Νικέλιο | Néckel | Никель | Konukōreko | Niken | Nichele | 自然镍 | Nikkel

### Statements

subclass of	01.AA.05 - Copper-cupalite family ▸ 2 references
	mineral material ▸ 1 reference

has chemical element	nickel ▸ 1 reference
----------------------	-------------------------

member of	Iron Group ▸ 1 reference
-----------	-----------------------------

same as	<a href="http://www.mindat.org/min-2895.html">http://www.mindat.org/min-2895.html</a> ▾ 0 references
	<a href="https://w3id.org/gso/mineral/nickel">https://w3id.org/gso/mineral/nickel</a> ▾ 0 references



# Approach – cont'd

- Step 2. Entity Linking

- Link the generated entities to a domain data-rich vocab (i.e., GeoKB)
  - Determine similarity by textual similarity (i.e., Jaccard)
  - Directly within SPARQL

"Clay, Fire (Refractory)"

```

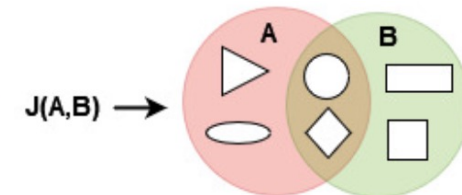
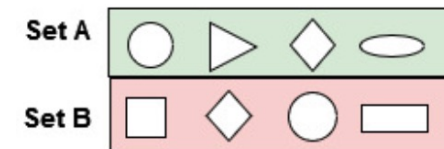
entity.value \
0 uri https://geokb.wikibase.cloud/entity/Q413
1 uri https://geokb.wikibase.cloud/entity/Q424
2 uri https://geokb.wikibase.cloud/entity/Q423
3 uri https://geokb.wikibase.cloud/entity/Q421
4 uri https://geokb.wikibase.cloud/entity/Q162319
5 uri https://geokb.wikibase.cloud/entity/Q425
6 uri https://geokb.wikibase.cloud/entity/Q426
7 uri https://geokb.wikibase.cloud/entity/Q428
8 uri https://geokb.wikibase.cloud/entity/Q427
9 uri https://geokb.wikibase.cloud/entity/Q429
    
```

```

1 SELECT ?entity ?entityLabel WHERE {
2   ?entity rdfs:label ?entityLabel.
3   ?entity gkbt:P1 gkbi:Q406. # instance of mineral commodity
4   FILTER(CONTAINS(LCASE(?entityLabel), "nickel")) }
    
```

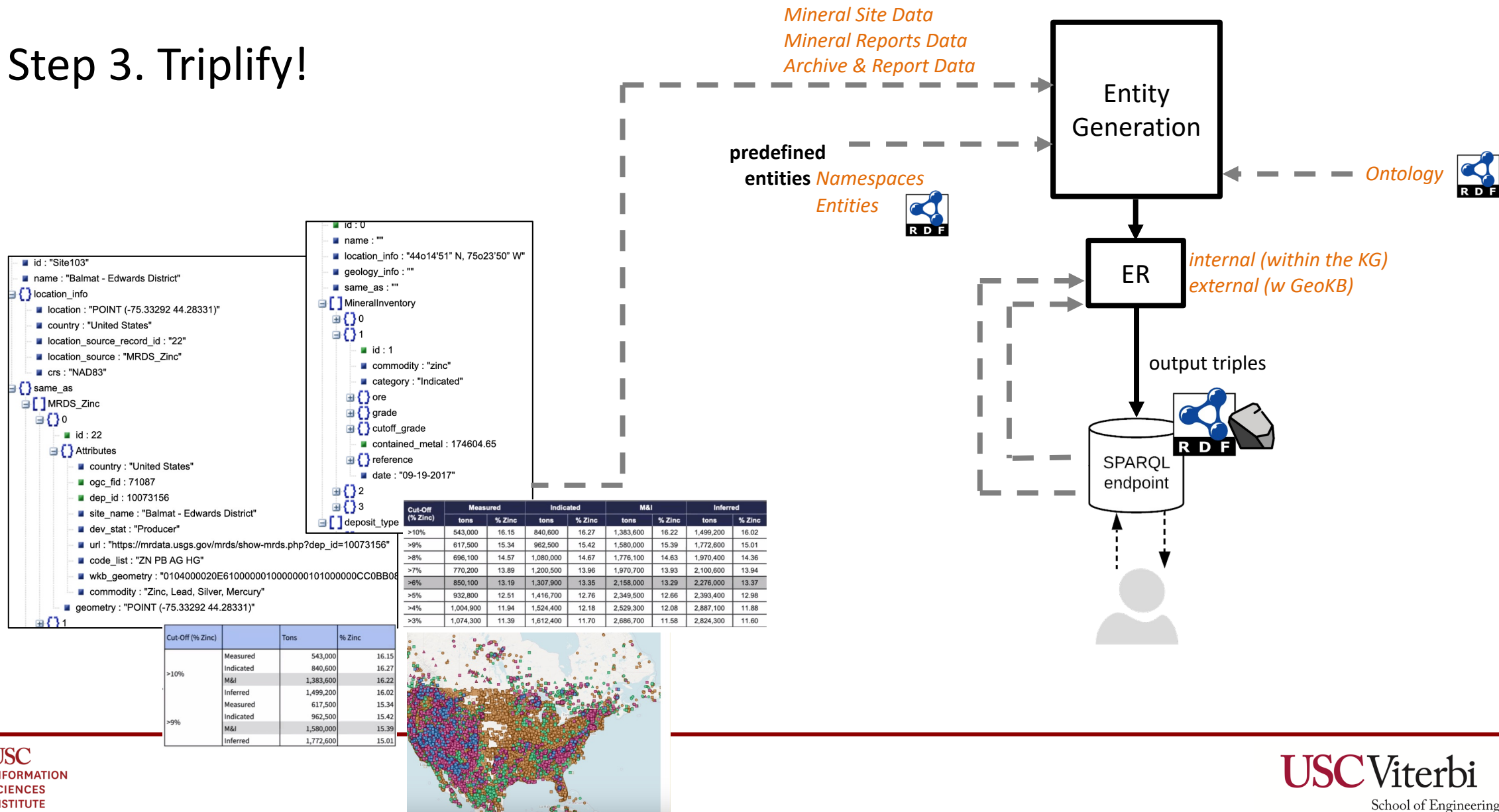
	entityLabel.xml:lang	entityLabel.type	entityLabel.value
0	en	literal	high alumina clay aluminum
1	en	literal	bloating material clay
2	en	literal	brick clay
3	en	literal	clay
4	en	literal	Clay
5	en	literal	bentonite clay
6	en	literal	chlorite clay
7	en	literal	fire (refractory) clay
8	en	literal	fullers earth clay
9	en	literal	glaucanite clay

Jaccard



# Approach - cont'd

- Step 3. Triplify!



# Agenda

- Intro
- Problem
- Approach
- Evaluation & Discussion ←
- Spatio-Temporal Analysis via GeoSPARQL
- Related Work
- Future Directions
- Conclusions

# Evaluation & Discussion

- Data completeness (SHACL)
- Entity linking
- Query performance

Characteristic	Count
Total Triples	2,397,708
Distinct Classes	16
Instances (Non-literals)	226,267
Geospatial Instances	2,884
Blank Nodes	1,518,981

Data: 2.4m triples // 135 commodities // focus on 2 critical mineral: nickel, zinc

Method	MRR	Hits@1	Hits@3	Hits@5
String search, then Jaro	0.557	0.459	0.659	0.659
String search, then Jaccard	0.648	0.637	0.659	0.659
Instance search, then Jaro	0.801	0.689	0.926	0.956
<b>Instance search, then Jaccard (proposed)</b>	<b>0.940</b>	<b>0.904</b>	<b>0.978</b>	<b>0.978</b>

```

1 ?ms :location_info/:location ?loc_wkt .
2 FILTER(geof:distance(?loc_wkt, "POINT(-118.57 47.56)"^^geo:wktLiteral, unit:mile) < 500)
    
```

Query Constraint Type	Avg	Min	Max
Textual	450	369	649
Temporal/Numeric	438	388	607
Spatial	708	501	811

# Agenda

- Intro
- Problem
- Approach
- Evaluation & Discussion
- Spatio-Temporal Analysis via GeoSPARQL ←
- Related Work
- Future Directions
- Conclusions

# Spatio-Temporal Analysis via GeoSPARQL

Grade	Assays					
	Cu (%)	Ni (%)	S (%)	Au (g/t)	Pt (g/t)	Pd (g/t)
Concentrate	7.16-10.1	1.66-2.20	18.4-21.5	0.65-1.28	1.17-1.59	5.76-6.71

```

1 SELECT
2   ?ms ?ms_name ?deposit_name ?loc_wkt ?total_tonnage_measured ?total_tonnage_indicated ?total_tonnage_inferred ?total_contained_measured
   ?total_contained_indicated ?total_contained_inferred
3   (?total_tonnage_measured + ?total_tonnage_indicated + ?total_tonnage_inferred AS ?total_tonnage)
4   (?total_contained_measured + ?total_contained_indicated + ?total_contained_inferred AS ?total_contained_metal)
5   (IF(?total_tonnage > 0, ?total_contained_metal / ?total_tonnage, 0) AS ?total_grade)
6 WHERE {
7   {
8     SELECT ?ms ?ms_name ?deposit_name ?country ?loc_wkt
9       (SUM(?tonnage_measured) AS ?total_tonnage_measured)
10      (SUM(?tonnage_indicated) AS ?total_tonnage_indicated)
11      (SUM(?tonnage_inferred) AS ?total_tonnage_inferred)
12      (SUM(?contained_measured) AS ?total_contained_measured)
13      (SUM(?contained_indicated) AS ?total_contained_indicated)
14      (SUM(?contained_inferred) AS ?total_contained_inferred)
15    WHERE {
16      ?ms :deposit_type [ rdfs:label ?deposit_name ] .
17      ?ms :mineral_inventory ?mi .
18      OPTIONAL { ?ms rdfs:label:name ?ms_name . }
19      ?ms :location_info/:location ?loc_wkt .
20      ?mi :category ?mi_cat .
21      ?mi :date ?date .
22    }
23    FILTER(geof:sfWithin(?loc_wkt, "POLYGON(...)" ) .
24    FILTER(?date >= "2000"^^xsd:gYear && ?date <= "2010"^^xsd:gYear) .
25    ?mi :ore [ :ore_value ?ore_val_raw; :ore_unit ?ore_unit ] .
26    ?mi :grade [ :grade_value ?grade_val; :grade_unit ?grade_unit ] .
27    BIND(IF(bound(?ore_val_raw), ?ore_val_raw, 0) AS ?ore_val_pre)
28    BIND(IF(?ore_unit = <http://data.nasa.gov/qudt/owl/unit#MetricTon>, ?ore_val_pre / 1e6, ?ore_val_pre)) AS ?ore_val
29    BIND(IF(CONTAINS(LCASE(STR(?mi_cat)), "measured"), ?ore_val, 0) AS ?tonnage_measured)
30    BIND(IF(CONTAINS(LCASE(STR(?mi_cat)), "indicated"), ?ore_val, 0) AS ?tonnage_indicated)
31    BIND(IF(CONTAINS(LCASE(STR(?mi_cat)), "inferred"), ?ore_val, 0) AS ?tonnage_inferred)
32    BIND(IF(CONTAINS(LCASE(STR(?mi_cat)), "measured") && ?grade_val > 0, ?ore_val * ?grade_val, 0) AS ?contained_measured)
33    BIND(IF(CONTAINS(LCASE(STR(?mi_cat)), "indicated") && ?grade_val > 0, ?ore_val * ?grade_val, 0) AS ?contained_indicated)
34    BIND(IF(CONTAINS(LCASE(STR(?mi_cat)), "inferred") && ?grade_val > 0, ?ore_val * ?grade_val, 0) AS ?contained_inferred)
35  }
36  GROUP BY ?ms ?ms_name ?deposit_name ?loc_wkt }

```

2<sup>nd</sup> aggregation:  
total grade & tonnage  
computation

`FILTER(geof:sfWithin(?loc_wkt, "POLYGON(...)" ) .`

`FILTER(?date >= "2000"^^xsd:gYear && ?date <= "2010"^^xsd:gYear) .`

`BIND(IF(?ore_unit = <http://data.nasa.gov/qudt/owl/unit#MetricTon>, ?ore_val_pre / 1e6, ?ore_val_pre)) AS ?ore_val`

1<sup>st</sup> aggregation:  
tonnage computation


```

qudt:hasDimension: "",
qudt:abbreviation: "g tonne-1",
..
qudt:hasPart: [
{
  qudt:hasDimension: "M",
  qudt:quantityKind: "http://data.nasa.gov/qudt/owl/unit#Gram",
  qudt:conversionMultiplier: 0.001,
  qudt:conversionOffset: 0.0,
  qudt:symbol: "g"
},
{
  ccut:exponent: "-1",
  qudt:hasDimension: "M",
  qudt:quantityKind: "http://data.nasa.gov/qudt/owl/unit#MetricTon",
  qudt:conversionMultiplier: 1000.0,
  qudt:conversionOffset: 0.0,
  qudt:symbol: "t"
}
]

```




# Agenda

- Intro
- Problem
- Approach
- Evaluation & Discussion
- Spatio-Temporal Analysis via GeoSPARQL
- Related Work 
- Future Directions
- Conclusions

# Related Work

- General geo KBs (Zhu 2017, Brodaric 2020)
  - Mostly encompasses **conceptual knowledge & data**
  - Does not address: quantitative data integration
- GeoKGs related to mineral data (Qun 2023)
  - Tailored for **geochemical data**
  - Does not address: quantitative data integration
- Information extraction for geo KGs (Wang 2018)
  - Focus is on the **data extraction**
  - Does not address: data integration & entity linking

# Agenda

- Intro
- Problem
- Approach
- Evaluation & Discussion
- Spatio-Temporal Analysis via GeoSPARQL
- Related Work
- Future Directions 
- Conclusions

# Future Directions

- Advanced **data modeling**
  - More modalities (remote sensing)
  - More data (e.g., rapidly changing geographies)
  - Uncertainty & probabilistic modeling
- **Enhanced embedding** techniques for **ER & EL**
  - Expand integration of textual data
  - Utilize subword information & deep learning attention mechanisms
- **KG** expansion
  - Extend KG linkage to additional KBs & LOD
  - Apply & integrate with additional domains like archaeology & environmental sciences
- Dynamic **semantic modeling**
  - Create more sophisticated & evolving semantic models for accurate representation across multiple domains

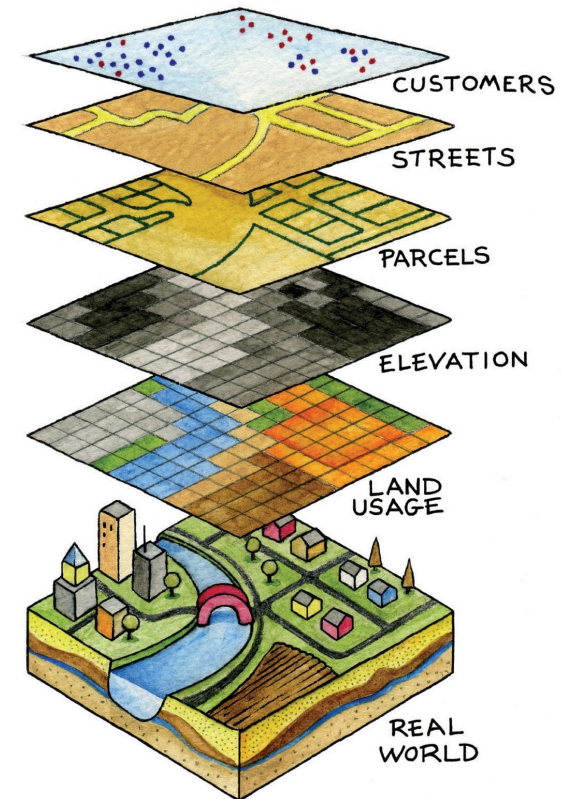



figure from *Essentials of Geographic Information Systems, Ch 7, Saylor Academy, 2012*

# Agenda

- Intro
- Problem
- Approach
- Evaluation & Discussion
- Spatio-Temporal Analysis via GeoSPARQL
- Related Work
- Future Directions
- Conclusions 

# Conclusions

- Presented a method for the construction of a **spatio-temporal KG** from geo-referenced **spatial entities in archive report**
- Contributions
  - **pipeline** for building a KG from extracted **quantitative, spatial & semantic information** from historical mining data archives
    - automatically, incrementally, follows LD & SW principles, linked to web
  - **method to identify & retrieve instances of a given type** from a **publicly available KG**
    - specifically, entity matching commodities with GeoKB
  - **spatio-temporal queries** to automatically **generate grade-tonnage models**
  - **publicly available** resulting **KG** in the form of linked data covering two critical minerals: Zinc & Nickel
    - queryable RDF via a **(Geo)SPARQL endpoint**



<https://minmod.isi.edu/>

<https://github.com/DARPA-CRITICALMAAS/ta2-minmod-data/>