# Deeply Integrating Linked Data with Geographic Information Systems

Esri UC 2019, July 9th 2019

Gengchen Mai<sup>1</sup> Krzysztof Janowicz<sup>1</sup> Bo Yan<sup>1</sup> Simon Scheider<sup>2</sup>



<sup>1</sup>STKO Lab, University of California, Santa Barbara

<sup>2</sup>Utrecht University, Netherlands

INTRODUCTION		
●0000000		

# Linked Data (LD)

- A set of design principles for sharing machine-readable interlinked data on the Web<sup>1</sup>
- Datasets constructed by following these principles



<sup>&</sup>lt;sup>1</sup>https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/

INTRODUCTION
0000000

Discussio 000 Conclusion 00

# LINKED OPEN CLOUD (LOD)



#### Geographic data play a prominent role in the Linked Data cloud as places act as central nexuses that interconnect events, people, and objects.

Introduction	Метноd	Discussion	
00000000	00000000	000	

#### INTRODUCTION

#### Linked Data and Knowledge Graphs power many applications:

- Intelligent assistance systems: Apple Siri
- Search engines: Google Search
- The Linked Data paradigm addresses many key challenges of GIScience and cyber-data infrastructures (Kuhn et al., 2014).



Figure From https://towardsdatascience.com/automatic-question-answering-ac7593432842

INTRODUCTION		
000000		

# The plethora of Geographic Linked Data

Despite all these success stories, from a GIS perspective, **Linked Data** seems almost like a one-way street.

- It is now easier than ever to publish and consume geo-data on the (Semantic) Web:
  - Converting shapefiles to RDF
  - Fusing geometries from different sources (Giannopoulos et al., 2014)
  - Discovering links (Mai et al., 2016)
  - Querying remote endpoints (Battle et al., 2012)
  - Computing geospatial properties on-demand (Regalia et al., 2016).
- All this work focuses merely on how to get geo-data out of data silos.
- The question of how to actually make use of this plethora of data remains largely unanswered.





Regalia et al., 2016

INTRODUCTION		
0000000		

# The plethora of Geographic Linked Data

- Two main ways of accessing and utilizing Geographic Linked Data:
  - Directly use GeoSPARQL-enabled SPARQL endpoint: very limited spatial analysis capability
  - Convert Linked Data to CSV/JSON and Import them into a GIS: Data are flattened and the link structure is largely lost
- Problem: While we can semantically enrich geo-data and publish them as Linked Data, consuming these data in a GIS and applying the vast toolboxes of modern spatial analysis is more difficult, especially if we aim at maintaining the link structure while doing so.
- Proposed Solution: a Linked Data connector framework for the deep integration of Linked Data and GIS

INTRODUCTION		
00000000		

# LINKED DATA CONNECTOR

- At first glance a software engineering task: add Linked Data as another data source of a GIS
- In reality a complicated research problem:
  - Linked Data is not a data format but a paradigm for Web-scale, distributed data infrastructures
  - Linked Data does not harmonize well with how we conceptualize data (exchange) in GIS



Figure made based on Esri and Ontotext images

INTRODUCTION		
00000000		

# LINKED DATA CONNECTOR

#### **Considerations:**

- How should GIS and its users interact with Linked Data?
- Which concrete benefits does Linked Data bring to the table with respect to spatial analysis?
- How can these key benefits of Linked Data be maintained during conversion and analysis without having to flatten the data back to a tabular format?

How to utilize the ontologies used to semantically lift Linked Data? Prototypical Implementation: we implement individual toolboxes for ArcGIS 10.4 as proof-of-concept by using ArcPy.



Introduction 00000000	Метноd 00000000	Discussion 000	

# Contribution

- We proposes Linked Data Connector framework a deep integration of Linked Data into GIS.
- We demonstrate how to utilize Semantic Web reasoning and ontologies to extract additional properties (e.g., subsumption reasoning and (inverse) partonomical relations).
- Our deep integration supports exploratory search via n-degree property path queries, a feature that is not typically found in a GIS environment.
- We demonstrate how the queried Linked Data can be seamlessly used to perform GIS analysis.



Method	
0000000	

#### Geographic Linked Data Retrieval

**Buffer Query**: Directly interact with Geographic Linked Data through a SPARQL endpoint

Oakland San Francisco		Sold and the second sec
•	San Leandro Tracy	3.7
	Luermare	- Jan
San Mateo	§ <sup>r</sup> Linked Data Spatial Query	
Re	Input Buffer Query Center	
	LinkedDataSpatialQuery::in_buf_query_center	- 🖻
	in_buf_query_center	
	Input Place Type (optional)	
	dty(wd:Q515)	•
	Disable Transitive Subclass Reasoning (optional)	
	Input Search Radius (mile)	
	Output Location	
	E:\	
	Output Point Feature Class Name	
	5°.	

LINKED DATA CONNECTOR

GENGCHEN MAI , KRZYSZTOF JANOWICZ , BO YAN , SIMON SCHEIDER

Method	
0000000	

#### PROPERTY ENRICHMENT

- Properties Selection: Identify meaningful properties
- Datatype Casting of Datatype
   Properties: Linked Data typically do not restrict datatype properties to a specific XSD data type
- Spatial Relation Normalization: handling N-to-N relation/property



Who was born in Berlin?

 INTRODUCTION
 METHOD
 Discussion
 Conclusic

 00000000
 00000000
 000
 000
 000

# PROPERTY ENRICHMENT

#### Get the **common properties** of these spatial entities and their **subdivisions**:

- Common properties
- Inverse common properties
- Expanded common properties
- Inverse expanded common properties

I Linked Data Location Entities Property Enrichment		
Input wikidata location entities Feature Class	<u>.</u>	Inverse Expanded
F-11CS8_STKO_Lab/STKO_Paraarch/caraarch/08padia	-Search of vin/test/oronartyEnrich ofb/SanMatanCounty	Common Properties
		(optional)
Common Properties (optional)		(opnorial)
dop:populationUensity5dM(1)	^	No description available
Deptomezone(1)		ite description enditation
diputitization departure	P1	
dowidiReceID(1)		
C doc.mor agenc(1)		
docrareal/(starSch0(1)		
depresenter(1)		
Thomak ParaPavision ID(1)	-	
4	m +	
Select All Unselect All	Add Value	
Get Inverse Common Properties (optional)		
Inverse Common Properties (optional)		
is dbo:sourceMountain Of (1)		
is dbo:origin Of (1)		
is dbo:locationOty Of (1)		
is dbo:deathPlace Of (1)		
is dbo:wikiPageRedirects Of (1)		
is dbp:county Of (1)		
is dbp:territory Of (1)	E	
is dbomouthPlace Of (1)		
is dbp:headquarters Of (1)	•	
	m ,	
Select All Unselect All	Add Value	
Get Expanded Common Properties by following Trans	stive Inverse Partonomical and Subdivision Paths (optional)	
Expanded Common Properties (optional)		
dbp:areaLandSqMi(33)	*	
dbo:populationTotal(33)		
dbp:areaWaterSqMi(33)		
dbo:thumbnai(33)		
coopropulatedPlace/areaTotal(33)		
g abosarea lota(33)		
dop-population ( otal (33)		
ubulareawater(33)	*	
( uppareavraueron(53)	m	
Select All Unselect All	Add Value	
Get Inverse Expanded Common Properties (optional)		
Inverse Expanded Common Properties (optional)		
is dbo:wikiPageRedirects Of (35)	×	
is dbo:wikPageDisambiguates Of (27)		
is dbo:location Of (23)		
is dbo:deathPlace Of (20)		
is dbo:city Of (20)	-	-
	OK Canal Environmente << little bielo	Tool Help
	Contra Contra Contraction Contraction	iournep

Method	
0000000	

# Non-functional Property Conversion

#### Non-functional Property Conversion



How many people were born in Berlin?

Input wikidata location entities Feature Class		
E:/UCSB_STKO_Lab\STKO Research\researc	ch\D8pedia-Search-plugin\test\test1.gdb\city	6
List of No-Functional Properties of Current Fe	ature Class	-
http://dbpedia.org/property/areaTotalKm		
List of Related Tables		
E: (JCS8_STKO_Lab \STKO Research \research	h\D8pedia-Search-plugin\test\test1.gdb\city_wikiURL_areaTotaKm	
<ul> <li>List of Merge Rules</li> </ul>		
1		-
SUM MIN MAX STDEV MEAN COUNT FIRST LAST EGREATEMATE		

Method	
00000000	

#### Relation Exploration



Humboldt?

Method	
00000000	

#### Relation Exploration

Query for 1-, 2-, and 3-degree sister city from the input spatial entities (Santa Barbara).

Input wikidata location entities Feature Class	<u>^</u>	The third degree
E: \UCSB_STKO_Lab\STKO Research\research\DBpedia-Search-plugin\test\propertyPath_2.gdb\SB		property (optional)
Relationship Degree		
3	-	No description available
The first degree property direction		
ORIGIN	-	
The first degree property (optional)		
sister city	-	
The second degree property direction (optional)		
ORIGIN	-	
The second degree property (optional)		
sister city	-	
The third degree property direction (optional)		
ORIGIN	-	
The third degree property (optional)		
sister dty	-	
The fourth degree property direction (optional)		
	· · ·	
The fourth degree property (optional)		
Output Location		
E: \UCSB_STKO_Lab \STKO Research \research \DBpedia-Search-plugin \test \propertyPath_2.gdb	<b>2</b>	
Output Triple Store Table Name		
SBPathQueryTripleStore		
Output Feature Class Name		
SBPathQueryLocation		

Method	
0000000	

#### **Relation Exploration**

Query for 1-, 2-, and 3-degree sister city from the input spatial entities (Santa Barbara).



Esri, HERE, DeLorme, MapmyIndia, @ OpenStreetMap contributors, and the GIS user community

Introduction	Метнор	Discussion	
00000000	00000000	•00	

# LINKED DATA CONNECTOR

- The presented Linked Data connector demonstrates a workflow for geographic Linked Data retrieval, attribute enrichment & conversion, and linkage exploration within a GIS.
- As far as we know, this is the first work about integrating Linked Data back to a GIS and making them ready for spatial analysis that does not simply flatten the data.
- The only system that supports a subset of the presented capabilities is the GeoEnrichment<sup>2</sup> service recently developed by Esri.

<sup>2</sup>https://developers.arcgis.com/rest/geoenrichment/api-reference/ geoenrichment-service-overview.htm

	DISCUSSION	
	000	

#### ESRI'S GEOENRICHMENT



#### The Enrich Layer toolbox in ArcGIS Pro.

4	OBJECTID	Stope	ld .	Label	URL	POINT_X	POINT_Y	sourceCountry	ORIG, D	areaType	buffer(	buffefA	bufferRadii	2000 Tetal Households	2000 Total Housing Units	2000 Total Population
		Point	1	San Francisco	http://www.wikidata.org/entity/Q62	-122.433333	37,756667	US		DriveTimeBuffer	esiMies	Miles	10	308675	407068	950993
	2	Point		Brisbane .	http://www.wikidata.org/entity/Q917671	-122,419167	37.680833	us	2	DriveTimeBatter	estMiss	Miles	10	403367	420446	1009438
	3	Point		Dely City	http://www.wikidata.org/entity/Q370825	-122.468333	37.686389	US	3	DriveTimeBatter	esiMies	Miles	10	425186	443492	1061784
	4	Point	1	Colma	http://www.wikidata.org/entity/Q987296	122,455556	37.678889	US	- 4	DriveTimeBatter	eriMie	Miles	10	416617	433722	1046568
	5	Point	1	South San Francisco	http://www.wikidata.org/entity/Q827122	-122.425556	37.656111	US	5	DriveTimeBuffer	siMe	Miles	10	279948	283756	777603

The enriched attribute table after executing the Enrich Layer toolbox.

Introduction	Метноd	Discussion	
00000000	00000000	00●	

# Comparison

Comparing GeoEnrichment Service with Linked Data connector:

- Well-curated dataset v.s. distributed, Web-scale, real-time knowledge graph
- Points-driving time zones-attributes v.s. points-spatial entities-properties
- The GeoEnrichment service supports datatype properties but not object type properties such as those linking actors, places, events, and objects together.
- From a question answering perspective, our Linked Data connector framework is more suitable to answer geographic queries such as about the oldest mission along California's coast.

Introduction	Метноd	Discussion	Conclusion
00000000	00000000	000	• O

# Conclusion

- We proposed and implemented a Linked Data connector workflow to deeply integrate Linked Data and GIS without simply flattening the retrieved data.
- Our system constantly creates new tables and schema for them, thereby enabling users to truly navigate the link structure of knowledge graphs and to query each node's datatype and object-type properties on-the-fly.
- We compare our framework with GeoEnrichment and show that both approaches are complementary.

Introduction	Метноd	Discussion	Conclusion
00000000	00000000	000	O

# Future Work

- We see the presented research as a starting point towards a more question answering oriented view on GIS.
- We have not addressed issues of scalability, user feedback and exception handling, provenance records, and so forth, but believe that they will be important steps towards a deployable add-on to GIS systems.