

The KnowWhereGraph



Pascal Hitzler

Data Semantics Laboratory (DaSe Lab)
Kansas State University

<http://www.daselab.org>

Data Semantics Lab, Kansas State University

- in Manhattan, Kansas
- <https://daselab.cs.ksu.edu/>
- 2 postdocs, 8 PhD students



- (Modular) Ontology Engineering and Knowledge Graphs
- Neuro-Symbolic Artificial Intelligence
- Explainable AI
- Knowledge Representation and Reasoning
- Application Partners: Digital Humanities, Agriculture, GeoSciences, Cognitive Science, Oceanography, ...

KnowWhereGraph



- 2 years, \$5M. Follows a \$1M, 1-year pilot.
- NSF “Open Knowledge Networks” (OKN) program. 21 phase 1 projects; 5 phase 2 projects.

Team and Partnership

PI: **Krzysztof Janowicz, UCSB**

Co-PIs: **Mark Schildhauer, Wenwen Li, Dean Rehberger, Pascal Hitzler**



KnowWhereGraph



Team



Shirley Ambrose
Ph.D. Student
ASU



Randall Barker
Senior Personnel
IN/IT



Bruno Basco
Senior Personnel
MSU



Ty Fitzpatrick
Senior Personnel
E&I



Catherine Foley
Senior Personnel
MSU



Sells Gonzalez
Lead of Software Development
M&E
Michigan State University



Paulina Oliva
Senior Personnel
UCB



Scott Robinson
Partner
Oliver Wyman's Commodity and Risk
Practices



Andrew Schroeder
VP of Research and Analysis
Direct Relief



Lynn Ueary
Senior Scientist
U.S. Geological Survey



Daria Vesselska
Research Scientist
U.S. Geological Survey



Sulu Wang
Ph.D. Student
The School of Computing, Informatics,
and Decision Systems Engineering
Arizona State University



Ling Cai
Student at STKO Lab
Department of Geography
UC Santa Barbara



Frank Davenport
Research Scientist
UCSD/Climate Research Center



Anthony D'Onofrio
Developer
MSU



Tony Hessemer
Senior Personnel
UCSB



Matt Jones
Senior Personnel
UCSB



Zilong Liu
Graduate Student
UCSB



Alica Shell
Current Manager
M&E
Michigan State University



Cogan Shimizu
Ph.D. Student
K-State



David Smith
Ph.D. Student
Oliver Wyman
New York



Zhanyu Wang
Graduate Student
UCSB



Dawn Wright
Ph.D. Student
School and Professor of Geography and
Geoinformatics
Oregon State University



Joseph Zalawski
Ph.D. Student
Civil and
Kansas State University



Karen Doehner
UCSB



Andrew Edis
Research Analyst
DataLab
Kansas State University



Colby Fisher
Senior Personnel
Oliver Wyman



Anna Lopez-Carr
Monitoring and Evaluation Specialist
Research and Analysis Group
Direct Relief



Gengchen Msi
Ph.D. Candidate
Space and Time for Knowledge
Cooperation Lab
Department of Geography
UC Santa Barbara



Tim Murphy
Senior Personnel
E&I



Thomas Theisen
Senior Personnel
Geography
UCSB



Yuanxuan Tian
Ph.D. Student
Geography
School of Geographical Sciences and
Urban Planning
Arizona State University



Austin Truchan
Developer
MSU



Lu Zhou
Ph.D. Student
K-State



Rui Zhu
Ph.D. Student
UCSB



Mike Mathews
Oliver Wyman

PIs



Krzysztof Janowicz
Principal Investigator
GeoInformatics
UC Santa Barbara



Pascal Hitzler
Computer Science
Kansas State University



Mark Schildhouse
Environmental Science
NCEAS UC Santa Barbara



Wenwan Li
Urban Planning and Geoinformatics
Arizona State University



Dean Rehberger
Hardware
Michigan State University



(some) project goals

- pushing the state of the art in spatiotemporal Knowledge Graph (KG) engineering
- transfer of KG technology towards adoptable practice
- application showcases



Addressing the bottleneck in data science:

**80% is data processing
20% is deriving insights**

<http://KnowWhereGraph.org/>

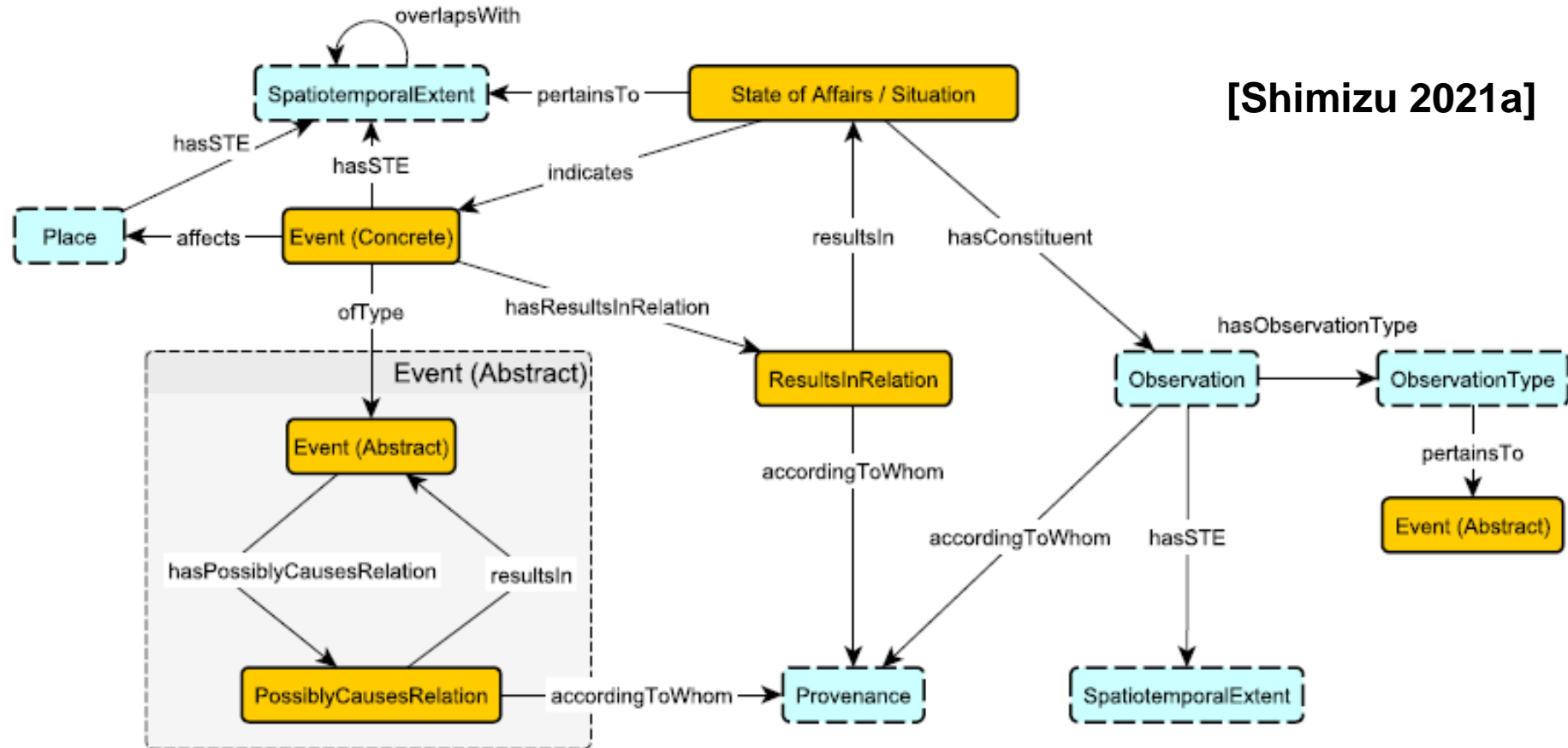
- 1. Schema design**
- 2. The graph and geo-enrichment**
- 3. Application use cases**

addressing cost and re-use [Shimizu 2021]



- **principled use of schema diagrams**
- **use of existing ontology design patterns (or make new ones)**
- **modular schema design**
 - **a module corresponds to a concept meaningful for a human**
 - **modules remain identifiable in the final schema**
 - **future changes remain “local” in the module sense**
 - **departures from ontology tradition:**
 - **modules are central, taxonomies are not**
 - **ontology shall not define, but constrain**

Causal Relations Pattern



InK Browser [Zalewski 2021a]



← → ↻ 🏠 127.0.0.1:5000/client ☆ 🔍 Search

InkBrowser

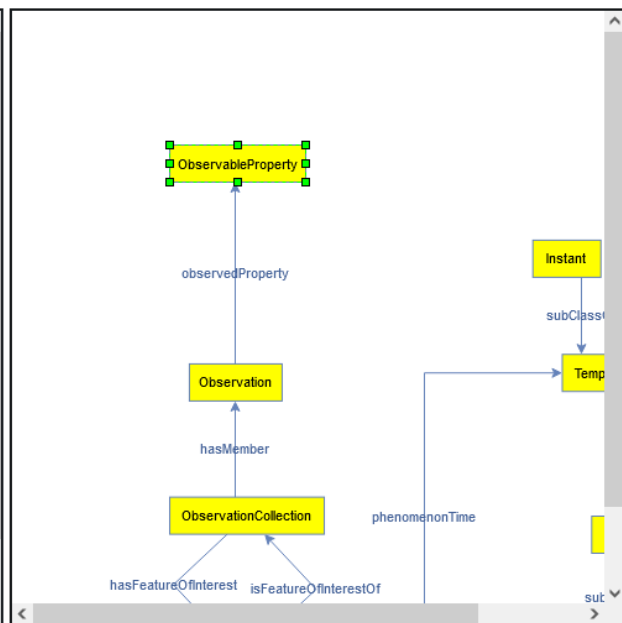
Class Hierarchy Client Statistics Search

Click Here

- <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.depth>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.depthError>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.dmir>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.gap>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.horizontalError>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.locationSource>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.mag>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.magError>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.magSource>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.magSubSource>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.net>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.nst>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.rms>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.statu>
 - <http://stko-roy.geog.ucsb.edu/od/resource/observableproperty.type>
- Load More Results

Predicate	Object
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/ns/sosa/ObservableProperty
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#NamedIndividual
http://www.w3.org/2000/01/rdf-schema#label	Horizontal distance from the epicenter to the nearest station (in degrees). 1 degree is approximately 111 kilometers. In general, the smaller this number, the more reliable is the calculated depth of the earthquake.

Load More Results



Look up a Schema Diagram:



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

CoModIDE (Protege plug-in) [Shimizu 2021]

The screenshot displays the CoModIDE Protege plug-in interface. The main window shows an ontology graph with classes (Person, Student, Teacher, Lecture, Course) and properties (dateOfBirth, name, attends, teaches, manages, startsAt, endsAt, awardsCredits). A red circle '1' highlights the 'Class' icon in the 'Core constructs' panel. A red circle '2' highlights the 'Identifier' pattern in the 'Pattern Library'. A red circle '3' highlights the 'Internal (keeping parent onto...)' option in the 'Module annotations placement' section of the 'CoModIDE Configuration' panel.

schoolExample (https://karlhammar.com/ontologies/schoolExample) : [/Users/karl/Desktop/SchoolExample.owl]

Active ontology x Entities x Individuals by class x DL Query x CoModIDE x

CoModIDE Renderer: Core constructs XSD datatypes

Class Property Subclass

232, 563

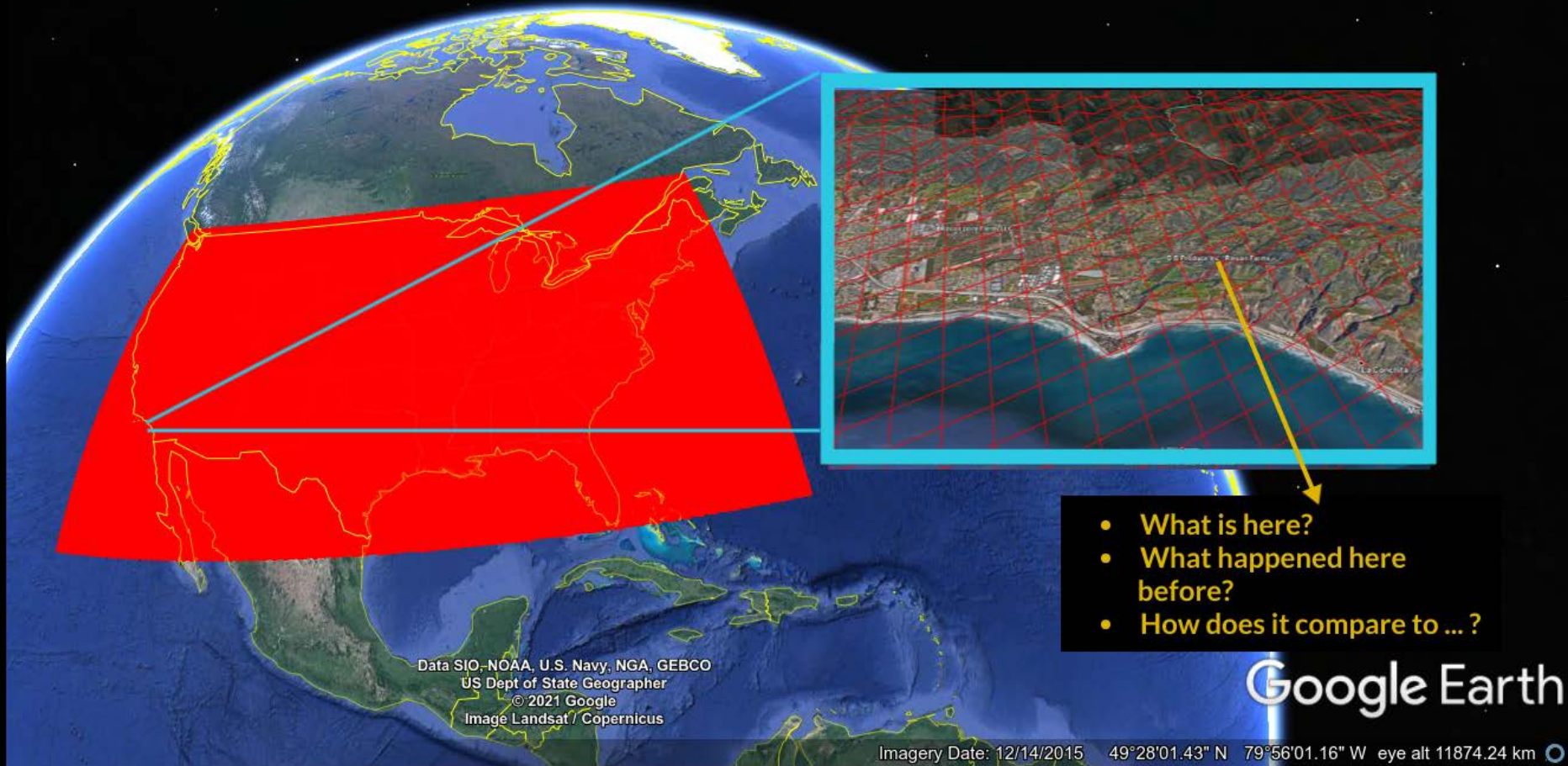
CoModIDE Pattern Selector: Pattern Library Pattern Category Sele... Any Patterns: Name Agent Role Agent Role Documentation Aggregation, Bag, ... Documentation Event Event Documentation Explicit Typin... Documentation Identifier Identifier Documentation Name Stub Name Stub Documentation Participant Role Participant Role Documentation Property Reification Property Reification Documentation Provenance Provenance Documentation Quantities and Units Quantities and Units Documentation Sequence, List Sequence, List Documentation Spatial Extent Spatial Extent Documentation Spatiotemporal Ex... Spatiotemporal Ex... Documentation Stubs Stubs Documentation Temporal Extent Temporal Extent Documentation

CoModIDE Configuration: Entity naming: Use target namespace Keep pattern namespace Module annotations placement: External (keeping parent onto... Internal (keeping parent onto...) Edge creation: RDFS Range RDFS Domain AllValuesFrom constraint SomeValuesFrom constraint

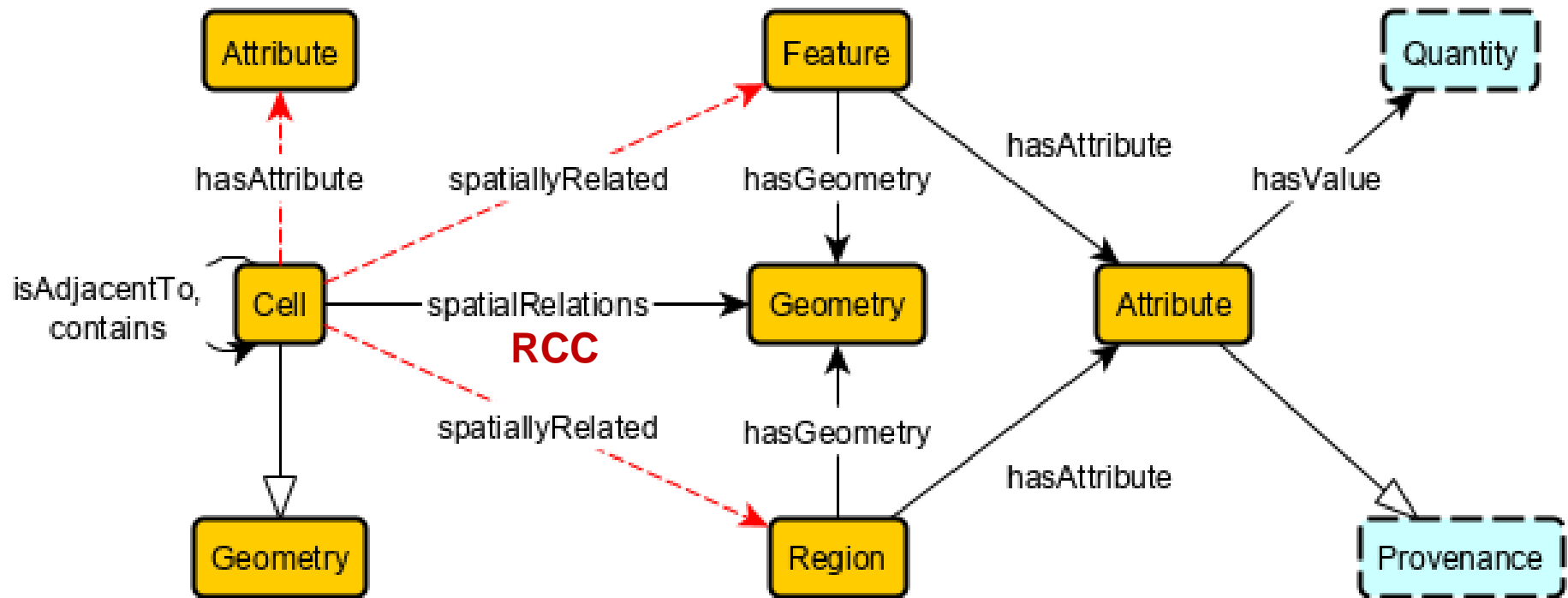
To use the reasoner click Reasoner > Start reasoner Show Inferences

Cells (S2)

Features as Collections of Cells in a Discrete Global Grid



Cells (S2)

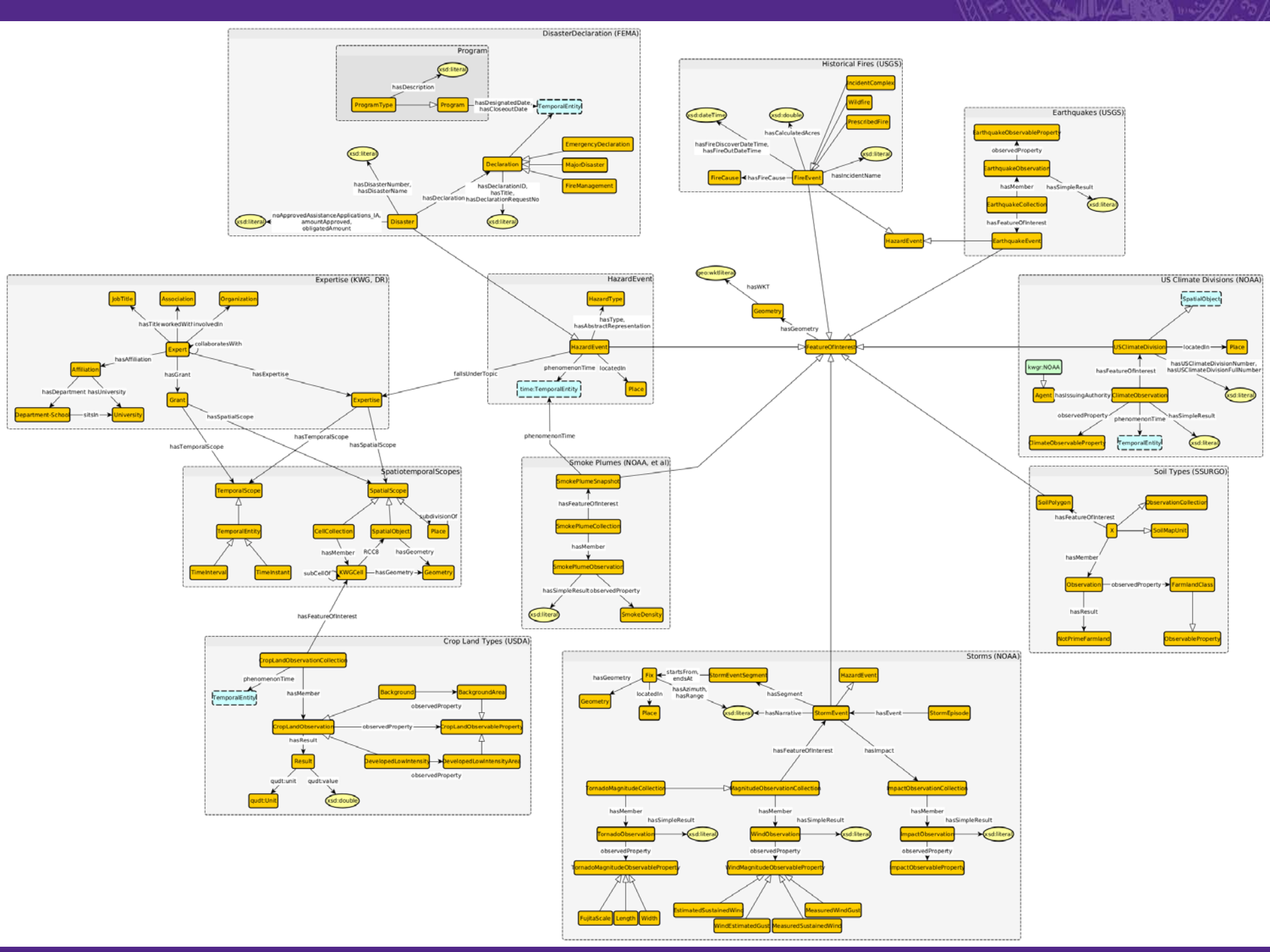


Cells and Region Connection Calculi



[Zalewski 2021]

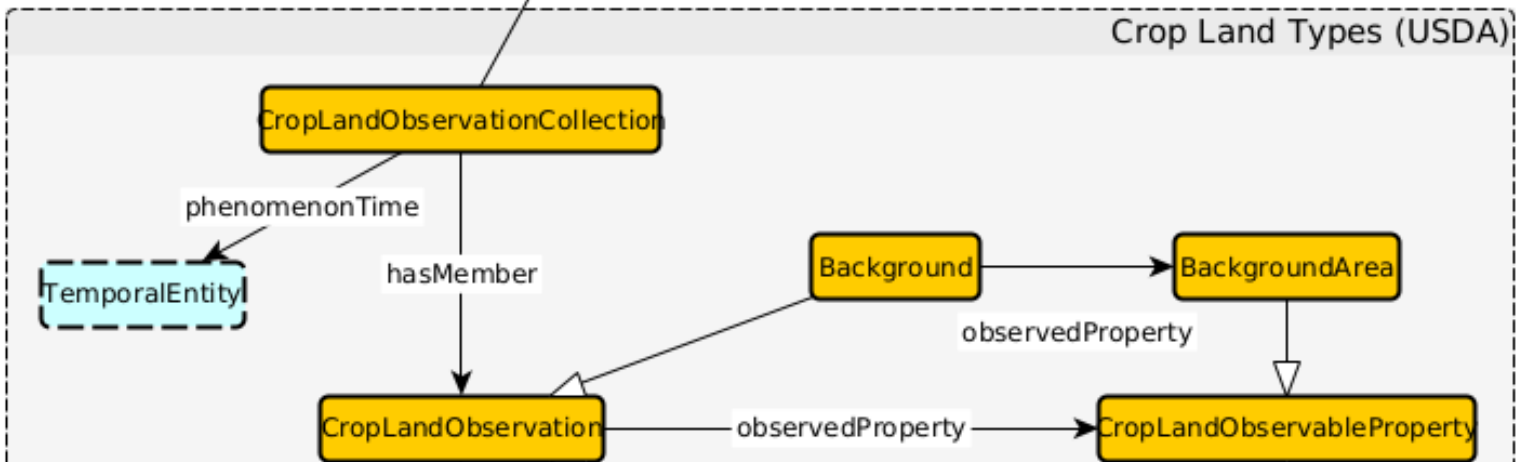
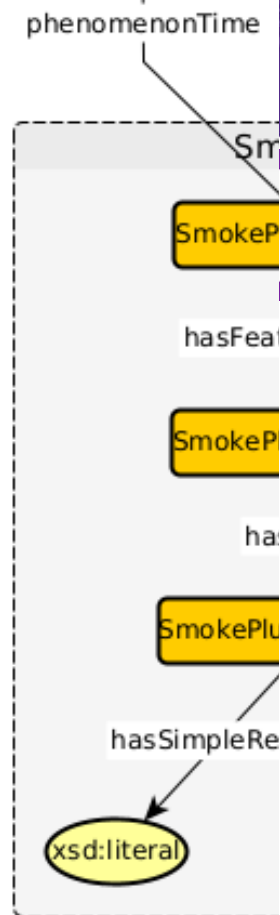
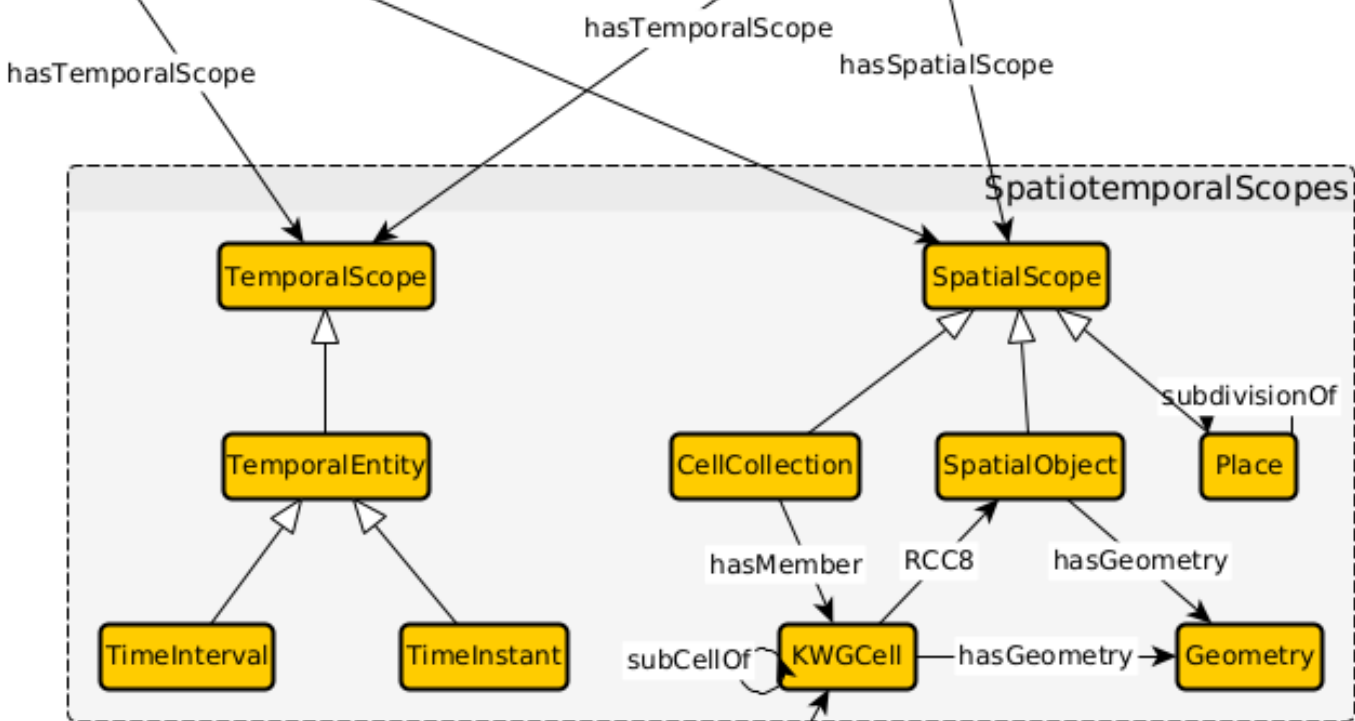
- **information inheritance (logical inference) along cell hierarchy: compressed data representation**
- **uniform graph representation of spatial data (using cells) for better querying**
- **graph representation that does not necessitate repeated costly calculations of region intersections.**
- **Comes with some loss of spatial accuracy, depending on cell level used.**



addressing waterfall design issue



- **we work both top-down (design first) and bottom-up (graph construction first) – but it's the same group of people.**
- **that way, perspectives converge over time, but we gain a stronger schema**



1. Schema design
2. **The graph and geo-enrichment**
3. Application use cases

The graph



- currently over 1B triples
- target 10B

Thematic Datasets					Place-Centric Datasets		
Dataset Name/ Theme	Source Agency	Key Attributes	Spatial Coverage	Temporal Coverage	Place-Centric Dataset	Defining Authority	Spatial Coverage
Soil Properties	USDA	soil type, farmland class	Targeted regions in US	Current	S2 Cells	Google	Lvl 9 (Global), Lvl 11 (US), Lvl 14 (CA)
Wildfires	USGS, USDA, USFS, NIFC	wildfire type, burn severity, num. acres burned, contained date	US	1984–current	Global Administrative Areas	University of Berkeley, Museum of Vertebrate Zoology and the International Rice Research Institute	Global
Earthquakes	USGS	magnitude, length, width, geometry	Global (mag. over 4.5)	2011-01-01 to 2021-02-10			
Climate Hazards	NOAA	injuries, deaths, property damages	US	1950–2020			
Expert - Covid-19 Mobility	Direct Relief (DR)	name, affiliation, expertise	Global	2021	US Federal Judicial District	DoJ, ESRI	US
Expert - General	KWG, UC System, DR	name, affiliation, expertise with spatiotemporal scopes	Global	unlimited	CodesNational Weather Zones	NOAA	US
Cropland Types	USDA	crop types (raster data)	CA (US)	2020	FIPS Codes	NRCS	US
Air Qual. Obs.	U.S. EPA	AQI value, CO concentration	US	1980–2021	DMA	Nielen	US
Smoke Plumes	NOAA	daily smoke plumes extent	US	2010-2021	ZIP	ZCTA	US
Climate Observations	NOAA	temperature, precipitation, PDSI, PHSI	US	1950 - 2020	Climate Division	NOAA	US

Graph Utilization

Graph-based GeoEnrichment in ArcGIS Pro

The screenshot displays the ArcGIS Pro interface with a map of the Santa Fe region. A blue polygon highlights a specific area on the map. The Geoprocessing pane on the right shows the 'Linked Data Relationship Finder' tool with the following parameters:

- Enter a SPARQL endpoint which supports GeoSPARQL: `http://stko-roy.geog.ucsb.edu:7200/repositories/kwg-seed-graph`
- Input geographic entities Feature Class: `soil`
- Relationship Degree: `2`
- The first degree property direction: `ORIGIN`
- The first degree property: `kwg-ontaffectedBy`
- The second degree property direction: `ORIGIN`
- The second degree property: `kwg-ontlocatedIn`
- Output Location: `kwg-test.gdb`
- Output Triple Store Table Name: `soilPathQueryTripleStore`
- Output Feature Class Name: `soilPathQueryGeographicEntity`

The table at the bottom shows the results of the query:

OBJECTID	Subject	Predicate	Object	Pred_Label	Degree
1	http://stko-roy.geog.ucsb.edu/od/resou...	http://stko-roy.geog.ucsb.edu/od/ontology/affect...	http://stko-roy.geog...	kwg-ontaffectedBy	1
2	http://stko-roy.geog.ucsb.edu/od/resou...	http://stko-roy.geog.ucsb.edu/od/ontology/locate...	http://www.wikidata...	kwg-ontlocatedIn	2
3	http://stko-roy.geog.ucsb.edu/od/resou...	http://stko-roy.geog.ucsb.edu/od/ontology/affect...	http://stko-roy.geog...	kwg-ontaffectedBy	1
4	http://stko-roy.geog.ucsb.edu/od/resou...	http://stko-roy.geog.ucsb.edu/od/ontology/locate...	http://www.wikidata...	kwg-ontlocatedIn	2
5	http://stko-roy.geog.ucsb.edu/od/resou...	http://stko-roy.geog.ucsb.edu/od/ontology/affect...	http://stko-roy.geog...	kwg-ontaffectedBy	1
6	http://stko-roy.geog.ucsb.edu/od/resou...	http://stko-roy.geog.ucsb.edu/od/ontology/affect...	http://stko-roy.geog...	kwg-ontaffectedBy	1
7	http://stko-roy.geog.ucsb.edu/od/resou...	http://stko-roy.geog.ucsb.edu/od/ontology/locate...	http://www.wikidata...	kwg-ontlocatedIn	2

- We are currently moving from ArcGIS Pro toolboxes to a full **ArcGIS Pro add-in**
- Interfaces for **QGIS**, **R**, etc.
- **Custom** use case-driven solutions, e.g., for Direct Relief, FMI, and FC

1. Schema design
2. The graph
3. **Application use cases**

Our Applications



Farm to Table Supply Chain & Sustainability

Enhance the sustainability, efficiency, and safety of consumer food supply in collaboration with the *Food Industry Association*.



Land Valuation and Risk of Default

Driver-based land potential assessment for model based valuation and risk assessment for agricultural credit applications & loan portfolio monitoring.



Humanitarian Aid

Apply our technologies to the humanitarian supply chain needs of *Direct Relief* during the COVID-19 crisis and help them to find experts.

Conclusions

- Lowering cost of spatiotemporal KG engineering
- Modular, extendable KG
- GIS interface to add KG-based data as needed



Our Value Proposition

KnowWhereGraph aims at **providing area briefings within seconds** for any region on Earth to answer questions such as

- *"What is here?"*
- *"What happened here before?"*
- *"Who knows more?"*
- *"How does it compare to other regions or previous events?"*

By doing so, we assist decision-makers and data scientists in rapidly enriching their data with millions of connected, up-to-date facts at the human-environment interface to gain the situational awareness required for good decision-making.

Our ultimate goal is to flip the 80/20 bottleneck on its head.



Thanks!

References (available online)

[Hitzler 2021] Hitzler, A review of the Semantic Web field. CACM 64 (2), 76-82, 2021.

[Shimizu 2021] Shimizu, Hammar, Hitzler, Modular Ontology Modeling, under review.

[Shimizu 2021a] Shimizu, Zhu, Schildhauer, Janowicz, Hitzler, A Pattern for Causal Relations Between Events, ISWC 2021 WOP.

[Zalewski 2021] Zalewski, Hitzler, Janowicz, Semantic Compression with Region Calculi in Nested Hierarchical Grids, SIGSPATIAL 2021.

[Zalewski 2021a] Zalewski, Zhou, Shimizu, Hitzler, InK Browser – The Interactive Knowledge Browser, ISWC 2021 P&D.

Zhu, Stephen Ambrose, Zhou, Shimizu, Cai, Mai, Janowicz, Hitzler, Schildhauer, Environmental Observations in Knowledge Graphs, ISWC 2021 DaMaLos.

