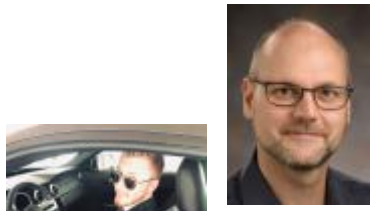
Knowledge Graphs

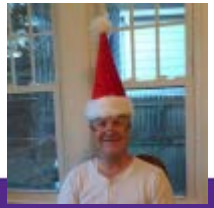
Pascal Hitzler
Data Semantics Laboratory (DaSe Lab)
Kansas State University

http://www.daselab.org

DaSeLab

- **Postdoc:**
  - **Cogan Shimizu**
- **PhD Students:**
  - **Aaron Eberhart**
  - **Abhilekha Dalal**
  - **Joseph Zalewski**
  - **Joshua Schwartz**
  - **Mohammad Saeid Mahdavinejad**
  - **Patrick Stingley**
  - **Reihaneh Amini**
  - **Rushrukh Rayan**
  - **Sanaz Saki Norouzi**
  - **Sulogna Chowdhury**
- **Undergrad Researchers:**
  - **Andrew Eells**
  - **Brayden Pankaskie**





Where (some) grad students went

- Industry
  - Amazon
  - IBM
  - Apple
  - GE Global Research
  - TigerGraph
- Academia
  - Wright State University, OH, USA
  - University of Hartford, NJ, USA
  - TU Dresden, Germany
  - IIIT Delhi, India
  - Universitas Indonesia, Jakarta
- Elsewhere
  - UN Headquarters, New York

Data, data, data!

- 2.5 * 10^18 (quintillion) bytes produced each day as of 2021
- 90% of the world's data was created in the last two years
- expected to double every two years

https://www.the-next-tech.com/blockchain-technology/how-much-data-is-produced-every-day-2019/

Data Integration and Re-use

- data coming
  - from different sources
  - in different formats
  - differently structured
- If you need to use it for a combined purpose, you need to understand and integrate this data
- which is a lot of work.

- E.g. thousands of mentions of “John Smith”
- What exactly is Kansas City? (Missouri, Kansas or metro?)
- What is a forest?
- What does wind direction mean?
- What is Q6499080?

Data management is 80% of the effort/cost of data analysis!

How to organize data to make this easier?

- formal language
- formal logic
- tables
- trees
- graphs

- Govern it all through standards and explicit agreements?
- Use of metadata? (data that explains other data)
  - wind direction is “to”
  - Kansas City typed as “metro area”

Knowledge Graphs

- **labeled directed graphs (“meaning” labels on both nodes and edges)**
- **nodes are “typed”**
- **additional complex metadata is often used but we will not get into this today ☺**

Google Knowledge Graph (live)

Laura Kelly

Governor of Kansas

Laura Kelly is an American politician serving as the 48th governor of Kansas since 2019. A member of the Democratic Party, she represented the 18th district in the Kansas Senate from 2005 to 2019. Kelly ran for governor in the 2018 election and defeated the Republican nominee, Kansas Secretary of State Kris Kobach. Wikipedia

Born: January 24, 1950 (age 69 years), New York, NY

Spouse: Ted Daughety

Party: Democratic Party

Office: Governor of Kansas since 2019

Education: Indiana University, Bradley University, Indiana University Bloomington

Children: Kathleen Daughety, Molly Daughety

hasEducation

Indiana University

iu.edu

Indiana University is a multi-campus public university system in the state of Indiana, United States. Indiana University has a combined student body of more than 110,000 students, which includes approximately 46,000 students enrolled at the Indiana University Bloomington campus. Wikipedia

Mascot: Referred to as "The Hoosiers"

Endowment: 1.986 billion USD

Students: 110,436 university-wide

President: Michael McRobbie

Academic staff: 8,733 university-wide

Subsidiaries: Indiana University Bloomington, MORE

hasPresident

Michael McRobbie

President of Indiana University

president.iu.edu

Michael Alexander McRobbie AO is an Australian-American computer scientist, educator and academic administrator. He became the eighteenth president of Indiana University on July 1, 2007. Wikipedia

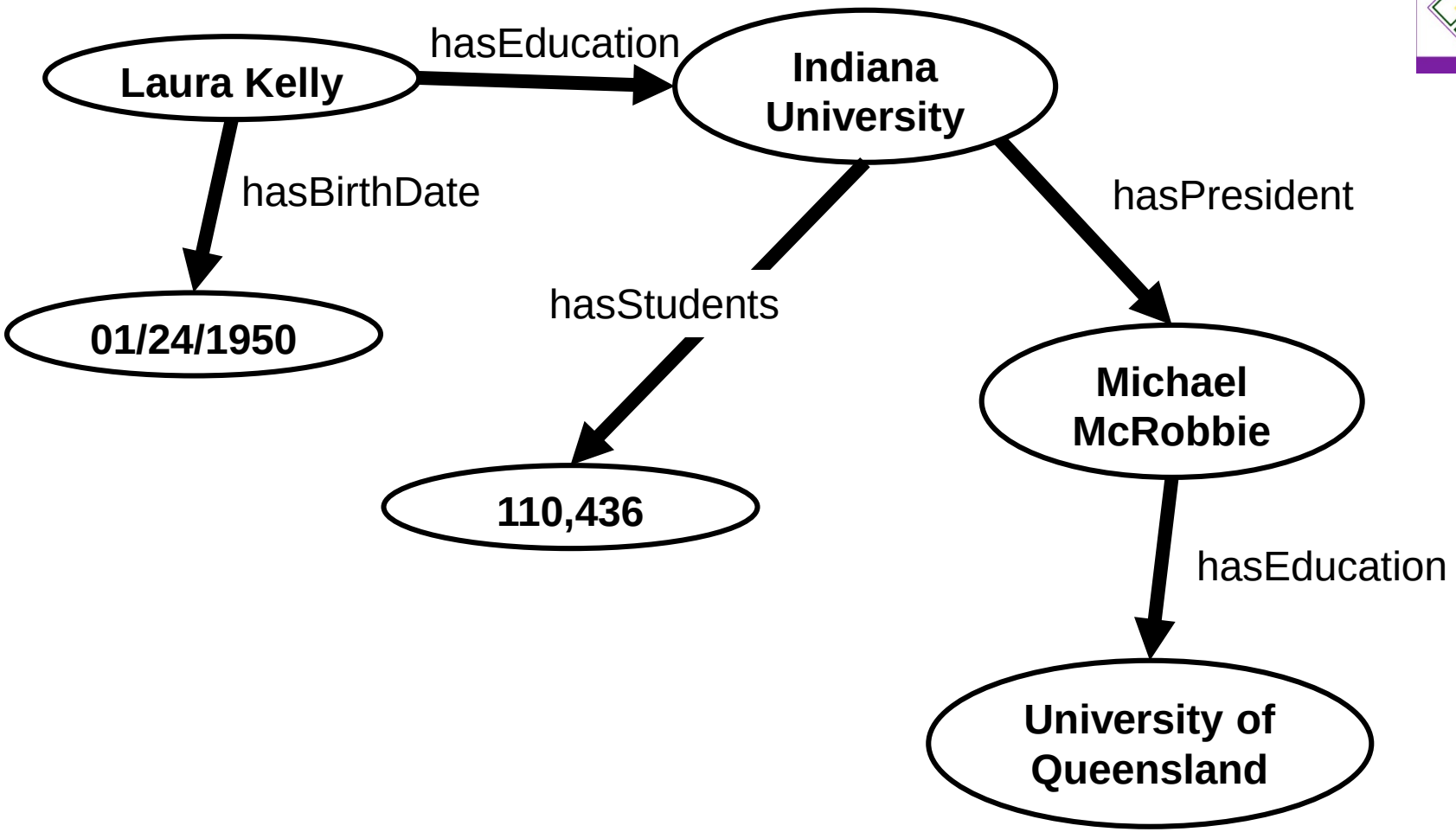
Born: October 11, 1950 (age 69 years), Melbourne, Australia

Spouse: Laurie Burns (m. 2005)

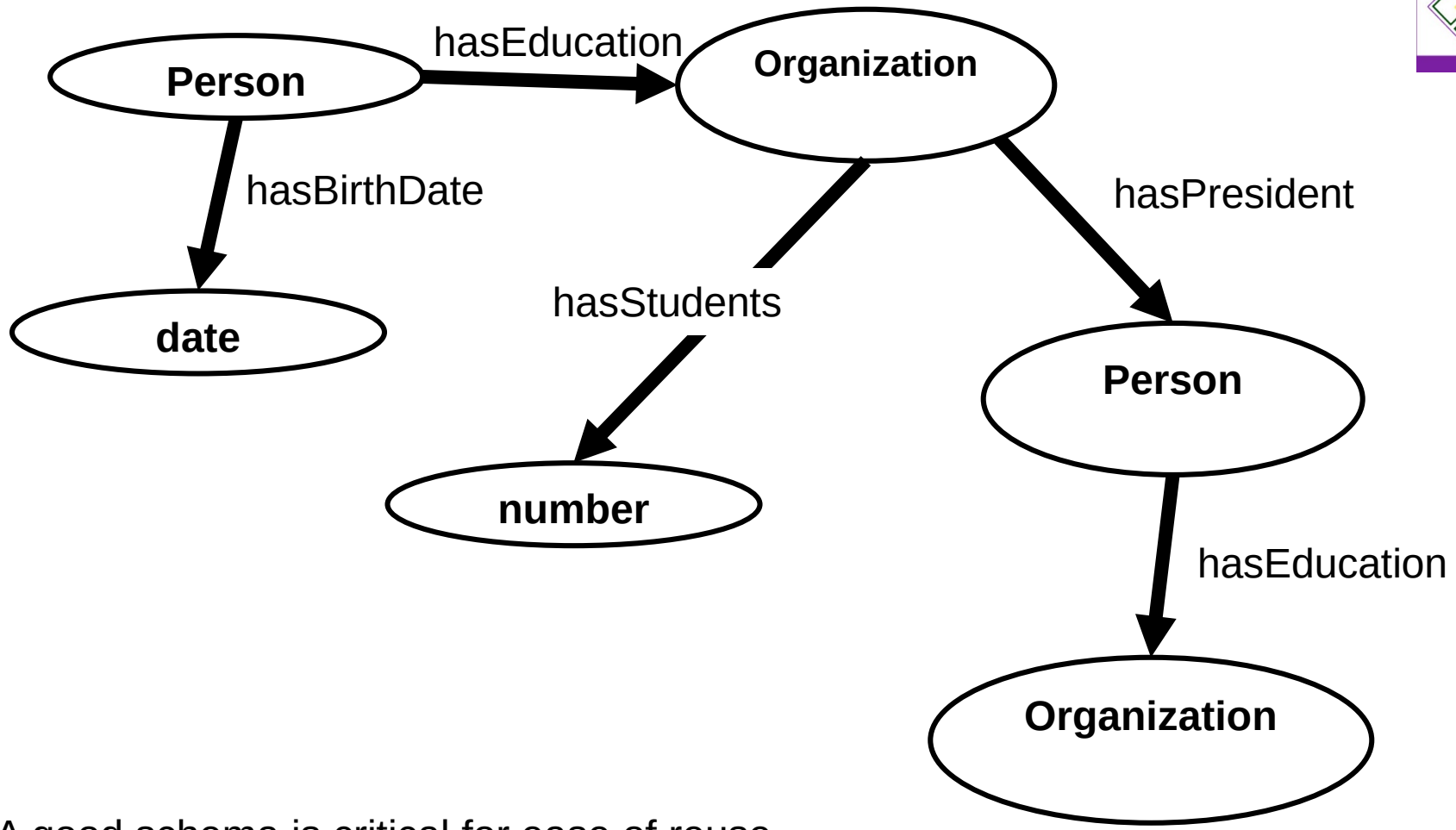
Education: The Australian National University, The University of Queensland

Books: Automated Theorem-proving in Non-classical Logics, Automated Deduction - Cade-13

Knowledge Graphs



Schema (as diagram)



A good schema is critical for ease of reuse

WIKIDATA

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data
Create a new Lexeme
Recent changes
Random Lexeme

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Wikidata item

In other projects
Wikimedia Commons
MediaWiki
Meta-Wiki
Multilingual Wikisource
Wikispecies
Wikibooks

collaborative

open

Welcome to Wikidata

the free knowledge base with 97,501,043 data items that anyone can edit.

Introduction • Project Chat • Community Portal • Help

free

multilingu

Want to help translate? Translate the missing messages.

Welcome!

Wikidata is a free and open knowledge base that can be read and edited by both humans and machines.

Wikidata acts as central storage for the **structured data** of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wiktionary, Wikisource, and others.

Wikidata also provides support to many other sites and services beyond just Wikimedia projects! The content of Wikidata is available under a free license, exported using standard formats, and can be interlinked to other open data sets on the linked data web.

Learn about data

New to the wonderful world of data? Develop and improve your data literacy through content designed to get you up to speed and feeling comfortable with the fundamentals in no time.

Item: *Earth (Q2)*

Property: *highest point*

(Live)

Laura Kelly (Q6499080)

48th and current governor of Kansas

Governor Laura Kelly | Laura Jeanne Kelly

edit

In more languages

Statements

instance of	human	edit
	1 reference	
	+ add value	
image		edit
educated at	Bradley University	edit
	academic degree: Bachelor of Science	
	academic major: psychology	
	1 reference	
	imported from Wikimedia project: English Wikipedia	
	+ add reference	
	Indiana University	edit
	academic degree: Master of Science	

Kansas State University (Q31249)

public university in the state of Kansas

Kansas State | K-State | Kansas State Agricultural College | Kansas Agricultural College | KSU

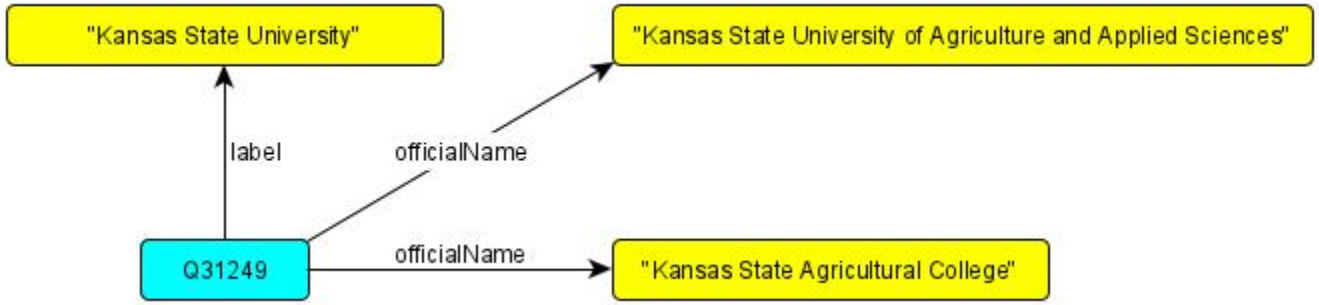
▸ In more languages

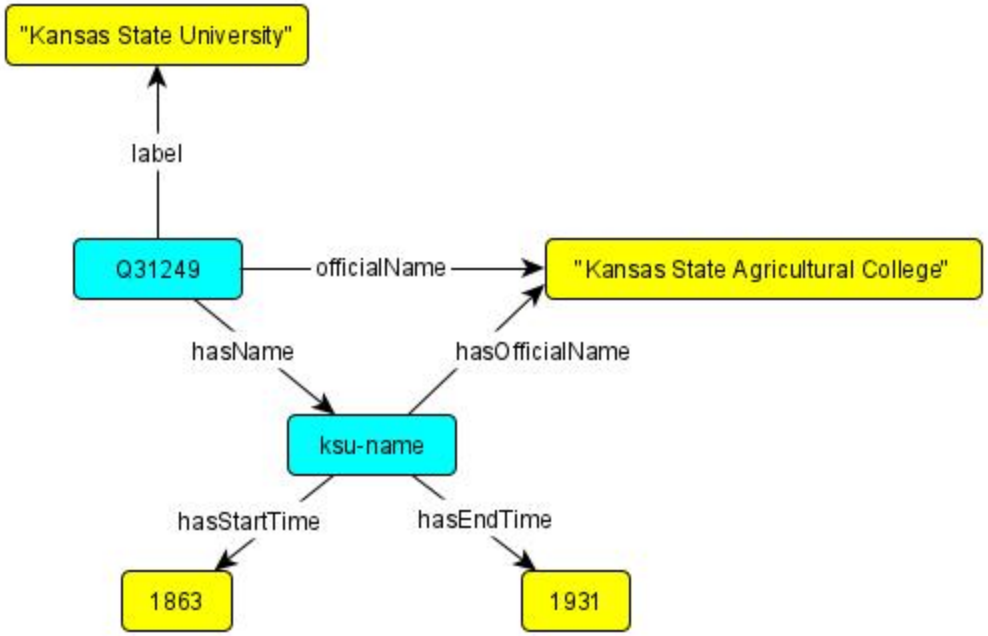
Statements

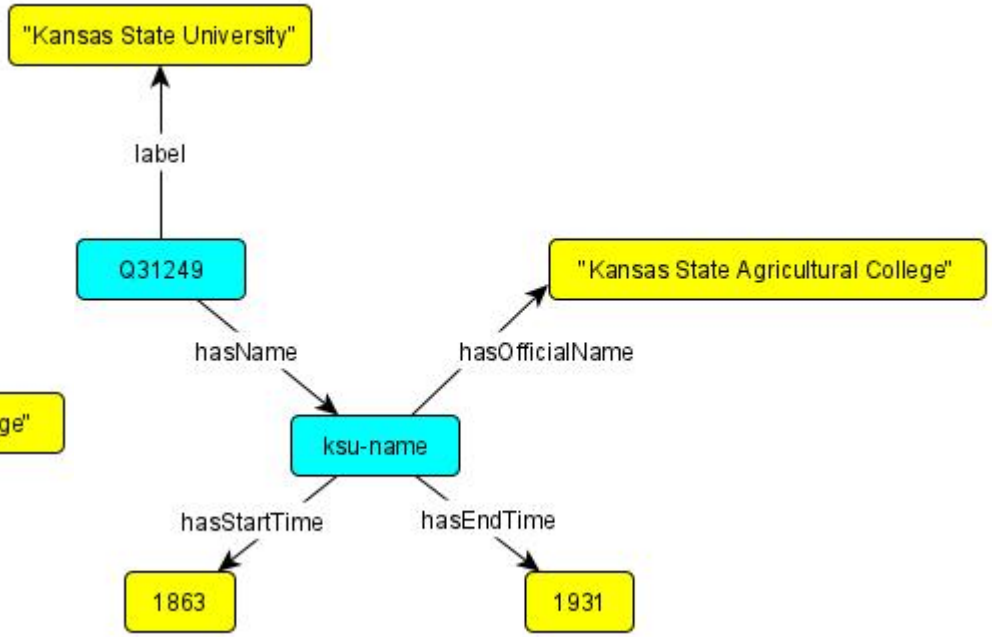
instance of	university ▾ 0 references land-grant university

official name	Kansas State Agricultural College (English) start time 1863 end time 1931 ▾ 0 references
	Kansas State College of Agriculture and Applied Science (English) start time 1931 end time 1959 ▾ 0 references
	Kansas State University of Agriculture and Applied Science (English) start time 1959

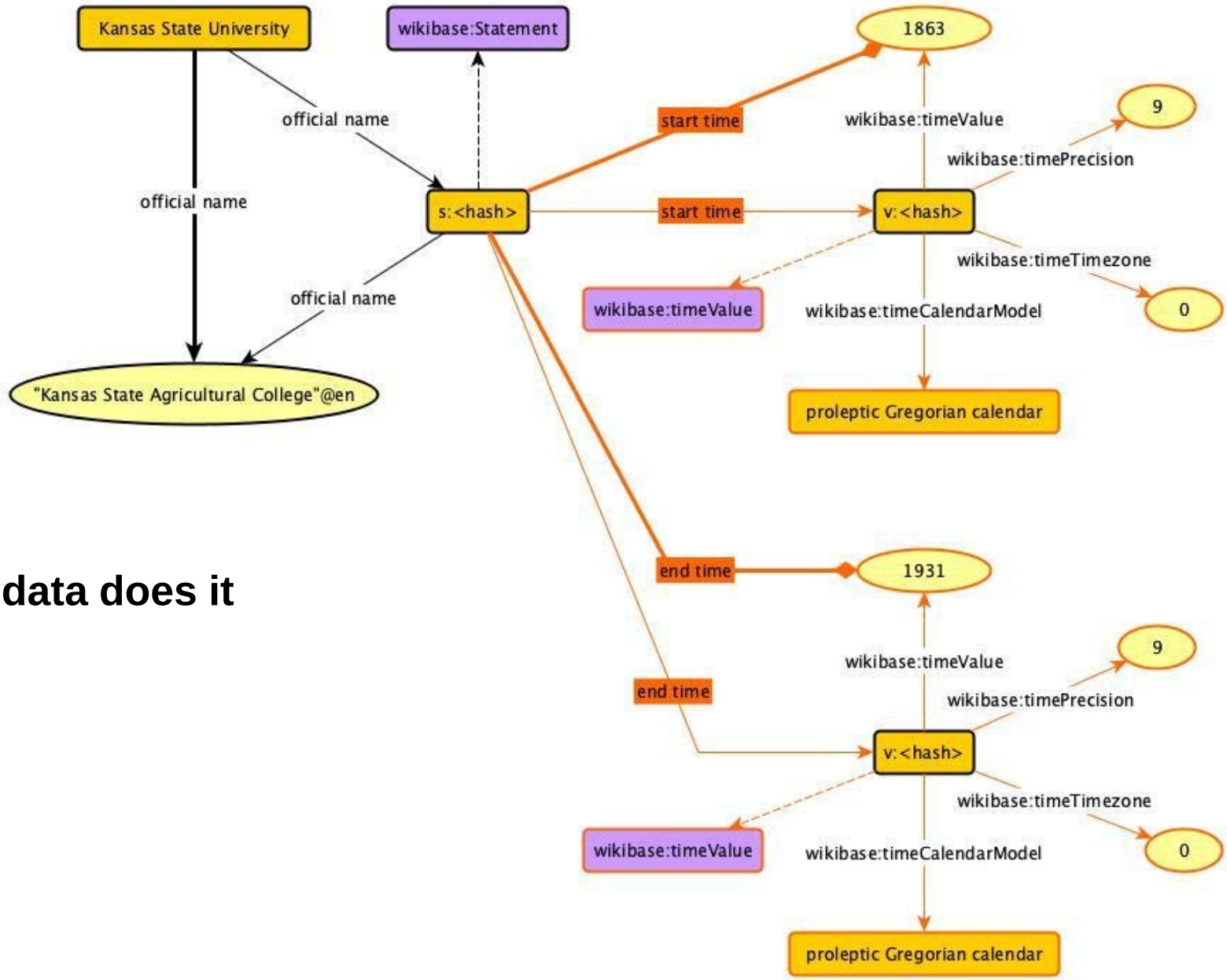
How to present context?







How to present context?



As Wikidata does it

Authors affiliations include: Google, Microsoft, IBM, Facebook, eBay

PRACTICE

Industry-Scale Knowledge Graphs: Lessons and Challenges

By Natasha Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, Jamie Taylor

Communications of the ACM, August 2019, Vol. 62 No. 8, Pages 36-43
10.1145/3331166

Comments

Credit: Adempercem / Stutterstock

Knowledge graphs are critical to many enterprises today: They provide the structured data and factual knowledge that drive many products and make them more intelligent and "magical."

In general, a knowledge graph describes objects of interest and connections between them. For example, a knowledge graph may have nodes for a movie, the actors in this movie, the director, and so on. Each node may have properties such as an actor's name and age. There may be nodes for multiple movies involving a particular actor. The user can then traverse the knowledge graph to collect information on all the movies in which the actor appeared or, if applicable, directed.

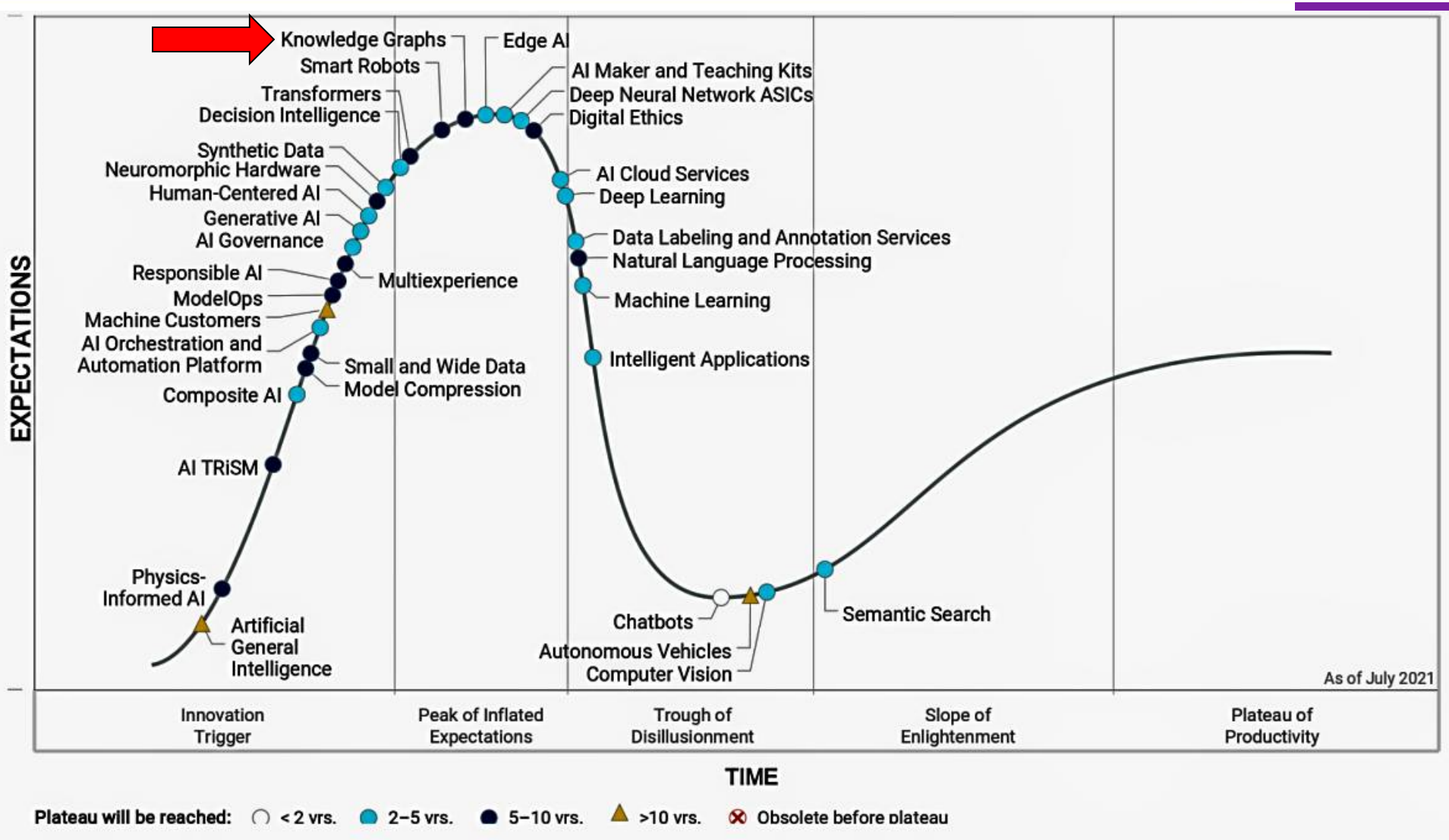
Many practical implementations impose constraints on the links in knowledge graphs by defining a *schema* or *ontology*. For example, a link from a movie to its director must connect an object of type Movie to an object of type Person. In some cases the links themselves might have their own properties: a link connecting an actor and a movie might have the name of the specific role the actor

SIGN IN for Full Access
User Name
Password
» Forgot Password?
» Create an ACM Web Account
SIGN IN

ARTICLE CONTENTS:
Introduction
What's In a Graph? Design Decisions
Challenges Ahead
Other Key Challenges
Conclusion
References
Authors

MORE NEWS & OPINIONS
MIT Robot Could Help People

Gartner, 2021



Knowledge Graph Standards

W3C Recommendation

RDF 1.1 Concepts and Abstract Syntax

W3C Recommendation 25 February 2014

This version:
http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/
Latest published version:
http://www.w3.org/TR/rdf11-concepts/
Previous version:
http://www.w3.org/TR/2014/PR-rdf11-concepts-20140109/
Previous Recommendation:
http://www.w3.org/TR/rdf-concepts
Editors:
Richard Cyganiak, DERI, NUI Galway
David Wood, 3 Round Stones
Markus Lanthaler, Graz University of Technology

DaSe Lab

Languages based on formal logic allow for automated (deductive) reasoning.
Corresponding algorithms are mathematically sophisticated and require formal correctness and complexity assessments.

Also:
The Standards need improvements!

W3C Recommendation

OWL 2 Web Ontology Language Primer (Second Edition)

W3C Recommendation 11 December 2012

This version:
http://www.w3.org/TR/2012/REC-owl2-primer-20121211/
Latest version (series 2):
http://www.w3.org/TR/owl2-primer/
Latest Recommendation:
http://www.w3.org/TR/owl-primer
Previous version:
http://www.w3.org/TR/2012/PER-owl2-primer-20121018/
Editors:
Pascal Hitzler, Wright State University
Markus Krötzsch, University of Oxford
Bijan Parsia, University of Manchester
Peter F. Patel-Schneider, Nuance Communications
Sebastian Rudolph, FZI Research Center for Information

Plenty of open questions

- What makes good knowledge graphs?
- What are good processes and tools for making them?
- What are strong intelligent algorithms for managing them, including
  - Automatic construction
  - Integration
  - Querying
- How do I make them self-explanatory?
- How do I use them in or with intelligent systems?
- What is the underlying theory/mathematics of the representation languages and (complex) algorithms?

Key research question (knowledge graphs)

- **Data management (discovery, integration, publishing, re-use) is a major cost factor in data-intensive applications.**
  - **In particular, if data is multi-sourced and heterogeneous.**

- **How can we save effort and cost for this data management?**

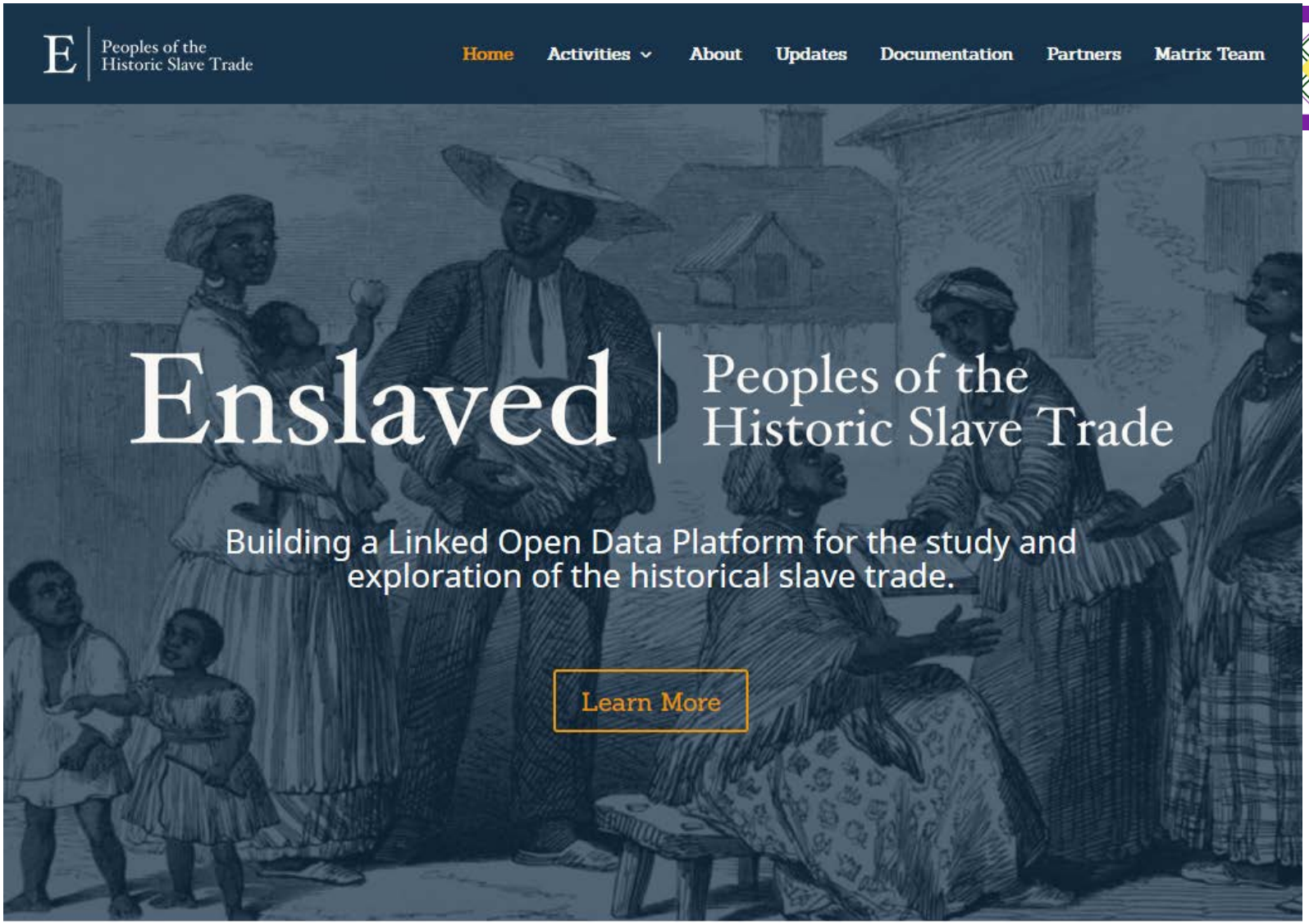
- **Research premise:**

  The principled use of Smart Data (knowledge graphs and ontologies) saves effort and cost.

  But how to exactly apply these methods best?

enslaved.org

https://enslaved.org/ https://lod.enslaved.org



DaSe Lab

enslaved.org process

1. **Quality Graph Design.**
2. **Realization in Wikibase. (Engine for Wikidata)**
3. **Knowledge graph construction and interaction through Wikibase as.**
4. **Additional front-end (simplified view)**

DaSe Lab

(4) https://enslaved.org/

(3) https://lod.enslaved.org/

>53M RDF triples from Wikibase export

PEOPLE

Gender

Age Category

Ethnodescriptor

Role Types

Occupation

Status

EVENT

Event Type

Date

PLACE

Place Type

Showing 20 of 552009 Results

Sort By

20 Per Page

People

Ardealah

Sex Female

Person Status Enslaved Person

Place Multiple

Person's Connections

4 3 3

Chemorowee

Sex Female

Person Status Enslaved Person

Place Multiple

Person's Connections

4 3 3

Allarbah

Sex Female

Person Status Enslaved Person

Place Multiple

Person's Connections

4 3 3

Arnahyajumah

Sex Female

Person Status Enslaved Person

Place Multiple

Person's Connections

4 3 3

Sannom (Q358958)

LSD-PER-075163
LSD-PER-075163

▾ In more languages
Configure

Language	Label	Description
English	Sannom	LSD-PER-075163

Statements

instance of	Person ▾ 0 references	
hasName	Sannom recordedAt	Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7) ▸ 1 reference
hasSex	Male recordedAt	Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7) ▸ 1 reference
hasPersonStatus	Enslaved Person hasStatusGeneratingEvent	Sale of unnamed

Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7) (Q310252)

LSD-EVE-SAL-27571
LSD-EVE-SAL-27571

▾ In more languages
Configure

Language	Label	Description	Also known as
English	Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7)	LSD-EVE-SAL-27571	LSD-EVE-SAL-27571

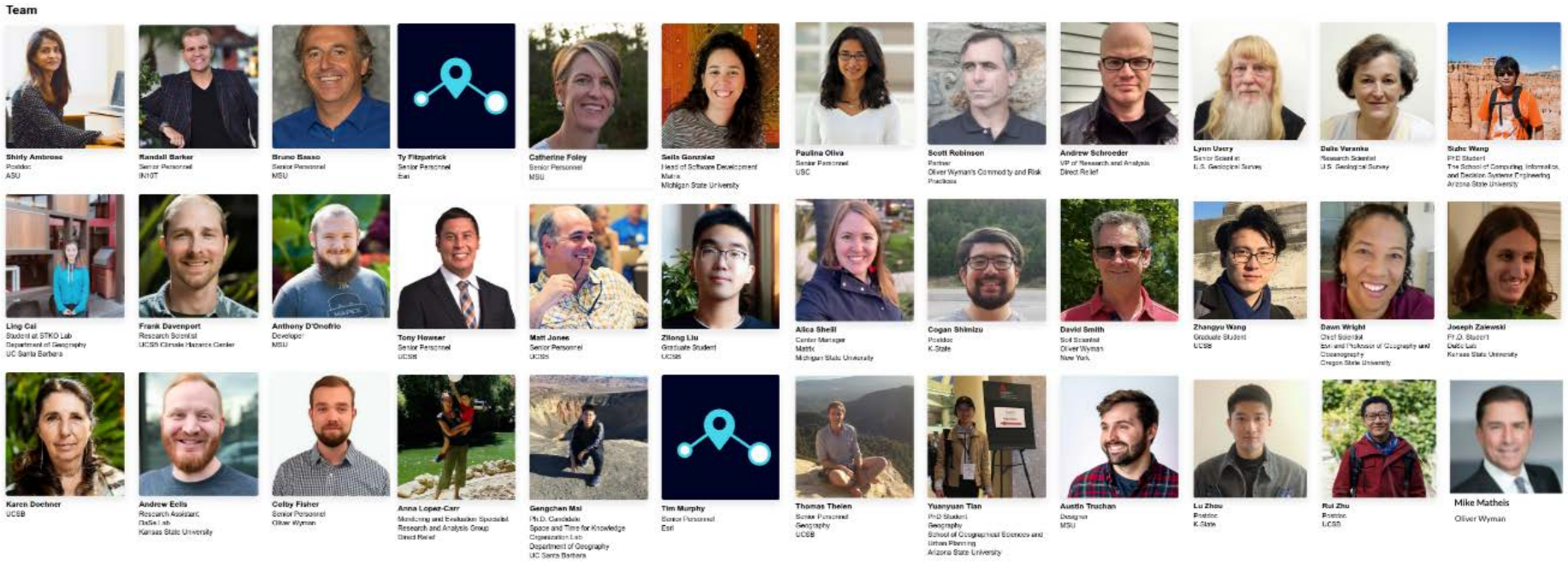
Statements

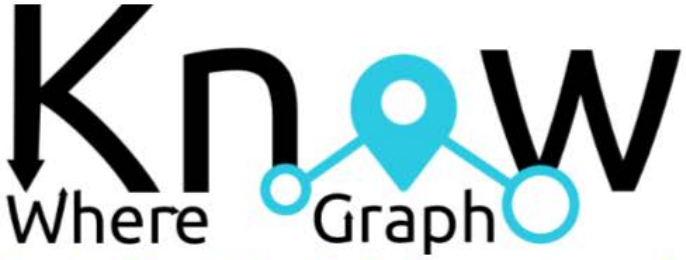
instance of	Event ▾ 0 references
hasName	Sale of unnamed enslaved persons by Mathurin Guerin and sons (1817-6-7) ▸ 1 reference
hasEventType	Sale ▸ 1 reference
date	7 June 1817 *Gregorian* ▸ 1 reference
atPlace	St. James ▸ 1 reference
providesParticipantRole	Sold Person hasParticipantRole: Isaac, James, John, Sarrah, Angelique, Congo Francoise, Alexandre

KnowWhereGraph

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

KnowWhereGraph





(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/

Public release

Forthcoming (very soon!) – public release

- **Knowledge Graph with >12B triples**
  - **One of the currently largest public knowledge graphs.**
  - **Focus on spatial data related to environment and natural disasters**
- **(somewhat later)**
  - **open source software for access and management**

http://knowwheregraph.org/

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 13 (US),
Wildfires	USGS, USDA, USFS, NIFC	wildfire type, burn severity, num. acres burned, contained date	US	1984–current	Global Administrative Regions	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 2022-01-18			
Climate Hazards	NOAA	injuries, deaths, property damages	US	1950–2022			
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, Semantic Scholar	name, affiliation, expertise with spatiotemporal scopes	Global	unlimited	National Weather Zones	NOAA	US
Cropland Types	USDA	crop types (raster data)	US	2008-2021	FIPS Codes	NRCS	US
Air Qual. Obs.	U.S. EPA	AQI value, CO concentration	US	1980–2022	Designated Market Area	Nielen	US
Smoke Plumes	NOAA	daily smoke plumes extent	US	2010-2022	ZIP	ZCTA	US
Climate Observations	NOAA	temperature, precipitation, PDSI, PHSI	US	1950 - 2022	Climate Division	NOAA	US
Disaster Declaration	FEMA	designated area, program, amount approved, program designated date	US	1953 - 2022	Census Metropolitan Area	US Census	US
Smoke Plume Extents	NOAA	Smoke extent	US	2017 - 2022	Drought Zone	NDMC, USDA,NOAA	US
BlueSky Forecasts	Bluesky	PM10, PM5	US	2022-03-07	Geographic Name Information System	USGS	US
Transportation (highway network)	DOT	road type, road length, road sign	US	2014			
Public Health	CDC, US Census	below poverty level percent, diabetes age adjusted 20 plus percent, obesity age adjusted 20 plus percent	US	2017			
Social Vulnerability	CDC/ATSDR	social vulnerability index	US	2018			
Hurricane Tracks	NOAA	max wind speed, min pressure	US	1851-2020			

DaSe Lab

Methods

- We develop and apply a whole range of techniques to problems around knowledge graphs, including
  - Deep learning
  - Natural language processing
  - Logic-based knowledge representation
  - Computational logic and automated reasoning
- We apply our methods to other fields
  - Intelligence data integration and analysis (DARPA)
  - Cognitive Agents (AFOSR)
  - Humanities (Mellon Foundation)
  - Explainable Deep Learning (OBOR)
  - Food Systems data (NIST / Department of Commerce)
  - Scientific data (NSF GEO)
  - Industry (several)

Thanks!

References

- Pascal Hitzler, Semantic Web: A Review of the Field. Communications of the ACM 64 (2), 76-82, 2021.
- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P., and Rudolph, S. (Eds.). *OWL 2 Web Ontology Language: Primer (2nd Ed.)*. W3C Recommendation 11 (Dec. 2012); http://www.w3.org/TR/owl2-primer/.
- Hitzler, P., Krötzsch, M., and Rudolph, S. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2010.
- Vrandecic, D. and Krötzsch, M. Wikidata: A free collaborative knowledgebase. *Commun. ACM 57*, 10 (Oct. 2014), 78–85.
- Michelle Cheatham, Adila Krisnadhi, Reihaneh Amini, Pascal Hitzler, Krzysztof Janowicz, Adam Shepherd, Tom Narock, Matt Jones, Peng Ji, The GeoLink Knowledge Graph. Big Earth Data 2 (2), 2018, 131-143.

References

- Cogan Shimizu, Pascal Hitzler, Quinn Hirt, Dean Rehberger, Seila Gonzalez Estrecha, Catherine Foley, Alicia M. Sheill, Walter Hawthorne, Jeff Mixter, Ethan Watrall, Ryan Carty, Duncan Tarr: The Enslaved ontology: Peoples of the historic slave trade. J. Web Semant. 63: 100567 (2020)
- Cogan Shimizu, Karl Hammar, Pascal Hitzler, Modular Ontology Modeling. Semantic Web. To appear.
- Krzysztof Janowicz, Pascal Hitzler, Wenwen Li, Dean Rehberger, Mark Schildhauer, Rui Zhu, Cogan Shimizu, Colby K. Fisher, Ling Cai, Gengchen Mai, Joseph Zalewski, Lu Zhou, Shirly Stephen, Seila Gonzalez, Bryce Mecum, Anna Lopez Carr, Andrew Schroeder, Dave Smith, Dawn Wright, SizheWang, Yuanyuan Tian, Zilong Liu, Meilin Shi, Anthony D'Onofrio, Zhining Gu, Know, Know Where, KnowWhereGraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications in Environmental Intelligence. AI Magazine.To appear.

References

- Pascal Hitzler, Cogan Shimizu, Modular Ontologies as a Bridge Between Human Conceptualizations and Data. In: Peter Chapman, Dominik Endres, Nathalie Pernelle: Graph-Based Representation and Reasoning - 23rd International Conference on Conceptual Structures, ICCS 2018, Edinburgh, UK, June 20-22, 2018, Proceedings. Lecture Notes in Computer Science 10872, Springer 2018, pp. 3-6.
- Pascal Hitzler, Aldo Gangemi, Krzysztof Janowicz, Adila Krisnadhi, Valentina Presutti (eds.), Ontology Engineering with Ontology Design Patterns: Foundations and Applications. Studies on the Semantic Web Vol. 25, IOS Press/AKA Verlag, 2016.
- Adila Krisnadhi, Pascal Hitzler, Modeling With Ontology Design Patterns: Chess Games As a Worked Example. In: Pascal Hitzler, Aldo Gangemi, Krzysztof Janowicz, Adila Krisnadhi, Valentina Presutti (eds.), Ontology Engineering with Ontology Design Patterns: Foundations and Applications. Studies on the Semantic Web Vol. 25, IOS Press/AKA Verlag, pp. 3-22.
- Cogan Shimizu, Karl Hammar, CoModIDE – The Comprehensive Modular Ontology IDE. In: 18th International Semantic Web Conference: Satellite Events, 2019, to appear.

Thanks!