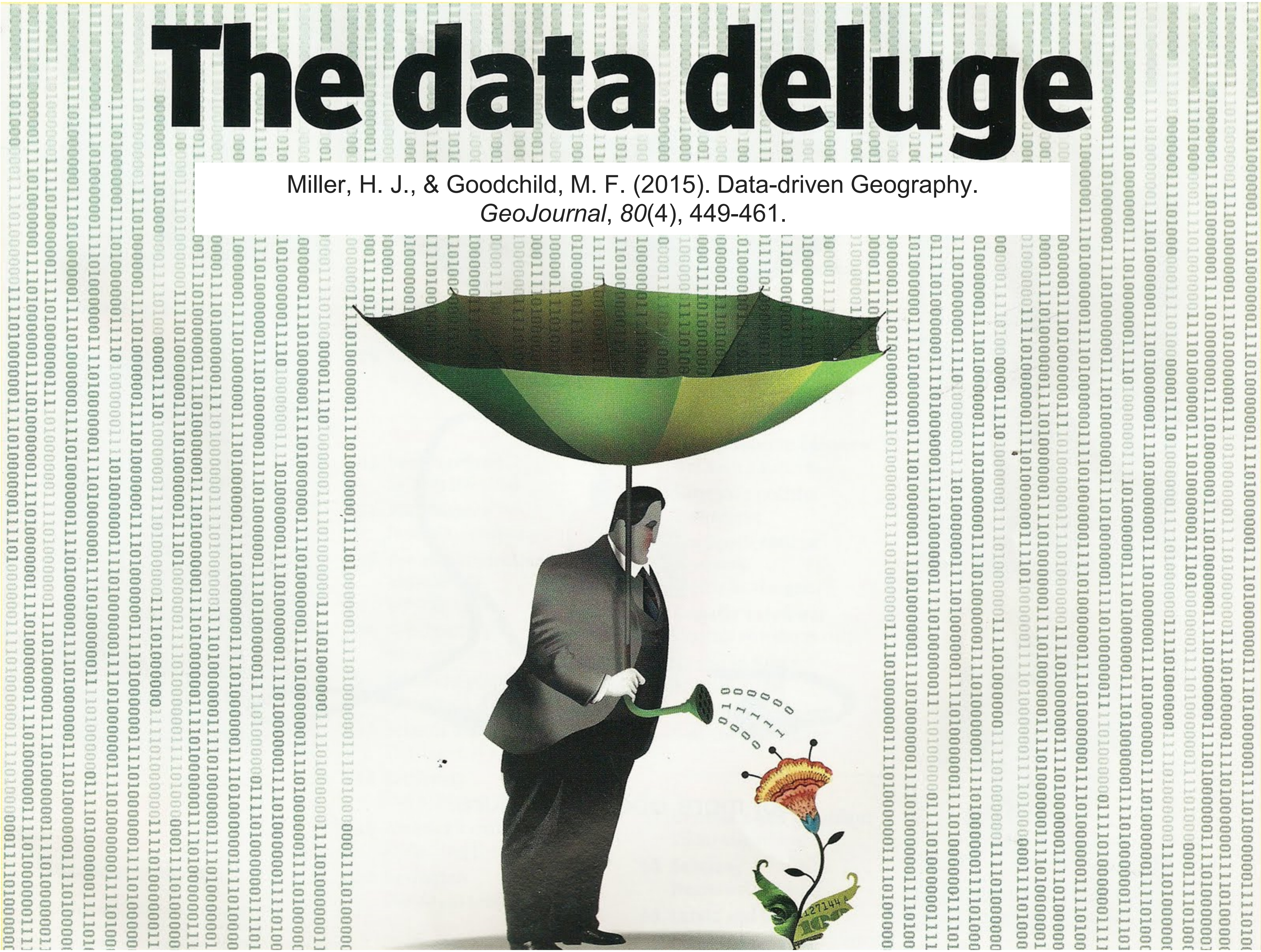
GeoAI, GeoKnowledge Graphs and GeoSemantics – Formalizing Geographic Reality

Johannes Scholz

TU Graz, Institute of Geodesy
Research Group Geoinformation

johannes.scholz@tugraz.at
ifg.tugraz.at || www.johannesscholz.net
@Joe_GISc @Joe_GISc@mastodon.online



Miller, H. J., & Goodchild, M. F. (2015). Data-driven Geography.
GeoJournal, *80*(4), 449-461.

Data Deluge and Artificial Intelligence?

Artificial Intelligence (AI) is:

*“a system's ability to correctly **interpret external data**, to **learn from such data**, and to **use those learnings** to **achieve specific goals** and tasks through **flexible adaptation**.”*

(Kaplan & Haenlein 2019)

Questions that surface …

What’s to come…

- **Methodological background**
  - **Geospatial AI :: a definition**
  - **Semantic Web & Knowledge Graphs**
  - **NoSQL Databases**
- **Integration of GeoAI, Knowledge Graphs & NoSQL?**
- **Selected Applications**
- **Conclusion**

Methodological Background
||
GeoAI

Geospatial AI

*"**Geospatial Artificial Intelligence (GeoAI)** as a subfield of **spatial data science** utilizes advancements in techniques and data cultures to support the creation of more intelligent geographic information as well as methods, systems, and services for a variety of downstream tasks.*
*These include **image classification**, **object detection**, **scene segmentation**, **simulation** and **interpolation**, **link prediction**, (natural language based) retrieval and **question answering**, **on-the-fly data integration**, **geo-enrichment**, and many others."*

(Janowicz et al. 2019)

Geospatial AI

- "**GeoAI** can be regarded as a study subject to **develop intelligent computer programs** to mimic the processes of **human perception**, **spatial reasoning**, and **discovery** about **geographical phenomena** and **dynamics**
  - to advance our knowledge,
  - to solve problems in human environmental systems and their interactions,
  - with a focus on spatial contexts and roots in geography or GIScience." (Gao, 2021)

- **Spatially explicit models incorporating spatial contexts** (Yan et al., 2018) can **outperform traditional nonspatial AI** models in many tasks:
  - image classification,
  - geographic knowledge graph summarization (Yan et al., 2019),
  - and geographic question-answering problems (Mai et al., 2019).

Geospatial AI :: history

- AI was born in 1956 at a workshop at Dartmouth College (McCarthy 1956)

- Development of AI
  - Early optimism (1960s and 70s)
  - AI winter followed thereafter – problem: lack of addressing real-world problems
  - After 2010: significant progress in AI research

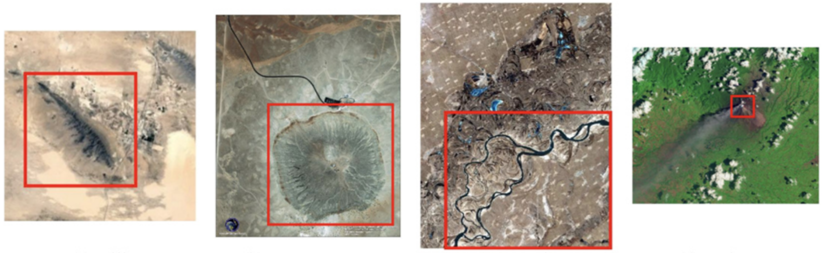
- Why progress after 2010:
  - Big data (user generated data, sensor data, high-quality labeled data)
  - Novel algorithms
  - Immense computational power

Geospatial AI :: history

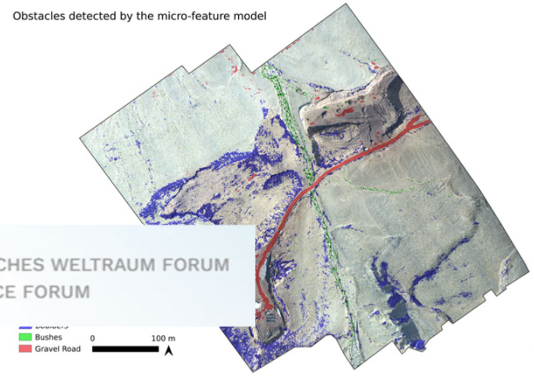
- Usage of AI technologies in Geography is not new
  - Openshaw & Openshaw (1997): Artificial Intelligence in Geography
  - Couclelis (1986) and Smith (1984) discussed the potential role of AI for geographic problem-solving

- AI technologies and geospatial "boom" relies on a change of culture (Janowicz et al. 2019)
  - Open-content mostly via APIs (100 APIs in 2005 vs. 22k in 2019)
  - Reusing data is the new normal
  - Data synthesis, alongside analysis >> one datasource can be used as proxy for the other one (which is maybe difficult to acquire)
  - From 2014 onwards – VGI was used to detect new insights (not only to confirm existing theories!) (e.g. Adams et al. 2014, Janowicz et al. 2014)

Detection of terrain features (Li and Hsu 2020, Prinz et al., 2022)

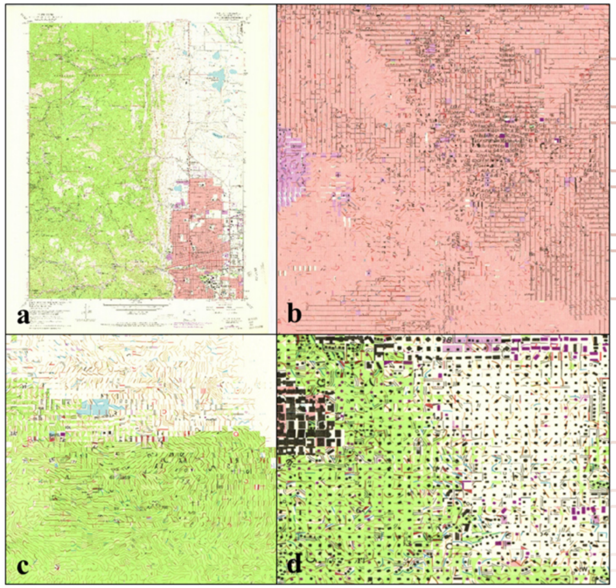


(a) Hill (b) Impact crater (c) Meander (d) Volcano



ÖSTERREICHISCHES WELTRAUM FORUM
AUSTRIAN SPACE FORUM

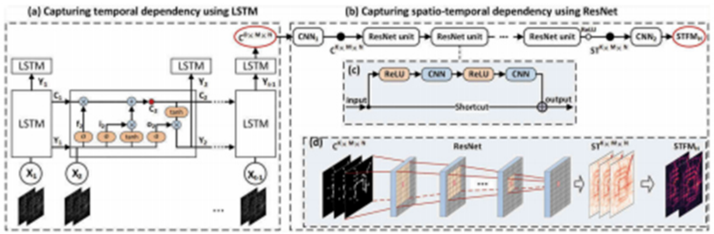
Information extraction from historical maps (Duan et al. 2020)



Building footprints (Xie et al. 2020)

- INSPIR

Traffic forecasting (Ren at al. 2020)



GeoAI :: Requirements

- High-quality data (i.e. high quality labels)

- Metadata are structurally incomplete and not detailed enough
  - Designed at a specific point in time > future use could not be foreseen
  - Data provenance and contextual information is necessary – and automatic workflows to create them!

- Data synthesis as new paradigm (Hey et al. 2009; Janowicz et al. 2015):
  - Semantics
  - Real-time data integration (semantic query language)

Methodological Background
||
Linked Data &
Knowledge Graphs

Linked Data & Knowledge Graphs

Linked Data describes a methodology of publishing **structured data** so that data from different sources can be **interlinked with typed links.**

- published in a **machine-readable form**
- published in a way that their **meaning is explicitly defined**
- linked to other data sets
- data that **can be linked from other data sets**

Paving the way from a *document oriented* Web to a *data driven* Web

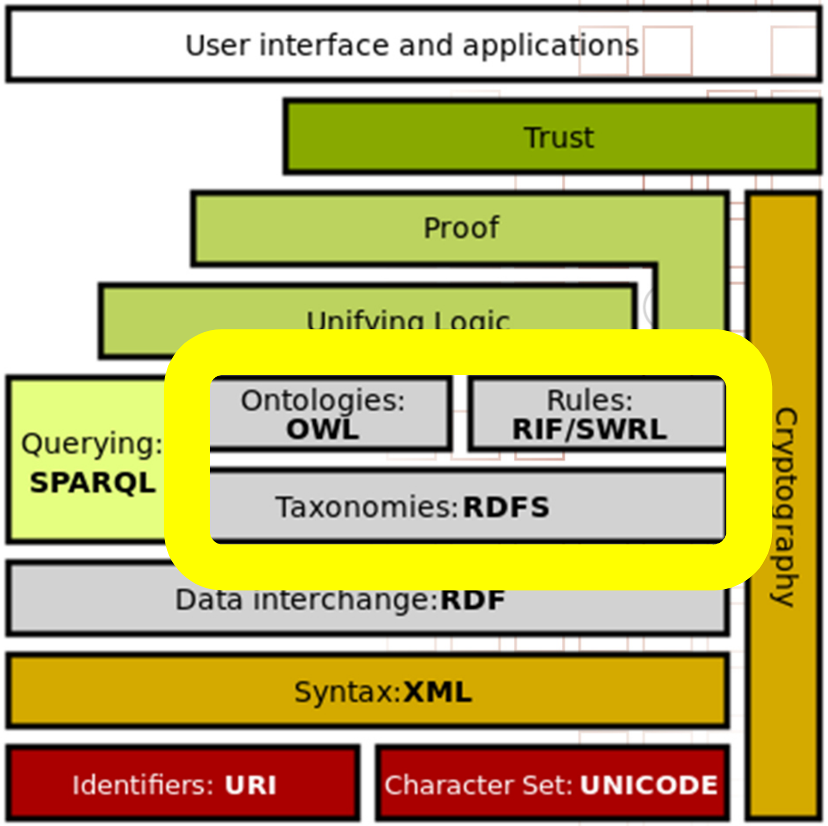
>> Web of Data <<

Geospatial Semantic Web

- Information seeking by allowing exploration, editing and interlinking of heterogeneous information sources with a spatial dimension (Janowicz et al. 2013; Egenhofer 2002).

- Combining Linked Data and Geoinformation can lead to a geospatially enriched Semantic Web
  - Geographic information can easily be integrated and processed.
  - But: requires semantics (Ontologies, Taxonomies)

- A number of Linked Data repositories with spatial data already available!

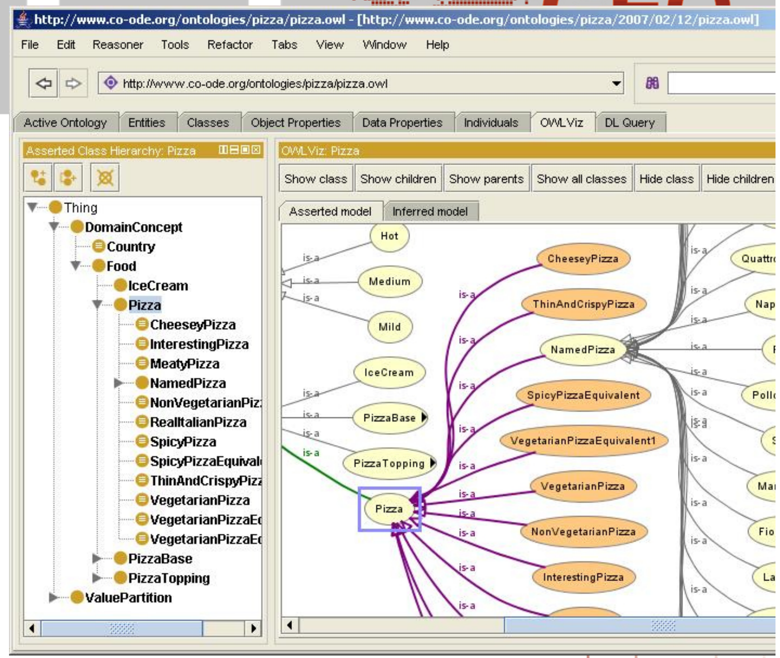


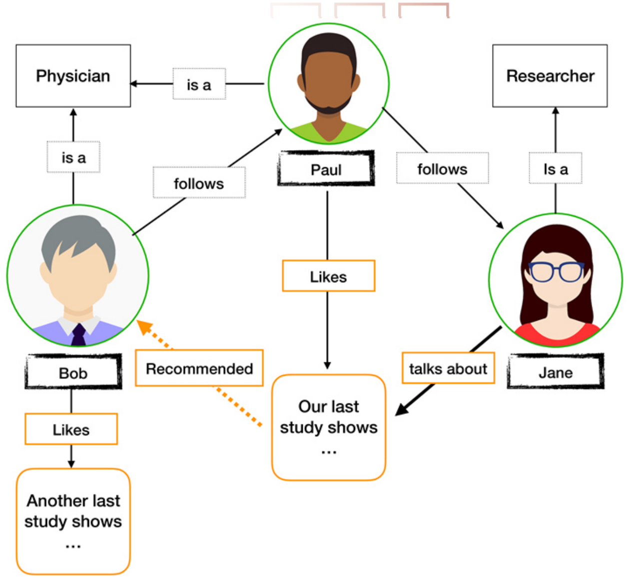
Knowledge Graphs & Ontologies

- Ontology:
  - Formal, explicit specification of a shared conceptualization (Gruber, 1993)
  - Description of the concepts and their relations existing in a Universe of Discourse (Uschold & Gruninger, 1996)

- Knowledge Graphs
  "A knowledge graph
  - (i) mainly describes real world entities and their interrelations, organized in a graph,
  - (ii) defines possible classes and relations of entities in a schema,
  - (iii) allows for potentially interrelating arbitrary entities with each other and (iv) covers various topical domains."

  (Paulheim 2017)

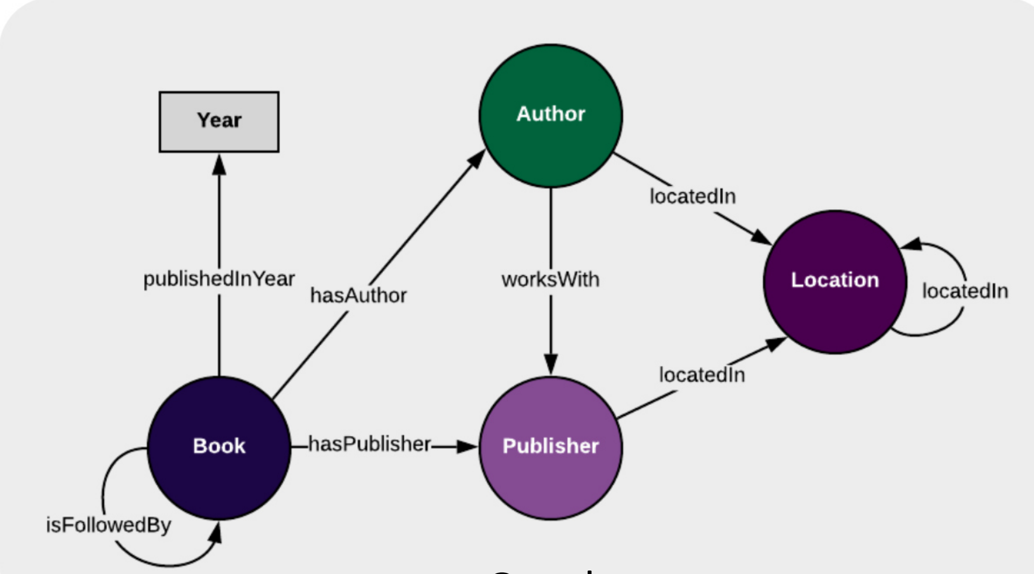




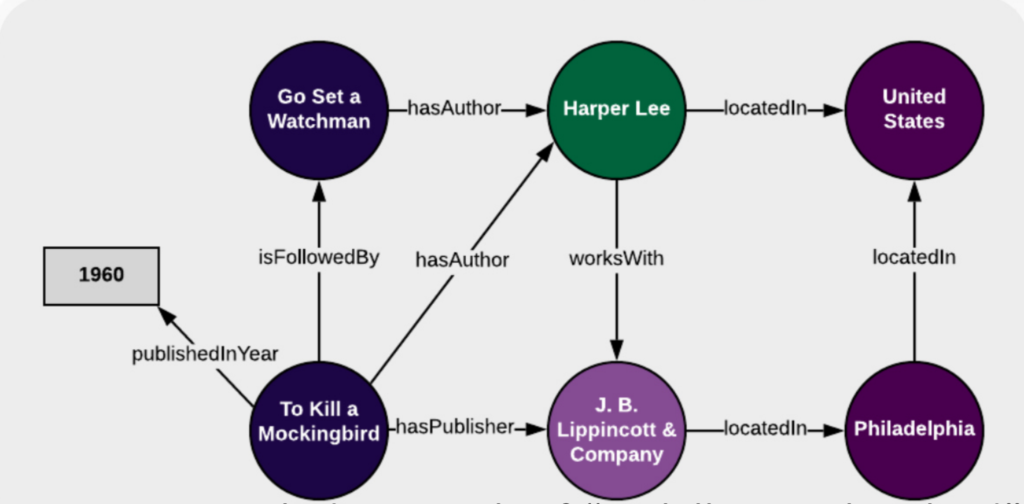
Knowledge Graphs & Ontologies

- Ontologies are used for
  - Definitions of shared vocabularies (>> Interoperability)
  - Actionable knowledge fragments (>> inferencing [i.e. creating new knowledge])

- Knowledge Graphs:
  - All "features" of ontologies
  - Create specific instances of each of the relationships



Ontology



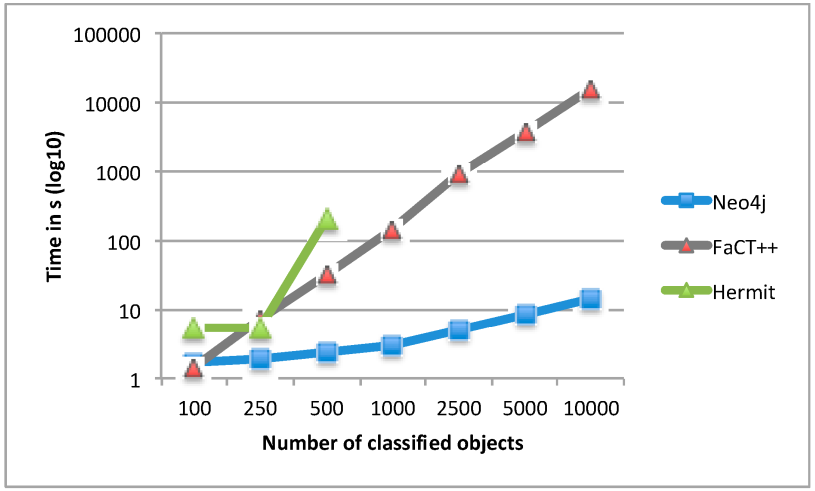
Knowledge Graph of "To kill a mockingbird"

Knowledge Graph Advantages

- Basic “equation”:

Ontology + Data = Knowledge Graph

- Graphs are an **efficient data structure** in terms of storage and analysis
- Graphs are supported by **Semantic Web approaches** and contemporary **NoSQL databases**
- In comparison to OWL-Ontologies and Reasoners the **reasoning speed is significantly higher** (see Lampoltshammer & Wiegand 2015)



Classification speed of EO data (Lampoltshammer & Wiegand 2015)

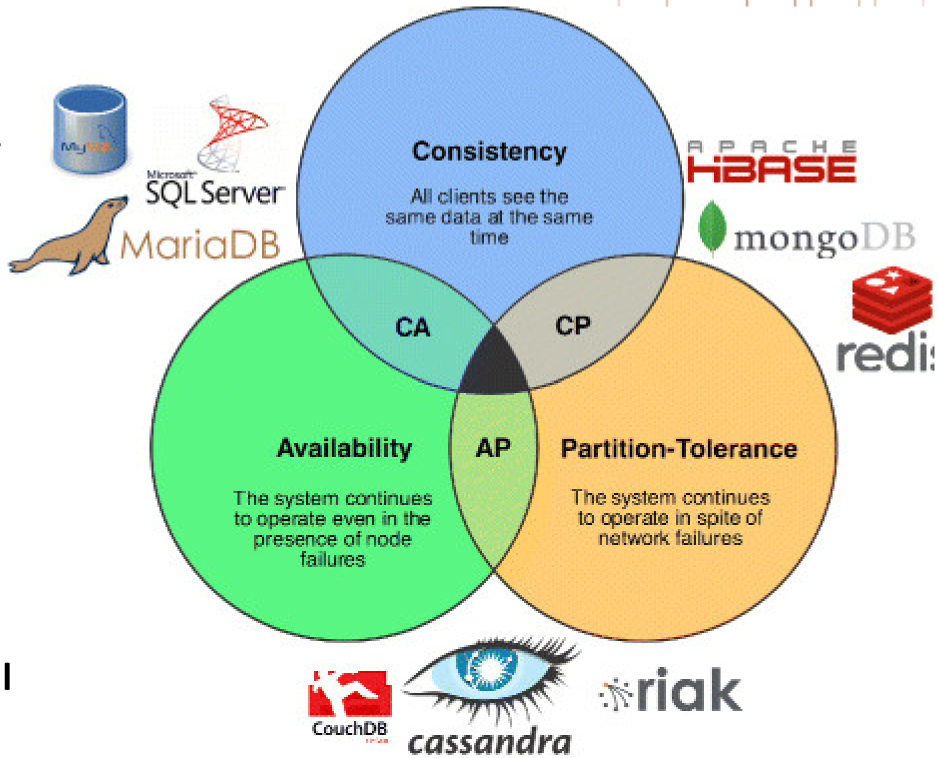
Methodological Background

||

NoSQL Databases

NoSQL Paradigm

- Not-only SQL (NoSQL) term emerged in 2009
- Umbrella term for a number of different database concepts (Friedland et al., 2011) with the following characteristics:
  - **Non-relational data model**
  - **Absence of ACID** (especially consistency – replaced with "*eventually consistent*")
    - Replaced with CAP theorem (Brewer 2000)
    - resulting in BASE (consistency & isolation are forfeited) (Pritchett 2008):
      - Basically available, Soft state, Eventual consistent (Vogel 2009)
  - **Flexible schema**: structure of data is not defined through explicit schemas; applications can store data as they desire;
  - **Tailored towards distributed** an **horizontal scalability**, **high data turnover** rates (Big Data)



Lourenço et al. (2015)

NoSQL types :: Overview

- **Column databases**
  - Tables, rows and columns – but columns can change
  - Apache Cassandra, Apache Hbase, Apache Accumulo, Google Bigtable
- **Key-value databases**
  - Key and associated value (similar to a hash), no relations
  - OrientDB, Dynamo (Amazon), Berkeley DB
- **Document databases**
  - Document metaphor – JSON, XML encodings to represent documents (absence of a schema!)
  - Apache CouchDB, MongoDB, CosmosDB (Microsoft), IBM Domi
- **Graph databases**
  - Representing data as **graphs in a database** (Robinson, Weber & Eifrem, 2015)
  - Graph DBs popular: Facebook Open Graph, Google Knowledge Graph, Twitter FlockDB (Miller, 2013)
- Multi-model databases

Integration?

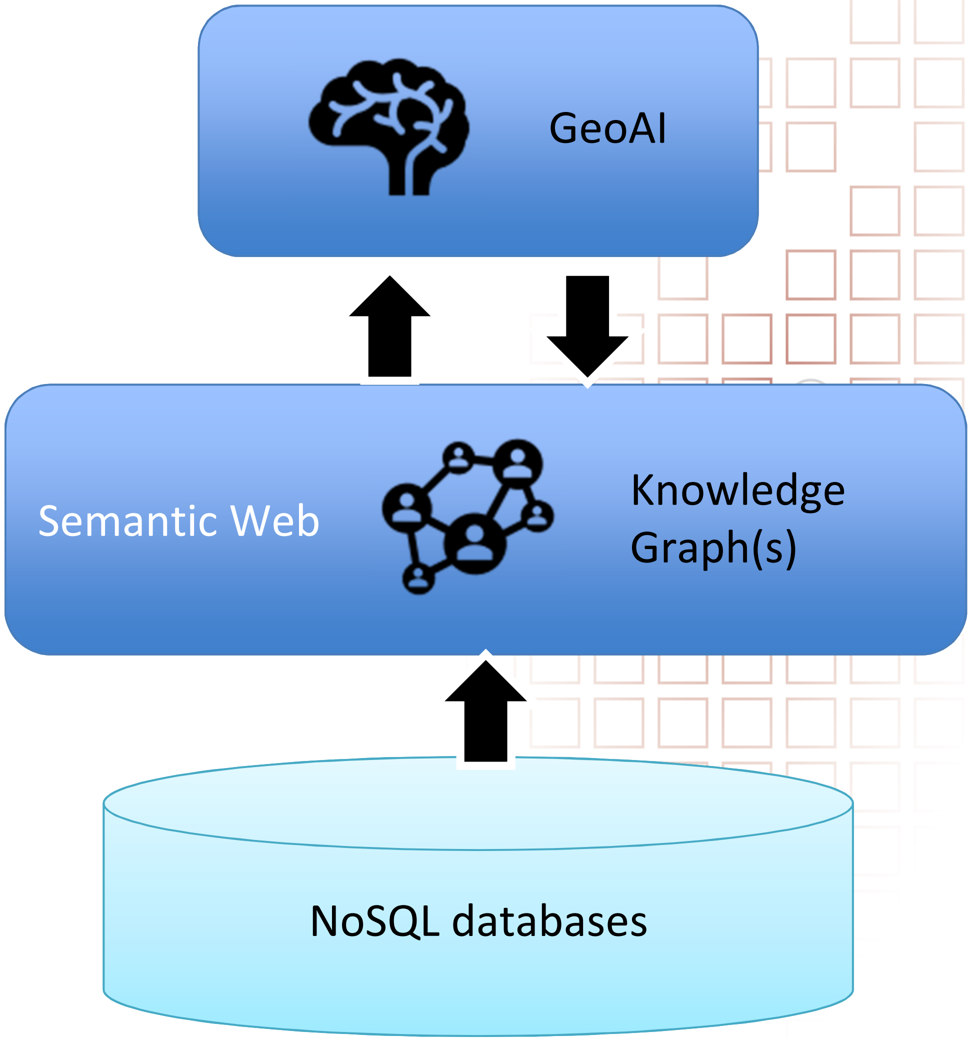
GeoAI || Knowledge Graphs || NoSQL

Connections?

- GeoAI can be fueled by (Geo)Knowledge Graphs

Why?

- Reusability of (geo)semantic queries (GeoSPARQL)
- Offers inference & reasoning
- Integration of heterogeneous data
- Geospatial knowledge graphs are symbolic representations of geospatial knowledge



Connection / Integration

- Knowledge graphs are **understood** by both **humans** and **machines**
  - Serve foundation for artificial intelligence (Semantic AI)
  - Facilitate applications such as **geospatial data integration** and **knowledge discovery**

- Spatial Linked Open Data cloud
  - Open-source cross-domain knowledge graph
  - Essential for describing events, people, and objects

- Geographic Question Answering (e.g. Mai et al. 2020):
  - Semantically enriched contextual data necessary
  - Data synthesis(!)

  >> (Geo)Knowledge Graphs can serve that functionality

Application Examples

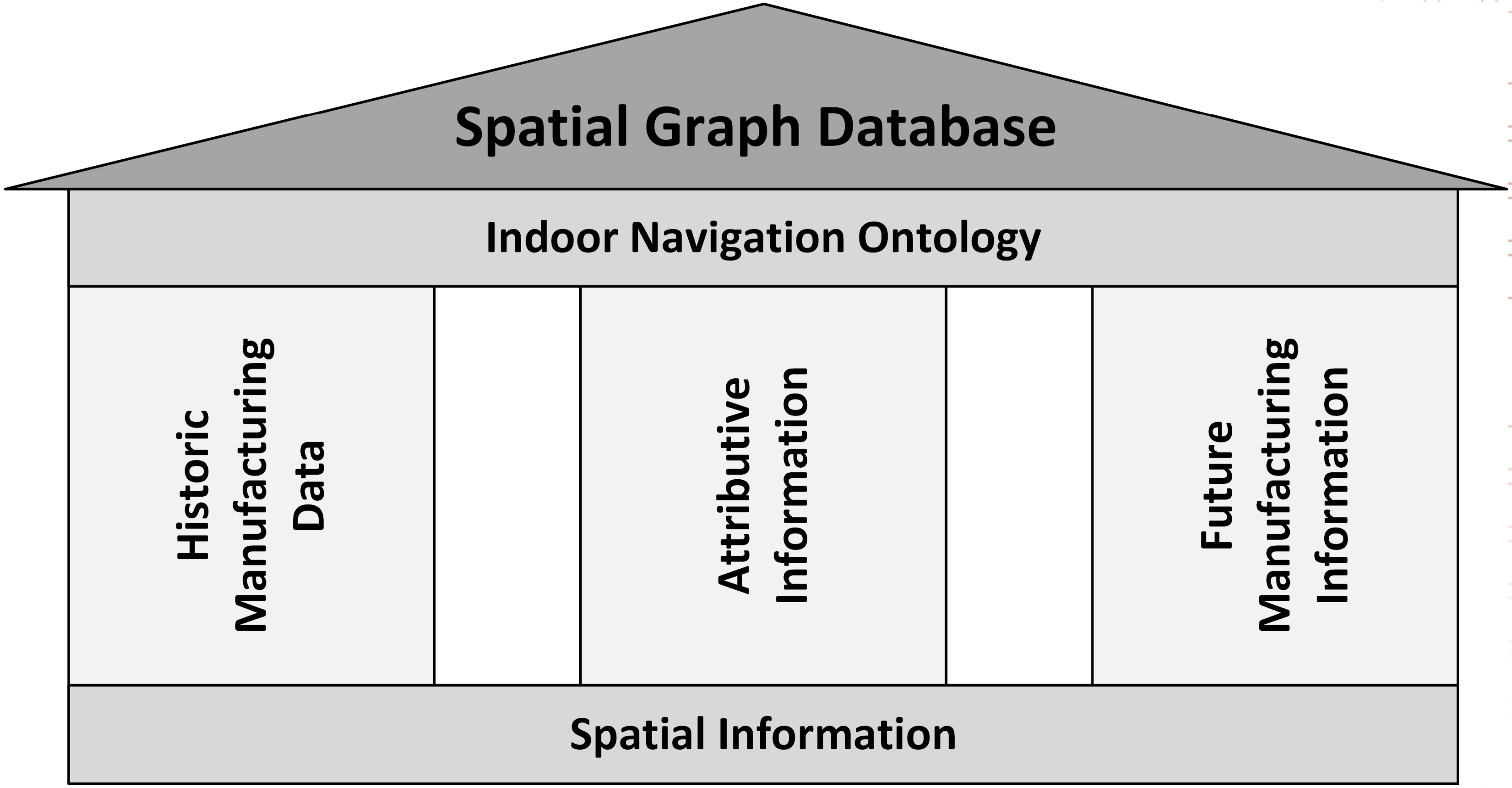
Indoor Geography and Smart Manufacturing

GEO INFORMATION

- Support for Decision-making in a semiconductor facility (Scholz & Schabus 2017; Schabus & Scholz 2017a; Schabus & Scholz 2017b)
  - Manufacturing purposes
  - Incident management

- Ontology for manufacturing data
  - Based on an indoor space ontology (Scholz & Schabus, 2014)
  - Spatial information
    - stored in classes position and graph
  - Temporal component
    - Historical information on production assets (spatial information [trajectory], sequence of manufacturing operations)

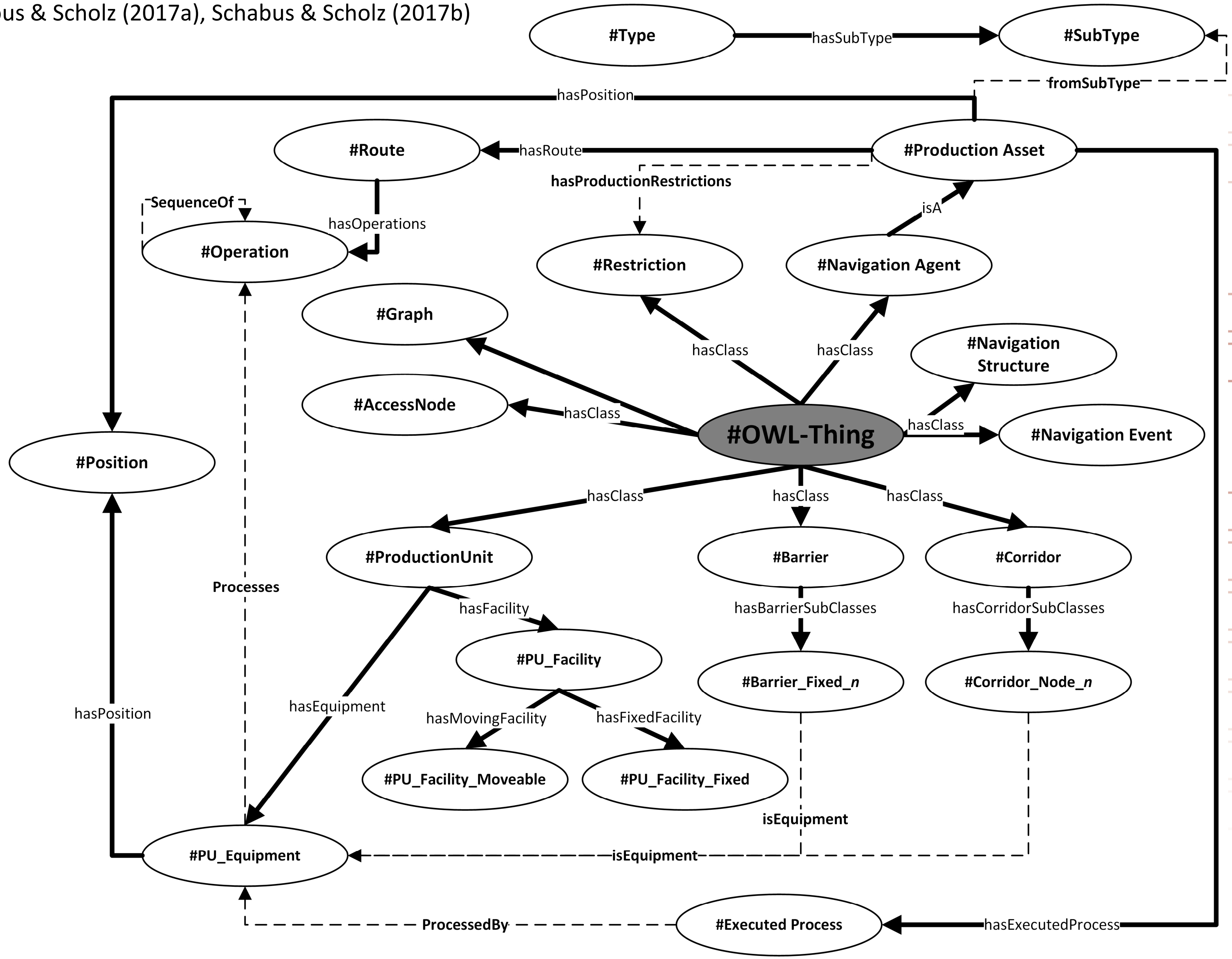
Indoor Geography and Smart Manufacturing



Schabus & Scholz (2017a), Schabus & Scholz (2017b)

Indoor Geography and Smart Manufacturing

Schabus & Scholz (2017a), Schabus & Scholz (2017b)



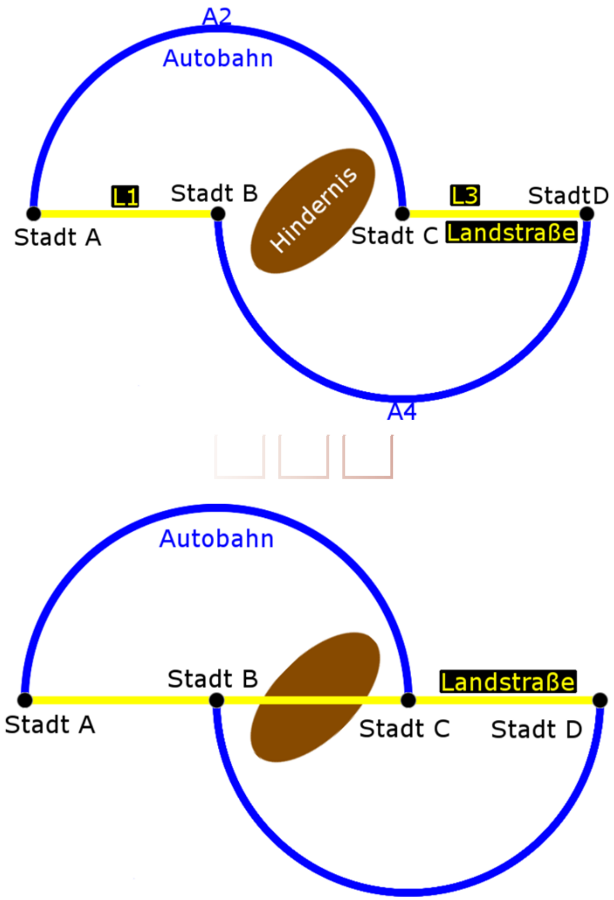
Selfish Routing & Agent-based Simulation

- **Selfish routing** is a result of different agents acting in a network, trying to find the **best route** from a **strictly personal viewpoint**, regardless of the consequence for other agents.

- Based on the **Braess Paradox** (Braess 1969, Roughgarden 2005)

- Result:
  >> selfish behaviour results in higher latency

- Objective:
  - **Selfish behaviour and uncertainty & influence of cognitive agents** (Scholz & Church 2018, Scholz 2015)



- **Predictive Memory** is a concept based on the recognition-prediction framework (Clark 2013; Hawkins & Blakeslee 2007):
  - matching sensory inputs with stored memory patterns
  - leads to predictions of what will happen in the future
  - involves constant learning from previous experiences

Selfish Routing & Agent-based Simulation

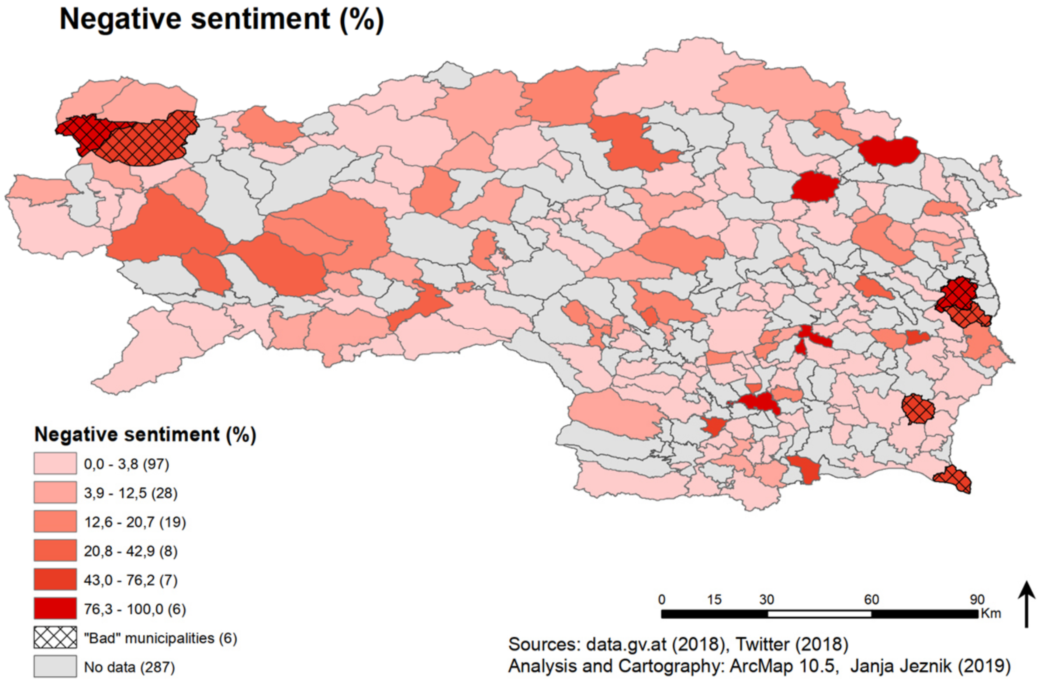
- Simulate such environments with cognitive agents in a spatial Agent-based model (ABM)

- Each agent is equipped with a predictive memory
  (Scholz 2015; Exenberger & Scholz forthcoming)
  - Graph-based memory structure (individual experiences and outcomes)
  - Reinforcement learning (i.e. Machine Learning) to match current traffic situations with historic experiences
  - Decision making based on historic experiences (and outcomes) and the current goal

Knowledge Discovery from geo-text data

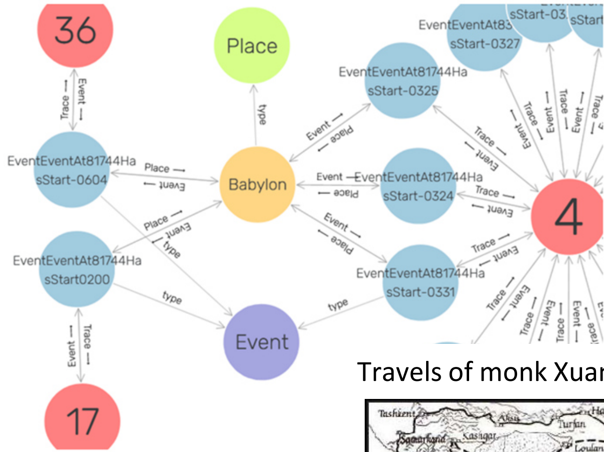
- Place opinions/emotions
  - Geo-text data contains words expressed by human beings
  - So there are some opinions and emotions involved as well ?
  - Analysing this is done with **Sentiment analysis** (Pang et al. 2008, Liu 2012)

- Analysis of crowd-sourced tourist data for the province of Styria
  (Scholz & Jeznik 2020)
  - MongoDB as basis for Sentiment analysis
  - Spatio-temporal analysis



GeoKGs for Digital Humanities and Supply Chain Visibility

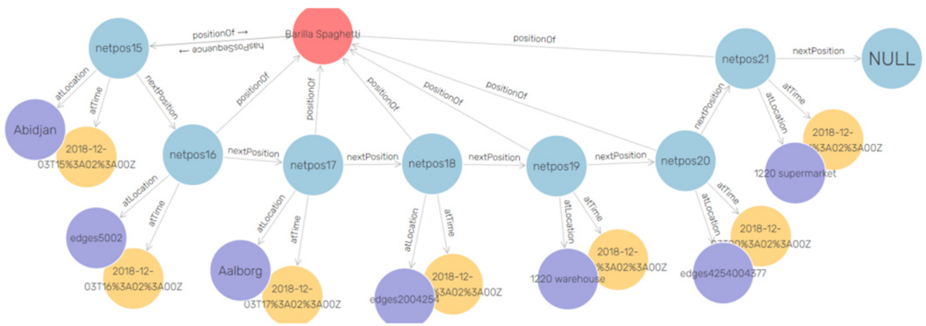
- GeoKGs for Digital Humanities (Hübl & Scholz, 2021) (CaGIS Rising Funding)
  - Linked Traces model the events of geographic movement (Grossner, 2021)
  - GeoKG of Linked Traces



Travels of monk Xuanzang (629AD-645AD)

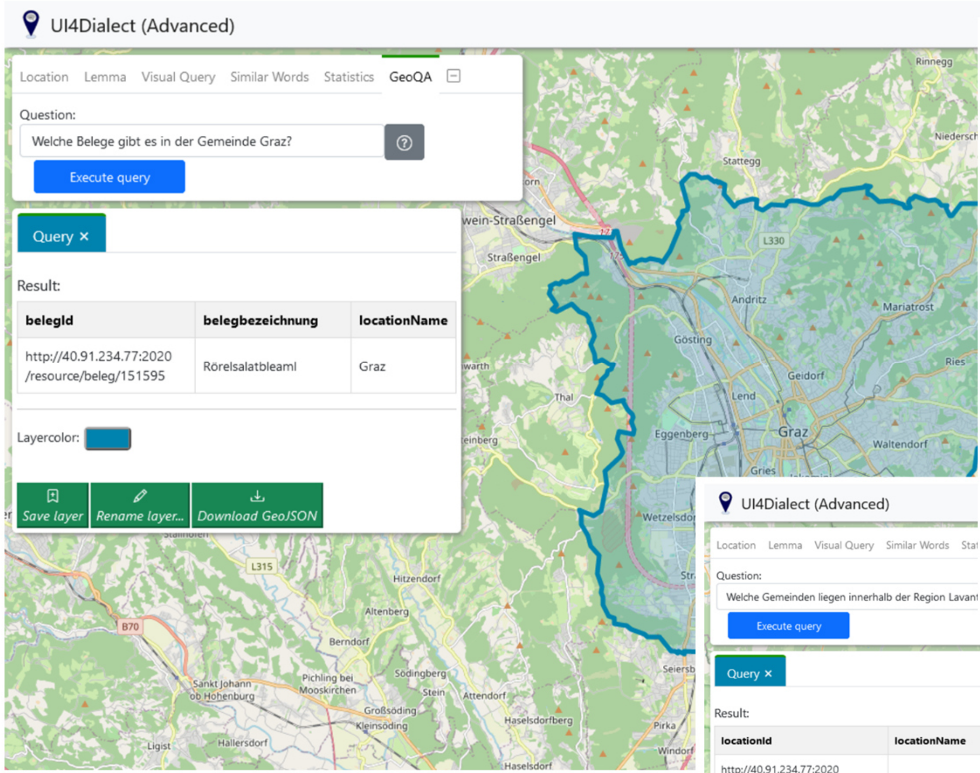
World Historical Gazetteer
linking knowledge about the past via place

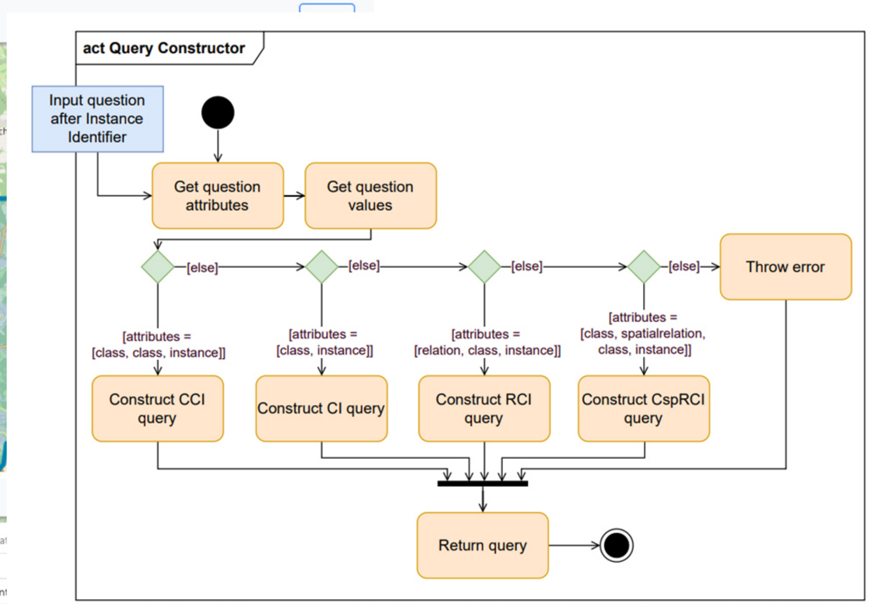
- GeoKGs for Supply Chain Visibility (Dopler & Scholz, 2021)

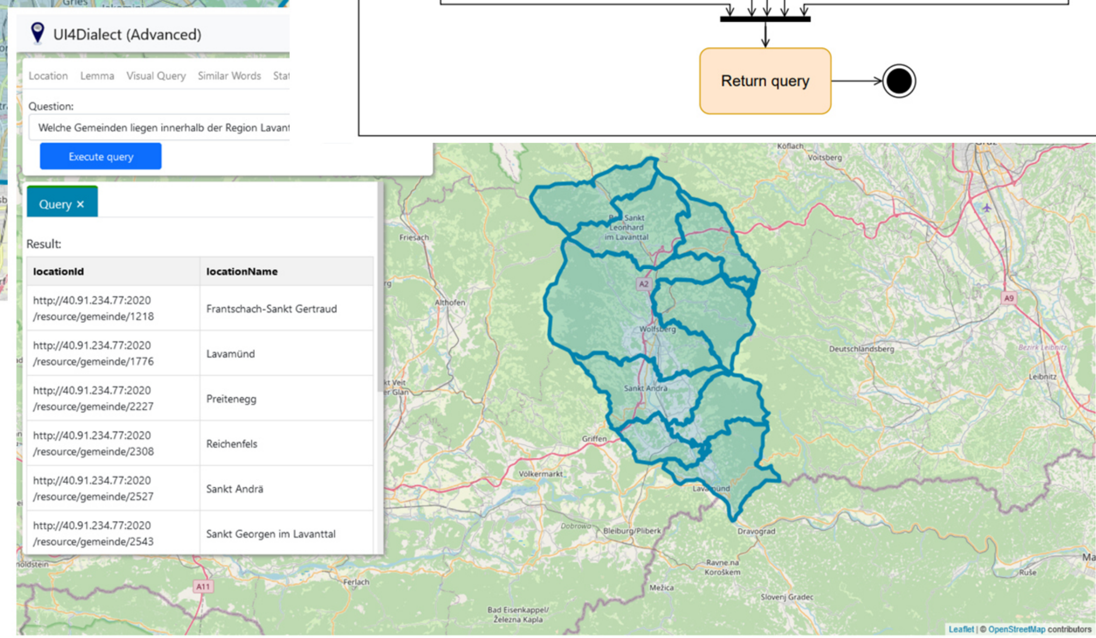


Geographic Question Answering

- **We utilize GeoKGs for GeoAI:**
  - GeoKGs help to denote the semantics of the digital abstraction of the reality
  - (Geo)Semantics helps to infer new knowledge!
    - Level of Detail (space and time)
    - Switching between spatio-temporal granularity levels.
  - Spatial Linked Data and Geospatial Semantic Web are crucial to **Geographic Question Answering** (e.g. Mai et al., 2020; Weinberger et al., 2022)







Weinberger et al., 2022

Causality & GeoAI

- ML is increasingly used for natural hazard risk assessment
  - Occurrence prediction
  - Impact prediction

- Predictive performance (e.g. accuracy) of ML models in this domain not sufficient

- Models should meet additional requirements
  - **Trustworthiness**: Users trust the model
  - **Informativeness**: Users can interpret model results properly
  - **Robustness**: Consistent and accurate generalization over time and space

Causality & GeoAI

Physical Model

- Fully specified data generating process
- Heavily relies on expert knowledge

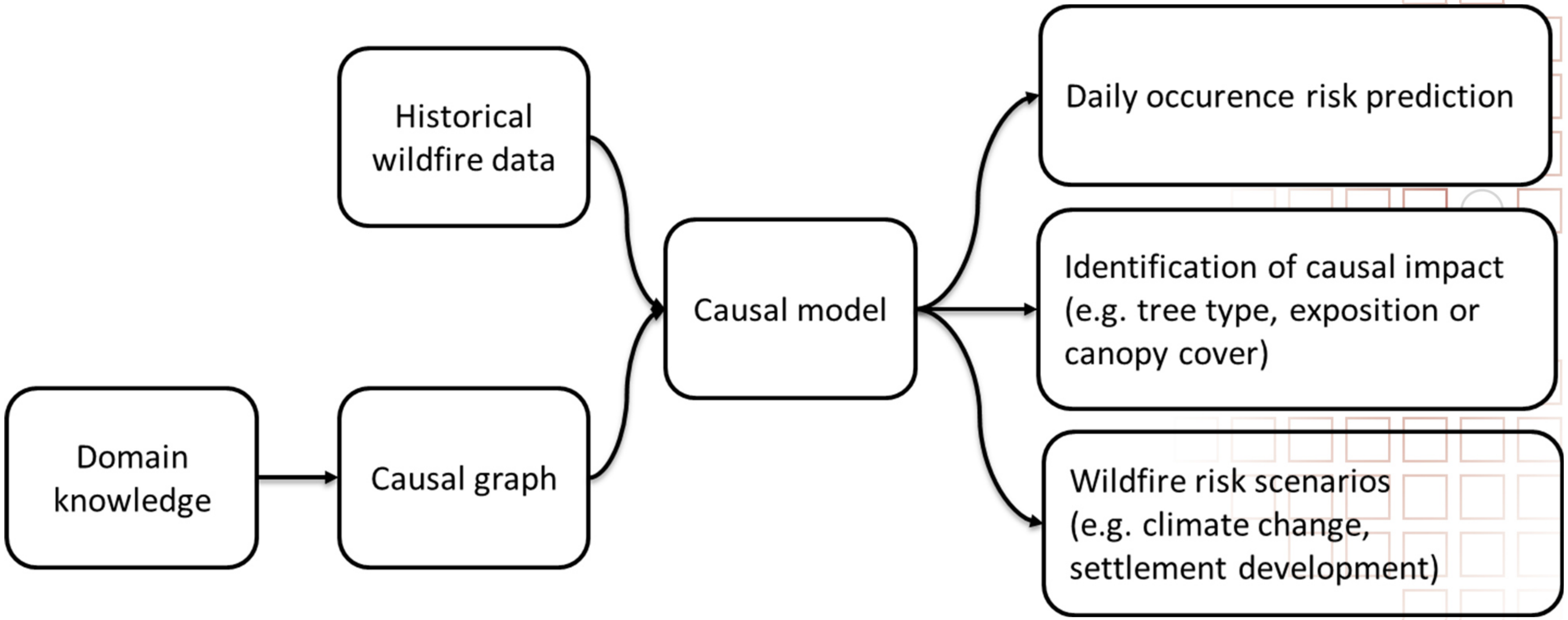
Causal Model

- Represents causal relationships of a system
- Relies on weak expert knowledge & observational data
- Prediction under distribution shift
- What-If Questions

Statistical Model

- Data-driven
- Relies on statistical dependcies
- Predictions in i.i.d setting

- Goal: Improved wildfire occurence risk assessment for Austria



Conclusion

Some lessons learned so far…

- GeoAI, GeoKGs and Geosemantics are closely related and of utmost interest for the GIScience community!

- Geoinformatics, GeoAI, GeoKGs and Geosemantics (can) deliver the methodological advance to
  - Make **significant contributions** to innovative **solutions of applied research questions** from **various scientific fields**
  - Serve as **“glue” between different scientific fields** in **interdisciplinary contexts**

- **Help us understand spatial phenomena in the data-driven era!**

Thanks to the team & collaborators

GEO INFORMATION

ISO

Open Geospatial Consortium

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

BOKU
Universität für Bodenkultur Wien
University of Natural Resources and Life Sciences, Vienna

University for Continuing Education Krems
DONAU UNIVERSITÄT KREMS

UC SANTA BARBARA Geography

TU WIEN
TECHNISCHE UNIVERSITÄT WIEN

FFG
Forschung wirkt.

dwh
technical solutions
simulation services

University of Ljubljana

European Environment Agency

EPFL

ÖAW
AUSTRIAN ACADEMY OF SCIENCES
AUSTRIAN CENTRE FOR DIGITAL HUMANITIES AND CULTURAL HERITAGE

PARIS
LODRON
UNIVERSITÄT
SALZBURG

… starting in Feb, 2024 you'll find me & group here:

Johannes Scholz

Paris-Lodron-Universität Salzburg
FB Geoinformatik (Z_GIS)

johannes.scholz@plus.ac.at
www.zgis.at || www.johannesscholz.net
@Joe_GISc @Joe_GISc@mastodon.online

Z_GIS

DEPARTMENT OF GEOINFORMATICS - Z_GIS

Next Generation Methodologies for Spatial Analysis and Spatial Simulation - Ontologies and Numerical Methods

Johannes Scholz

TU Graz, Institute of Geodesy
Research Group Geoinformation

johannes.scholz@tugraz.at
ifg.tugraz.at || www.johannesscholz.net
@Joe_GISc @Joe_GISc@mastodon.online