New Research Frontiers:
GeoAI, (Geo-)Knowledge Graphs and NoSQL Databases

GI Research Colloquium 2020 @ CUAS

Johannes Scholz

Institute of Geodesy – TU Graz
Head of RG Geoinformation

Mail: johannes.scholz@tugraz.at
Web: ifg.tugraz.at || www.johannesscholz.net
Twitter: @Joe_GISc

The data deluge

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

Datafication

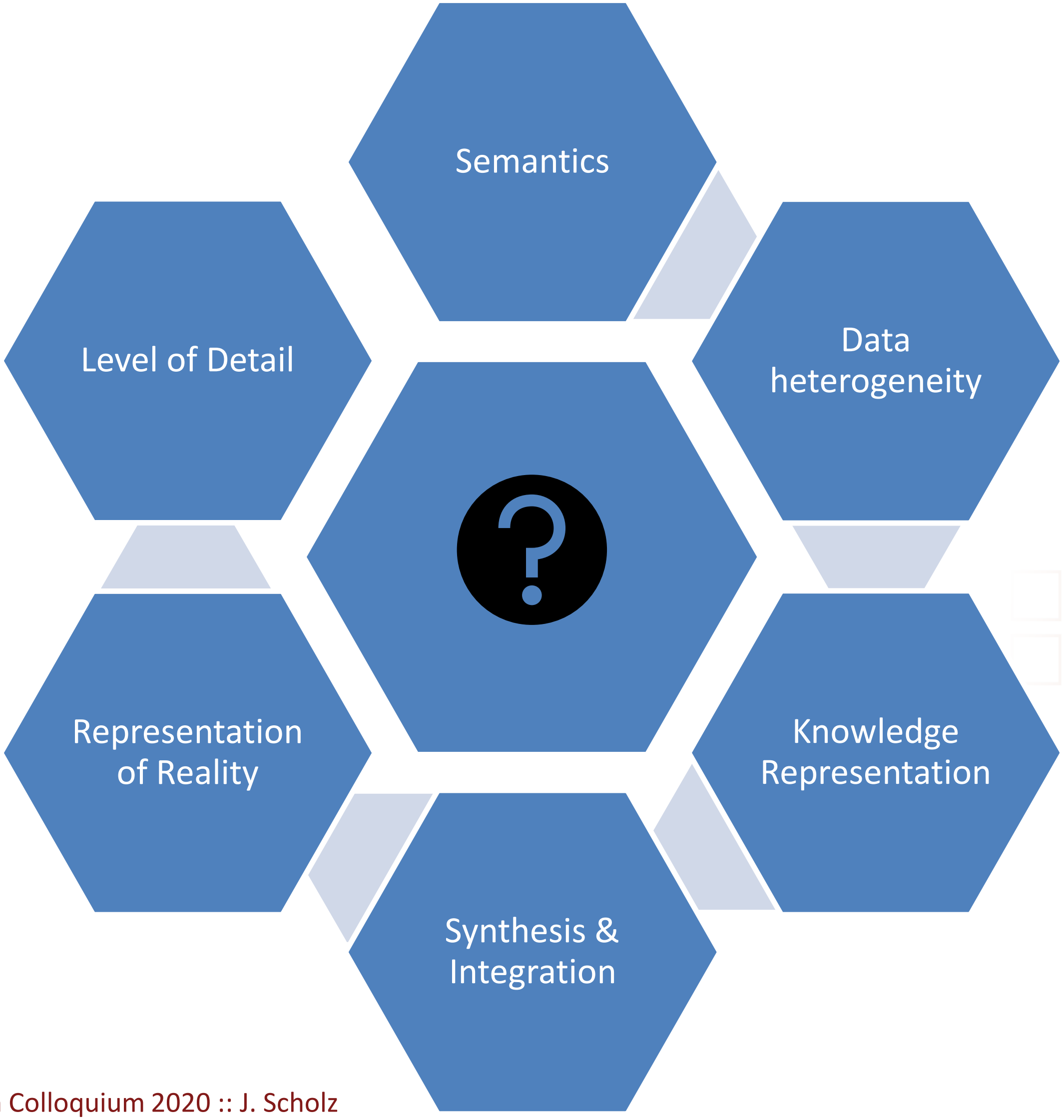
Data Deluge and Artificial Intelligence?

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
- Research Frontiers

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 :: 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

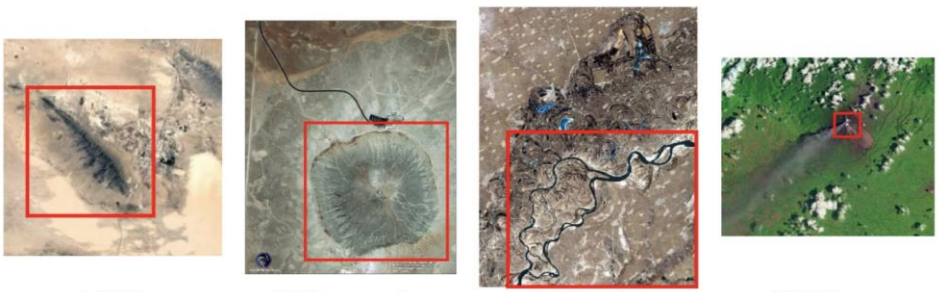
Geospatial AI :: history

- 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)

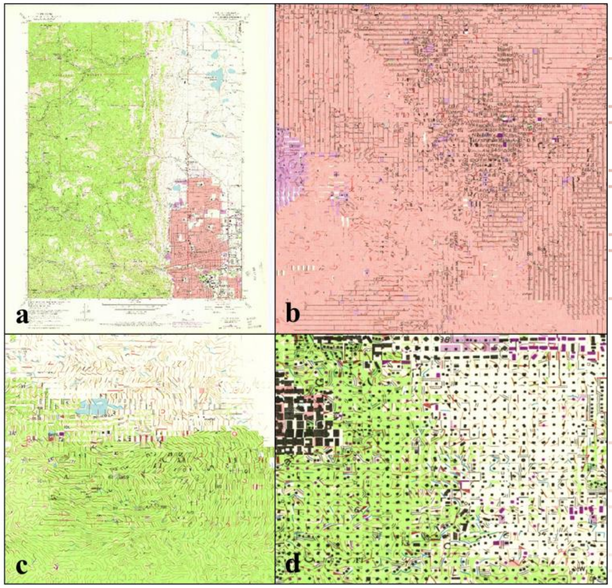
Geospatial AI :: Success Stories

Detection of terrain features (Li and Hsu 2020)



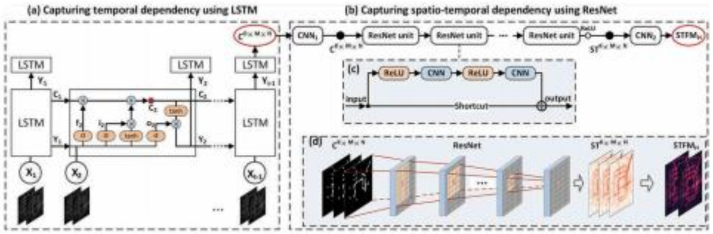
(a) Hill (b) Impact crater (c) Meander (d) Volcano

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



Building footprints (Xie et al. 2020)

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 fourth 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

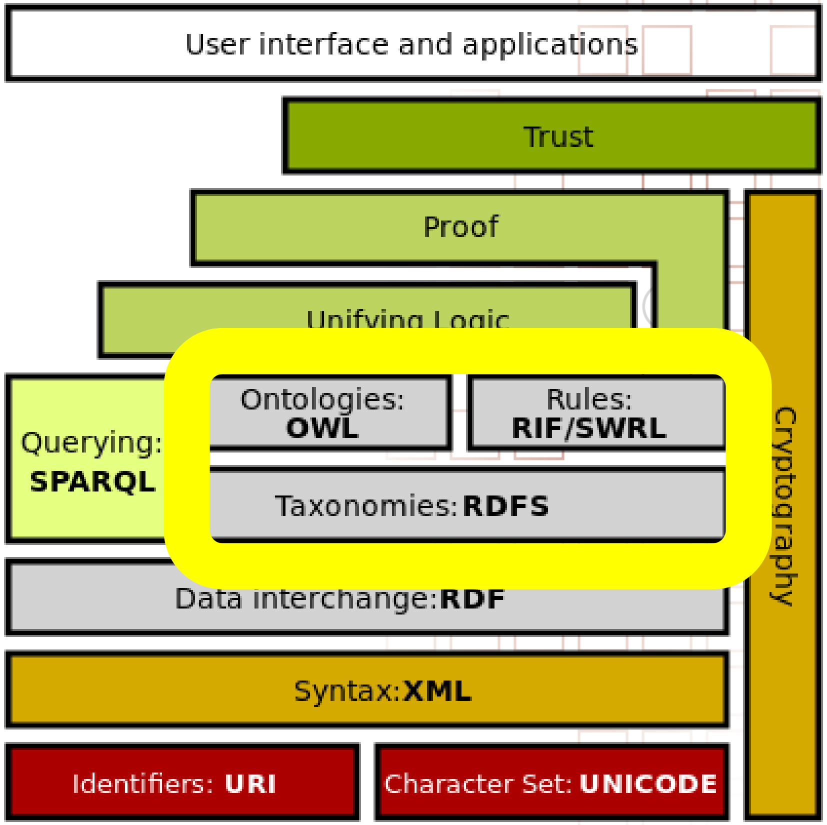
>> Web of Data <<

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).

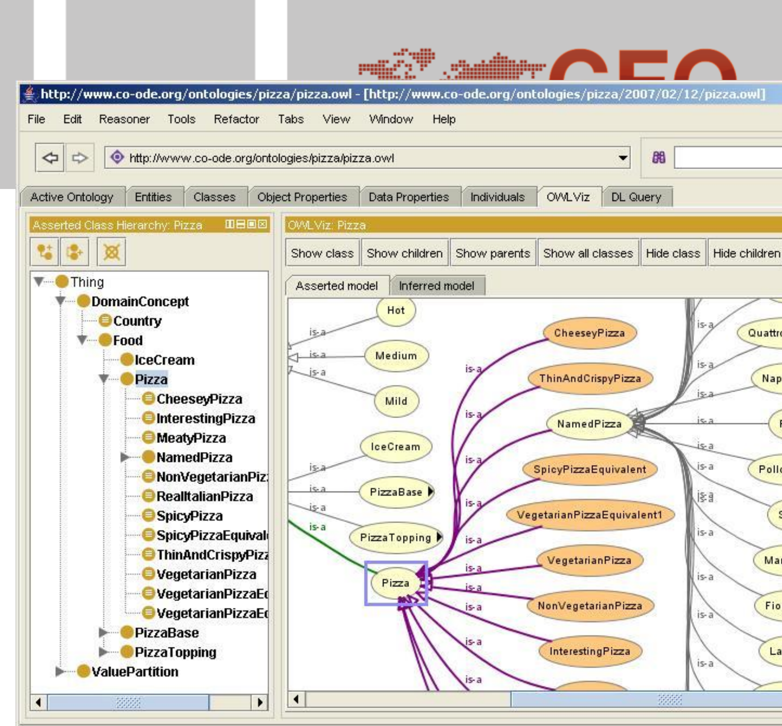
- Combining Linked Data and Geoinformation can lead to a geospatially enriched Semantic Web
  - Geographic information can easily be integrated and processed.
  - But: requires semantics (Ontologies, Taxonomies)

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



Knowledge Graphs & Ontologies

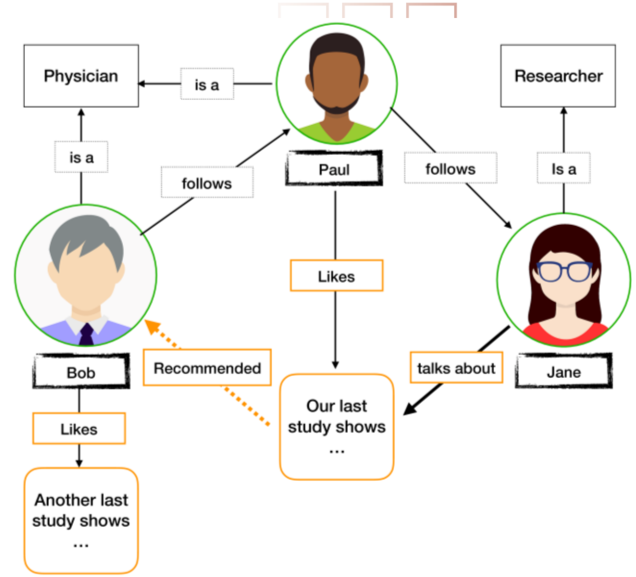
- Ontology:
  - Formal, explicit specification of a shared conceptualization (Gruber, 1993)
  - Description of the concepts and their relations existing in a Universe of Discourse (Uschold & Gruninger, 1996)



- 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."

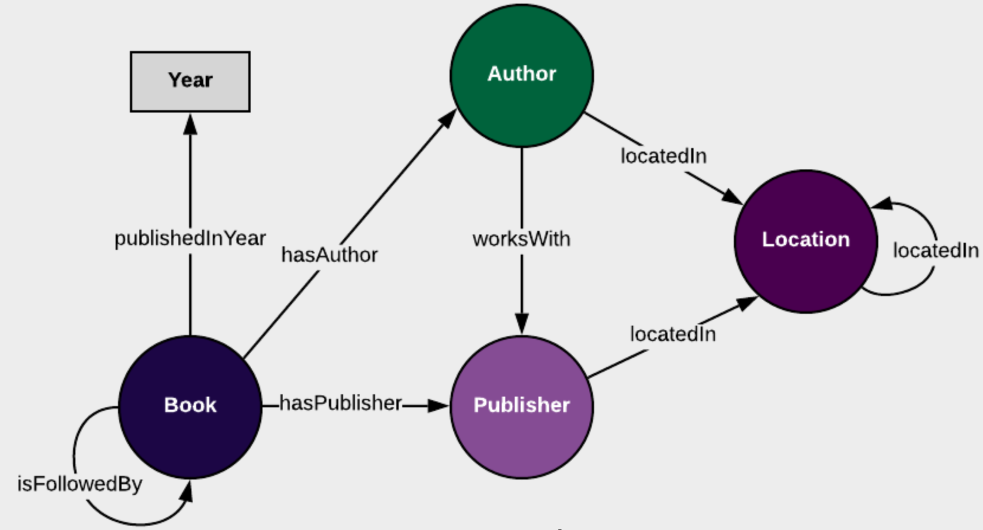
  (Paulheim 2017)



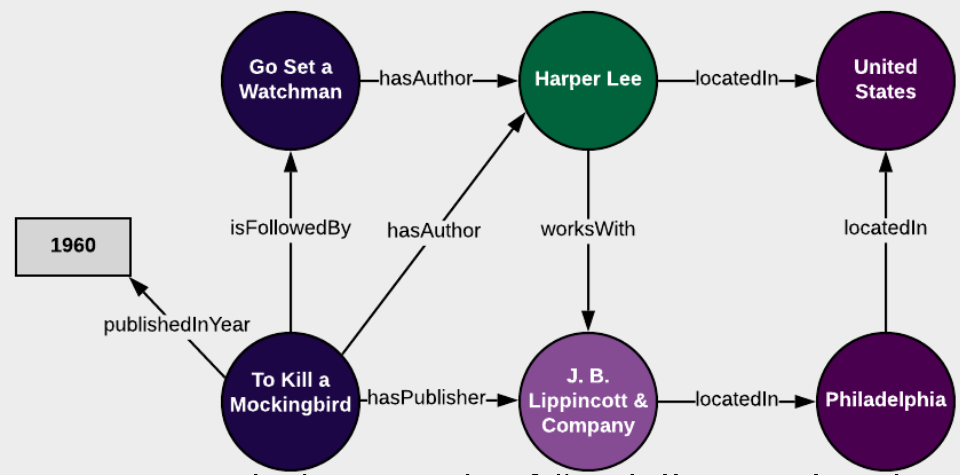
Knowledge Graphs & Ontologies

- 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



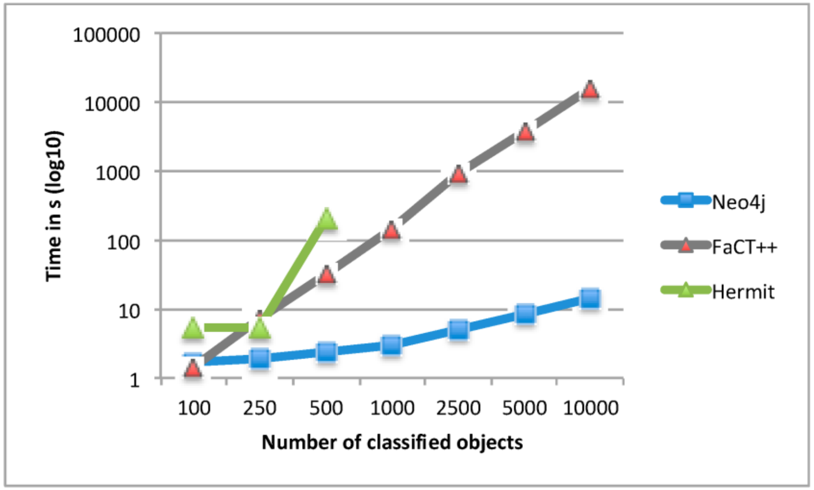
Knowledge Graph of “To kill a mockingbird”

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)

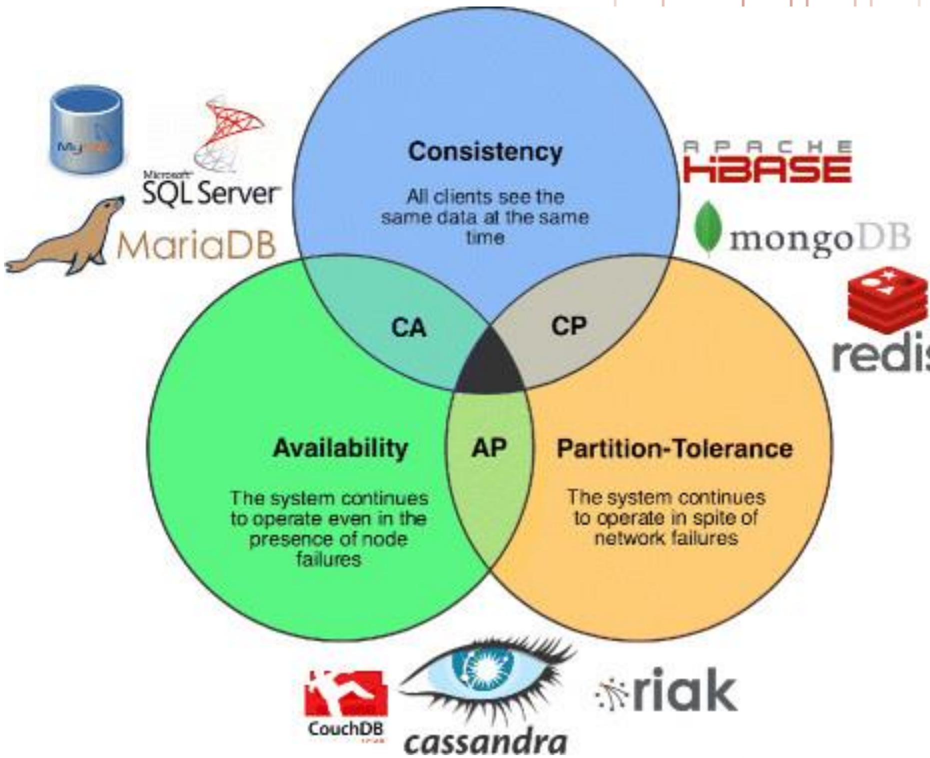


Classification speed of EO data (Lampoltshammer & Wiegand 2015)

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 Domino
- **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?

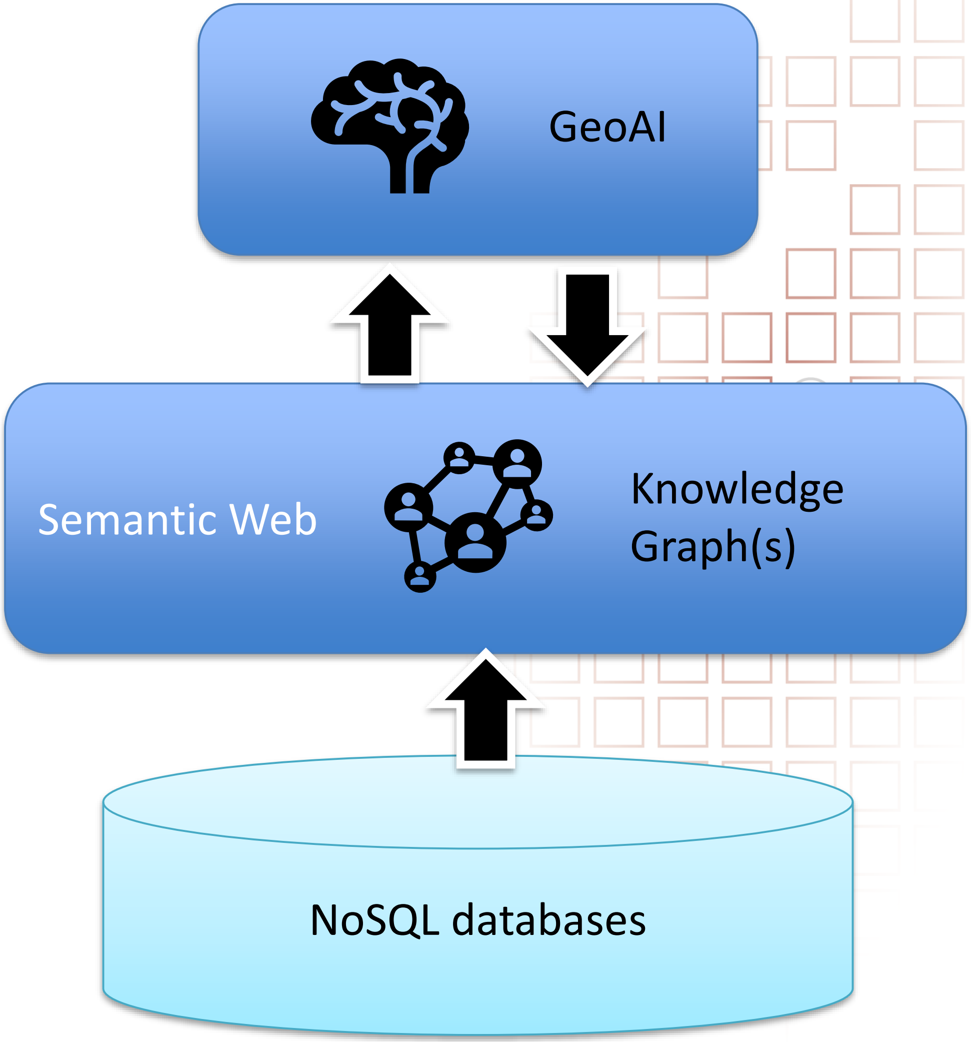
GeoAI || Knowledge Graphs || NoSQL

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

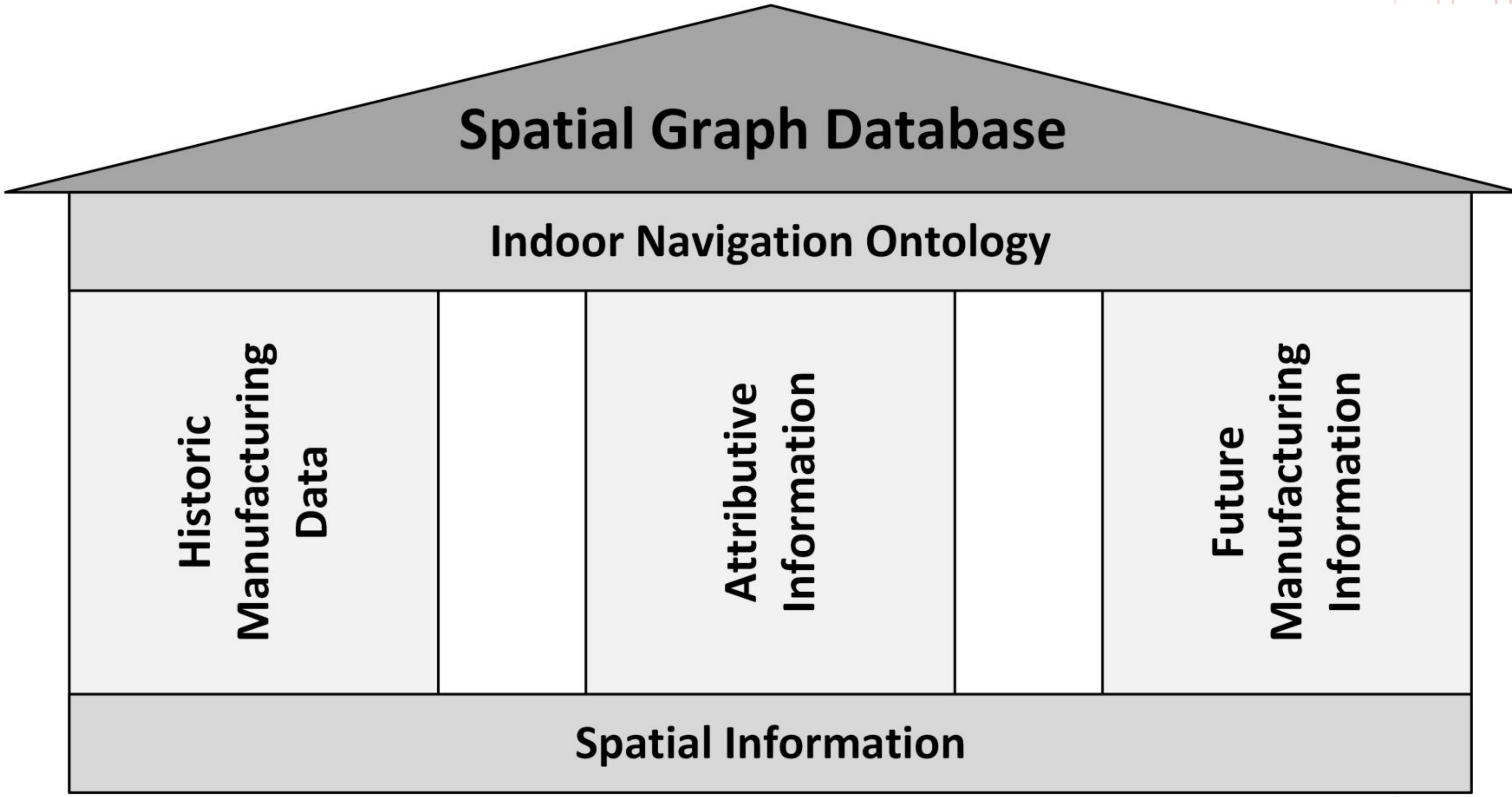
Application Examples

Indoor Geography and Smart Manufacturing

- 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)

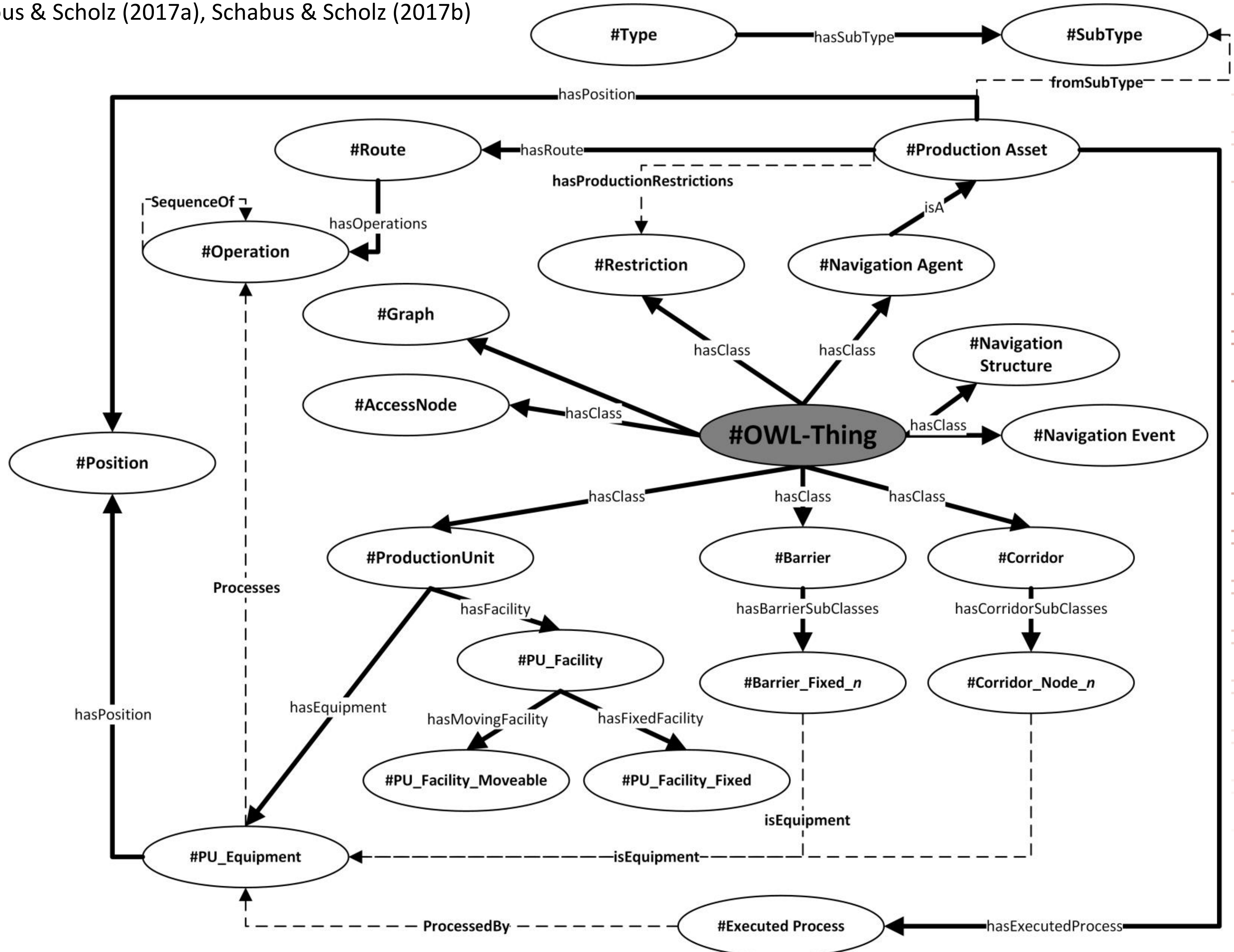
Indoor Geography and Smart Manufacturing



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

Indoor Geography and Smart Manufacturing

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



Selfish Routing & Agent-based Simulation

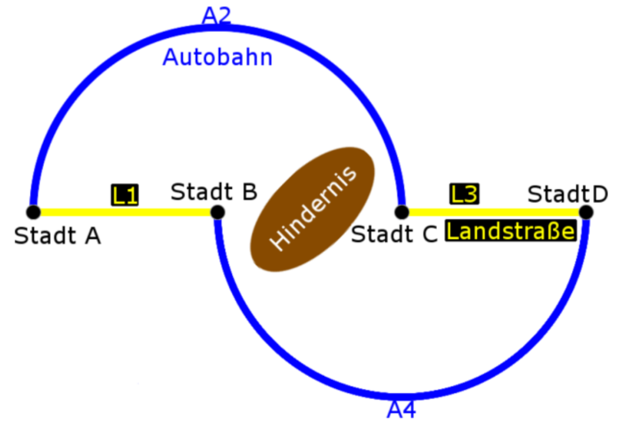
- **Selfish routing** is a result of different agents acting in a network, trying to find the **best route** from a **strictly personal viewpoint**, regardless of the consequence for other agents.

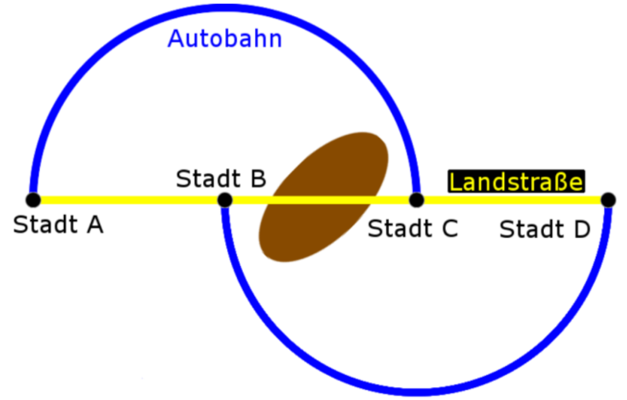
- Based on the **Braess Paradox** (Braess 1969, Roughgarden 2005)

- 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)

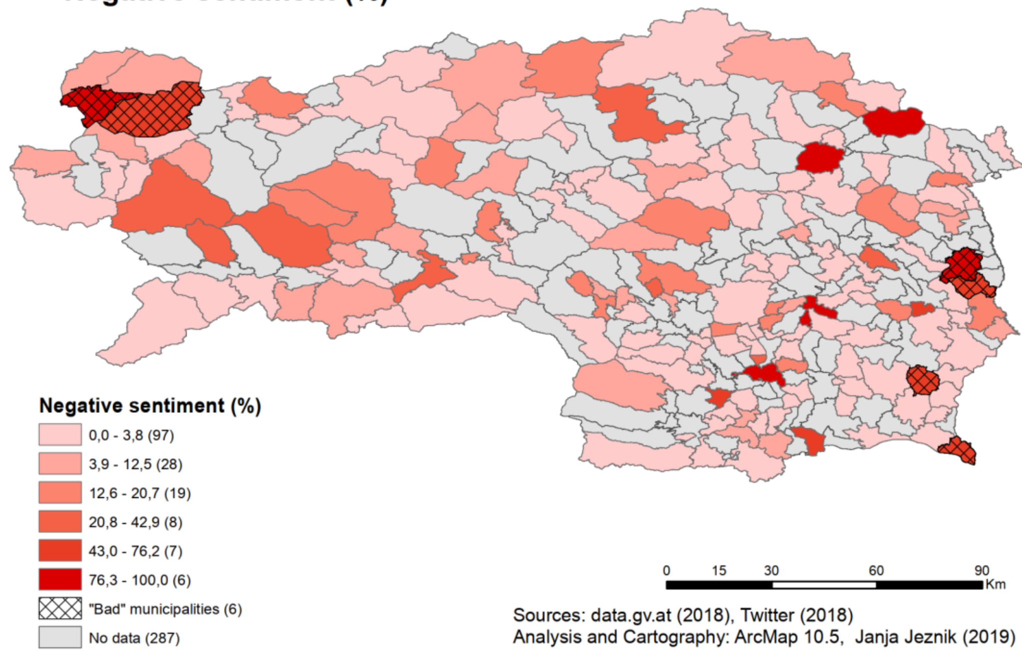
- Each agent is equipped with a predictive memory
  (Scholz 2015; Exenberger & Scholz forthcoming)
  - Graph-based memory structure (individual experiences and outcomes)
  - Reinforcement learning (i.e. Machine Learning) to match current traffic situations with historic experiences
  - Decision making 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 forthcoming)
  - MongoDB as basis for Sentiment analysis
  - Spatio-temporal analysis

Negative sentiment (%)



Research Frontiers

Some lessons learned so far…

- Relational machine learning models treat
  - **Geographic entities** as **ordinary** entities
  - hence **spatial footprints** of places are **neglected**
  - and the **distance decay** effect is **ignored**.

  >> suboptimal performance in: geospatial knowledge graph completion, geographic question answering, geographic entity alignment, as well as geographic knowledge graph summarization

- Large scale neural symbolic reasoning based on unstructured text is still to be developed

- Automatic (Geo)knowledge Graph construction is still in it's infancy

Frontiers of GeoKG & GeoAI

The 1st International Workshop on

Methods, Models, and Resources for Geospatial Knowledge Graphs and GeoAI

co-located with GIScience 2020, Poznań, Poland

Workshop Date Update:
Due to the uncertain impacts of COVID-19 in next months, the organizing committee has decided to postpone the workshop (GeoKG & GeoAI 2020) in conjunction with the GIScience conference until Fall 2021.

Call For Paper

The rapid increase in high-quality data, advanced machine learning algorithms, and the availability of fast hardware have largely contributed to a renewed interest in Artificial Intelligence (AI). Despite many successful stories in computer vision, natural

https://stko.geog.ucsb.edu/geokg-geoai2020/

Frontiers of GeoKG & GeoAI

Selected Topics from CfP of GeoKG & GeoAI Workshop:

- Deep Learning and Reinforcement Learning on Geospatial Knowledge Graphs
- GeoKG Construction & GeoOntology Engineering
- Geographic Information Retrieval and Geo-Text Analysis
- GeoAI Resources and Infrastructures
- Other GeoAI Topics
  - Spatial Optimization
  - Spatial Simulation
- Combination of
  - representation learning techniques (Connectionist Artificial Intelligence)
  - with symbolic representation and reasoning associated with knowledge graphs (Symbolic Artificial Intelligence)

  to develop scalable and interpretable machine learning models

Selected References

- Egenhofer, M. J. (2002). Toward the semantic geospatial web. In *Proceedings of the 10th ACM international symposium on Advances in geographic information systems* (pp. 1-4). ACM.
- Duan, W., Chiang, Y.-Y., Leyk, S., Uhl, J., and Knoblock, C. Automatic alignment of contemporary vector data and georeferenced historical maps using reinforcement learning. International Journal of Geographical Information Science, 824-849, 2020
- Hey, A. J., Tansley, S., Tolle, K. M., and others, . The fourth paradigm: data-intensive scientific discovery, volume 1. Microsoft research Redmond, WA, 2009.
- Janowicz, K., van Harmelen, F., Hendler, J. A., and Hitzler, P. Why the data train needs semantic rails. AI Magazine, 36(1):5–14, Mar. 2015
- Janowicz, K., Scheider, S., & Adams, B. (2013). A geo-semantics flyby. In *Reasoning web. Semantic technologies for intelligent data access* (pp. 230-250). Springer Berlin Heidelberg.
- Lampoltshammer, T. J., & Wiegand, S. (2015). Improving the computational performance of ontology-based classification using graph databases. *Remote Sensing*, *7*(7), 9473-9491.
- Li, W. and Hsu, C.-Y. Automated terrain feature identification from remote sensing imagery: a deep learning approach. International Journal of Geographical Information Science, pages 1–24, 2020.
- Lourenço, J. R., Cabral, B., Carreiro, P., Vieira, M., & Bernardino, J. (2015). Choosing the right NoSQL database for the job: a quality attribute evaluation. *Journal of Big Data*, *2*(1), 18.
- Mai G., Yan B., Janowicz K., Zhu R. (2020) Relaxing Unanswerable Geographic Questions Using A Spatially Explicit Knowledge Graph Embedding Model. In: Kyriakidis P., Hadjimitsis D., Skarlatos D., Mansourian A. (eds) Geospatial Technologies for Local and Regional Development. AGILE 2019. Lecture Notes in Geoinformation and Cartography. Springer, Cham
- Paulheim, Heiko. "Knowledge graph refinement: A survey of approaches and evaluation methods." *Semantic web* 8.3 (2017): 489-508.
- Ren, Y., Chen, H., Han, Y., Cheng, T., Zhang, Y., and Chen, G. A hybrid integrated deep learning model for the prediction of citywide spatio-temporal flow volumes. International Journal of Geographical Information Science, pages 1–22, 2020
- Scholz J., Church RL. (2018). Shortest Paths from a Group Perspective—A Note on Selfish Routing Games with Cognitive Agents. ISPRS International Journal of Geo-Information 7, no. 9: 345. DOI: 10.3390/ijgi7090345
- Scholz, J.(2015): Shortest Paths for Groups: Introducing a Predictive Memory for Cognitive Agents. GI_Forum 2015 - Geospatial Minds for Society (Journal): 571-574. DOI:10.1553/giscience2015s571
- Scholz, J.: Shortest Paths from a Group Perspective - a Note on Selfish Routing Games with Cognitive Agents. Proceedings of ICA Workshop on Street Networks and Transport 2013.
- Scholz, J. and Schabus, S. (2014): An Indoor Navigation Ontology for Production Assets in a Production Environment. In: Stewart, K., Pebesma, E., Navratil, G., Fogliaroni, P., Duckham, M. (Eds.): Geographic Information Science 2014, Lecture Notes in Computer Science, LNCS 8728, pp. 204–220, Springer
- Scholz, J., and Schabus, S. (2017). "Towards an Affordance-Based Ad-Hoc Suitability Network for Indoor Manufacturing Transportation Processes." ISPRS Int. J. Geo-Inf. 6, no. 9: 280. DOI: 10.3390/ijgi6090280
- Schabus, S. and Scholz, J. (2017a). Spatially-Linked Manufacturing Data to Support Data Analysis. GI_Forum 2017 (Journal) 1:126 - 140. DOI:10.1553/giscience2017_01_s12
- Schabus S., Scholz J. (2017b). Semantically Annotated Manufacturing Data to support Decision Making in Industry 4.0: A Use-Case Driven Approach. In: Haber P., Lampoltshammer T., Mayr M. (eds) Data Science – Analytics and Applications. pp. 97-102. Springer Vieweg, Wiesbaden. DOI: https://doi.org/10.1007/978-3-658-19287-7_14
- Xie, Y., Cai, J., Bhojwani, R., Shekhar, S., and Knight, J. A locally-constrained yolo framework for detecting small and densely-distributed building footprints. International Journal of Geographical Information Science, pages 777-801, 2020

New Research Frontiers:
GeoAI, (Geo-)Knowledge Graphs and NoSQL Databases

GI Research Colloquium 2020 @ CUAS

Johannes Scholz

Institute of Geodesy – TU Graz
Head of RG Geoinformation

Mail: johannes.scholz@tugraz.at
Web: ifg.tugraz.at || www.johannesscholz.net
Twitter: @Joe_GISc