"Natural Data Understanding": LLMs, Agents, Knowledge Graphs - why they need each other to solve (your?) Data problems?

Axel Polleres
Inst. for Data, Process & Knowledge Management, WU Wien

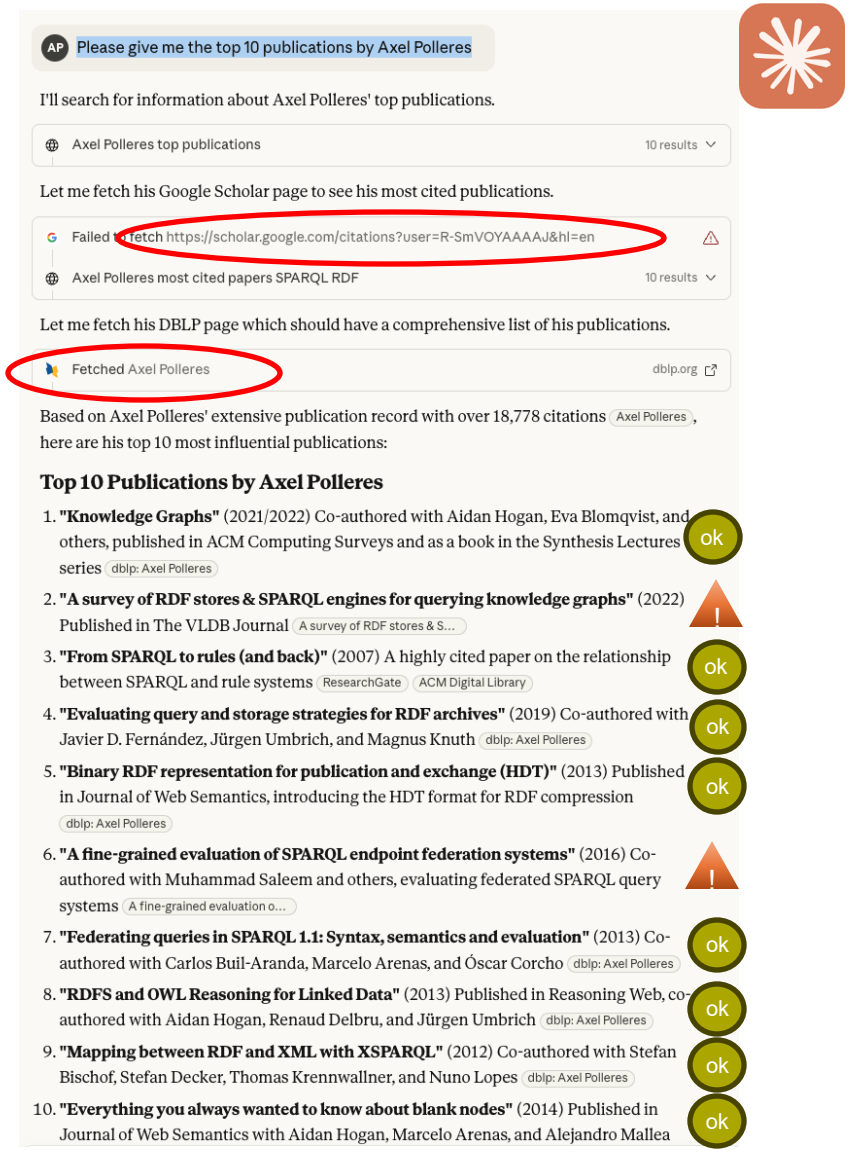
Funded by:
FWF Austrian Science Fund

CLUSTER OF EXCELLENCE

BilAI
Bilateral Artificial Intelligence

bilateral-ai.net

LLMs get better … but still (plausibly) hallucinate “are supposed to be PhD level”



AP Please give me the top 10 publications by Axel Polleres

I'll search for information about Axel Polleres' top publications.

Axel Polleres top publications — 10 results

Let me fetch his Google Scholar page to see his most cited publications.

Failed to fetch https://scholar.google.com/citations?user=R-SmVOYAAAAJ&hl=en

Axel Polleres most cited papers SPARQL RDF — 10 results

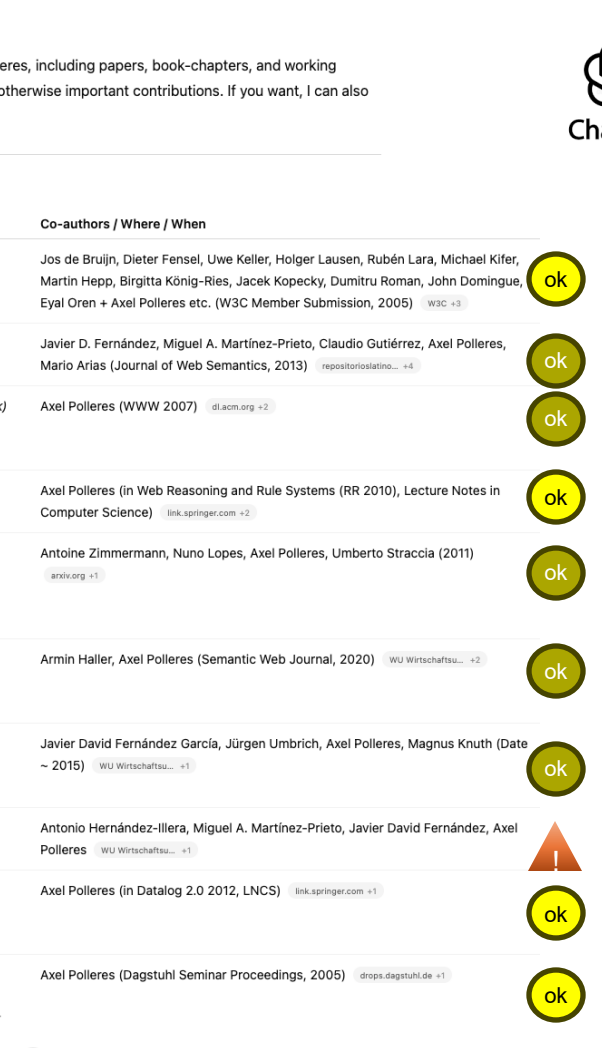
Let me fetch his DBLP page which should have a comprehensive list of his publications.

Fetched Axel Polleres — dblp.org

Based on Axel Polleres' extensive publication record with over 18,778 citations Axel Polleres, here are his top 10 most influential publications:

Top 10 Publications by Axel Polleres

1. **"Knowledge Graphs"** (2021/2022) Co-authored with Aidan Hogan, Eva Blomqvist, and others, published in ACM Computing Surveys and as a book in the Synthesis Lectures series dblp: Axel Polleres — ok
2. **"A survey of RDF stores & SPARQL engines for querying knowledge graphs"** (2022) Published in The VLDB Journal A survey of RDF stores & S… — !
3. **"From SPARQL to rules (and back)"** (2007) A highly cited paper on the relationship between SPARQL and rule systems ResearchGate ACM Digital Library — ok
4. **"Evaluating query and storage strategies for RDF archives"** (2019) Co-authored with Javier D. Fernández, Jürgen Umbrich, and Magnus Knuth dblp: Axel Polleres — ok
5. **"Binary RDF representation for publication and exchange (HDT)"** (2013) Published in Journal of Web Semantics, introducing the HDT format for RDF compression dblp: Axel Polleres — ok
6. **"A fine-grained evaluation of SPARQL endpoint federation systems"** (2016) Co-authored with Muhammad Saleem and others, evaluating federated SPARQL query systems A fine-grained evaluation o… — !
7. **"Federating queries in SPARQL 1.1: Syntax, semantics and evaluation"** (2013) Co-authored with Carlos Buil-Aranda, Marcelo Arenas, and Óscar Corcho dblp: Axel Polleres — ok
8. **"RDFS and OWL Reasoning for Linked Data"** (2013) Published in Reasoning Web, co-authored with Aidan Hogan, Renaud Delbru, and Jürgen Umbrich dblp: Axel Polleres — ok
9. **"Mapping between RDF and XML with XSPARQL"** (2012) Co-authored with Stefan Bischof, Stefan Decker, Thomas Krennwallner, and Nuno Lopes dblp: Axel Polleres — ok
10. **"Everything you always wanted to know about blank nodes"** (2014) Published in Journal of Web Semantics with Aidan Hogan, Marcelo Arenas, and Alejandro Mallea — ok



Here are **10 influential publications** by Axel Polleres, including papers, book-chapters, and working documents. These are among his most cited or otherwise important contributions. If you want, I can also fetch full download links or more recent ones.

Top 10 Publications by Axel Polleres

#	Title	Co-authors / Where / When	
1	*Web Service Modeling Ontology (WSMO)*	Jos de Bruijn, Dieter Fensel, Uwe Keller, Holger Lausen, Rubén Lara, Michael Kifer, Martin Hepp, Birgitta König-Ries, Jacek Kopecky, Dumitru Roman, John Domingue, Eyal Oren + Axel Polleres etc. (W3C Member Submission, 2005) W3C +3	ok
2	*Binary RDF representation for publication and exchange (HDT)*	Javier D. Fernández, Miguel A. Martínez-Prieto, Claudio Gutiérrez, Axel Polleres, Mario Arias (Journal of Web Semantics, 2013) repositoriolatino… +4	ok
3	*From SPARQL to Rules (and back)*	Axel Polleres (WWW 2007) dl.acm.org +2	ok
4	*SPARQL1.1: New Features and Friends (OWL2, RIF)*	Axel Polleres (in Web Reasoning and Rule Systems (RR 2010), Lecture Notes in Computer Science) link.springer.com +2	ok
5	*A General Framework for Representing, Reasoning and Querying with Annotated Semantic Web Data*	Antoine Zimmermann, Nuno Lopes, Axel Polleres, Umberto Straccia (2011) arxiv.org +1	ok
6	*Are We Better Off With Just One Ontology on the Web?*	Armin Haller, Axel Polleres (Semantic Web Journal, 2020) WU Wirtschaftsu… +2	ok
7	*Evaluating Query and Storage Strategies for RDF Archives*	Javier David Fernández García, Jürgen Umbrich, Axel Polleres, Magnus Knuth (Date ~ 2015) WU Wirtschaftsu… +1	ok
8	*Serializing RDF in Compressed Space (HDT++)*	Antonio Hernández-Illera, Miguel A. Martínez-Prieto, Javier David Fernández, Axel Polleres WU Wirtschaftsu… +1	!
9	*How (Well) Do Datalog, SPARQL and RIF Interplay?*	Axel Polleres (in Datalog 2.0 2012, LNCS) link.springer.com +1	ok
10	*Semantic Web Languages and Semantic Web Services as Application Areas for Answer Set Programming*	Axel Polleres (Dagstuhl Seminar Proceedings, 2005) drops.dagstuhl.de +1	ok

ChatGPT

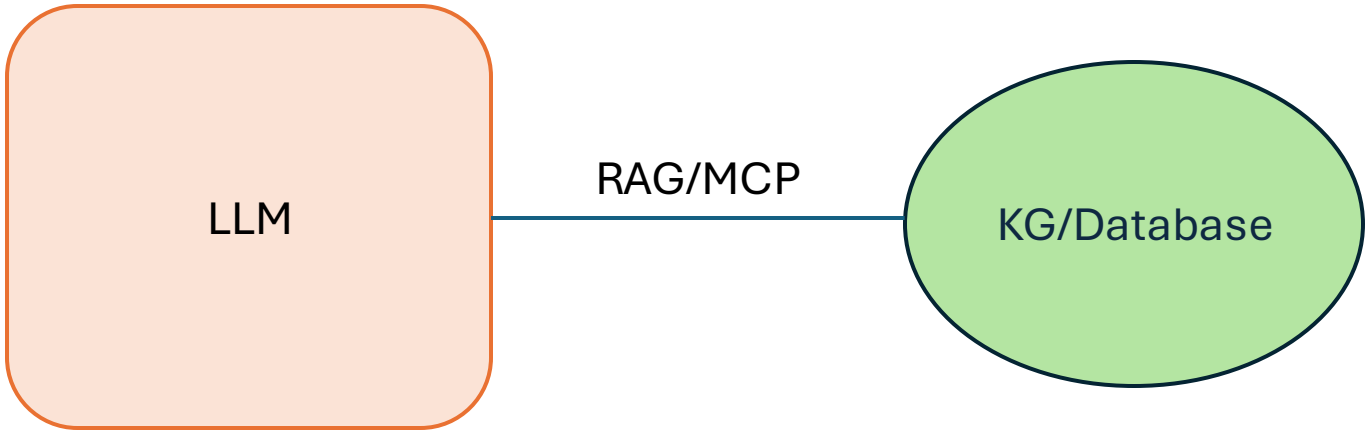
Inspired by Georg Gottlob’s SEMANTiCS2025 keynote

Antother key insight from SEMANTiCS – Heng Ji’s keynote

Larger models are not necessarily better!

Funded by: FWF Austrian Science Fund

Knowledge Graphs (and Databases) to the rescue...



Funded by:
FWF Austrian Science Fund

Knowledge Graphs (and Databases) to the rescue…

Closed KGs

- Google KG
- Enterprise KGs
- …

Open KGs

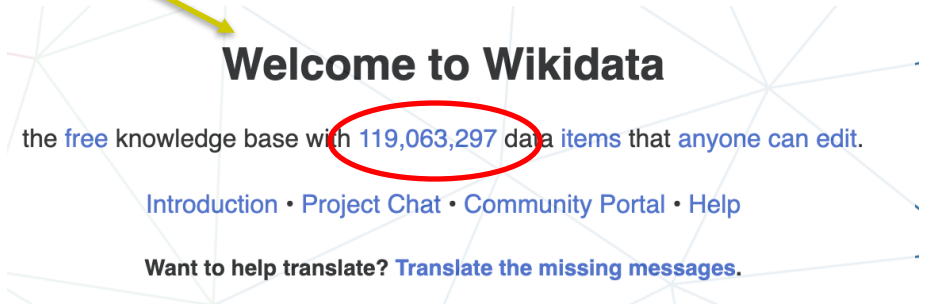
- Wikidata
- DBpedia
- DBLP
- …

Open KGs (April 2021)	
DBpedia	~4.58m entities, ~9.25GB
Yago4	~50m entities, ~18.4GB
Wikidata	~93m entities, ~99GB

Closed KGs (~2019)	
Microsoft	~2bn entities, ~55bn facts
Google	~1bn entities, ~70bn assertions
Facebook	~50m entities, ~500m assertions
eBay	~1bn triples
IBM	~100m entities, 5bn relationships

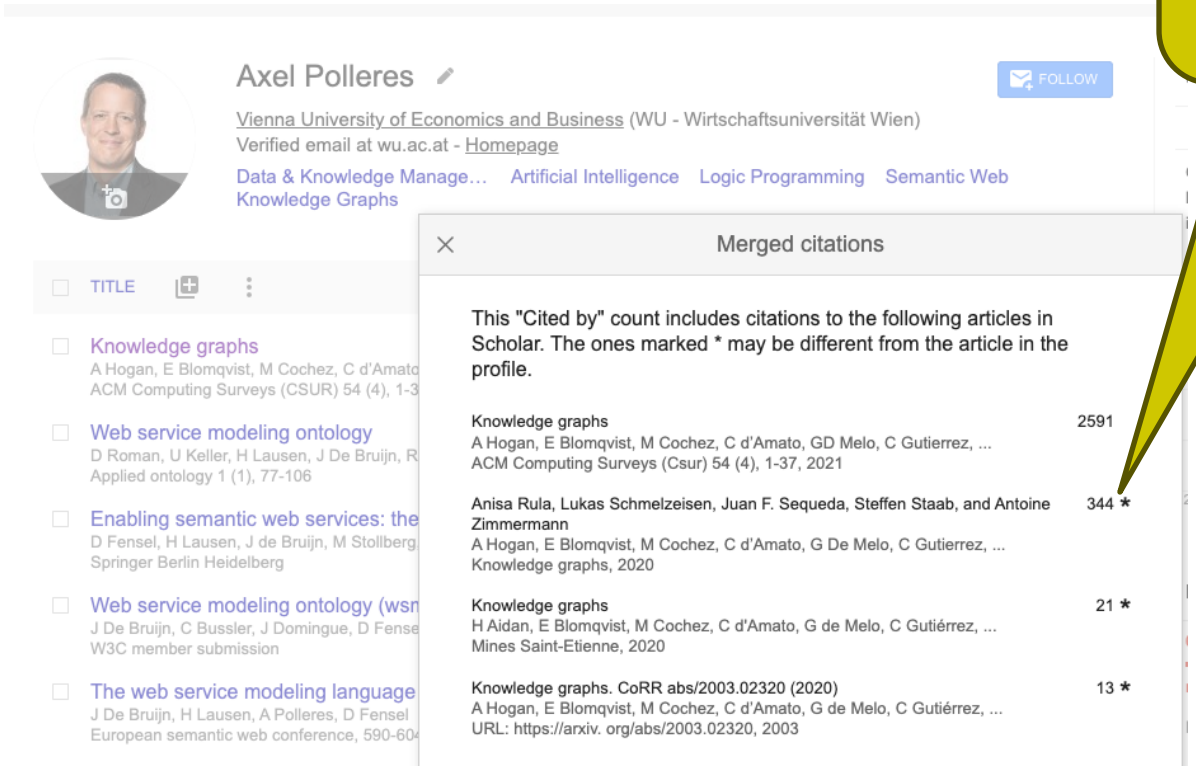
N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, J. Taylor: Industry-scale Knowledge Graphs: Lessons and Challenges. ACM Queue 17(2), 2019.
A. Hogan et al.: Knowledge Graphs. CoRR, 02320, 2020.

Rapidly growing!



Funded by: FWF Austrian Science Fund

But are (your?) KGs & Databases actually better?



Google asks it's users to manually correct/merge results

DBLP or scopus attempt to curate ambiguities by constraints (not always successful..)

Funded by:
FWF Austrian Science Fund

But are (your?) KGs & Databases actually better?

Example from Wikidata…

Tilikum (Q127236)

Item Discussion Read View history ☆

dugout canoe edit

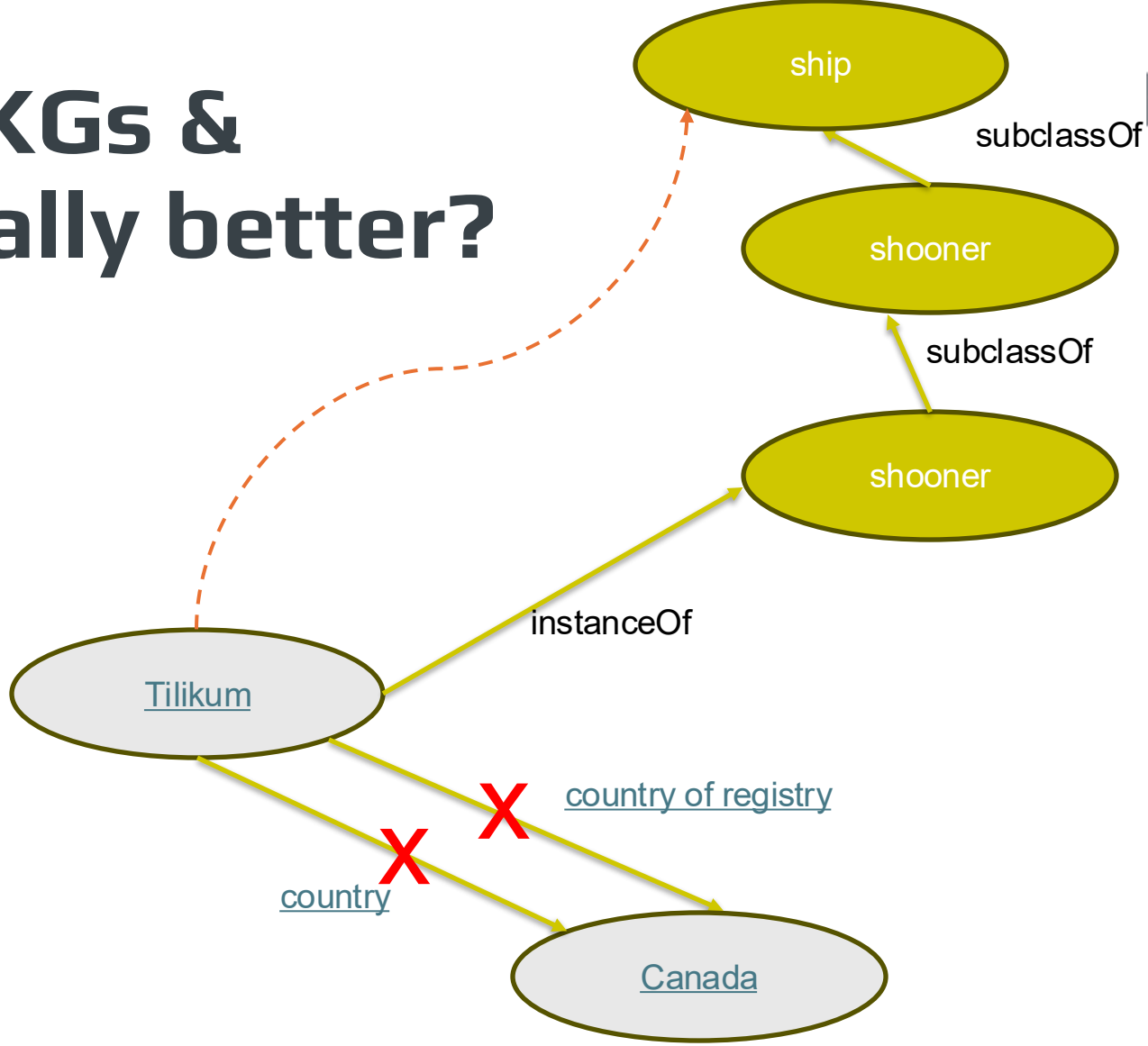
▾ In more languages

Configure

Language	Label	Description	Also known as
default for all languages	No label defined	–	
English	Tilikum	dugout canoe	
British English	Tilikum	No description defined	
American English	No label defined	No description defined	
German	Tilikum	Segelboot	

Funded by: FWF Austrian Science Fund

But are (your?) KGs & Databases actually better?



Ontologies +

Constraints,

such as:

~~*"country conflicts with country of registry"*~~

- *"i.e., an entity should not have both a "country" and a "country of registry"*

Funded by: FWF Austrian Science Fund

Why KGs can't (scalably) be correct either:

Collaborative Editing

Multi-sourced Automatic Extraction

Knowledge is distributed and evolving!

- ***Remote changes affect consistency***
- ***Ontologies evolve as well***

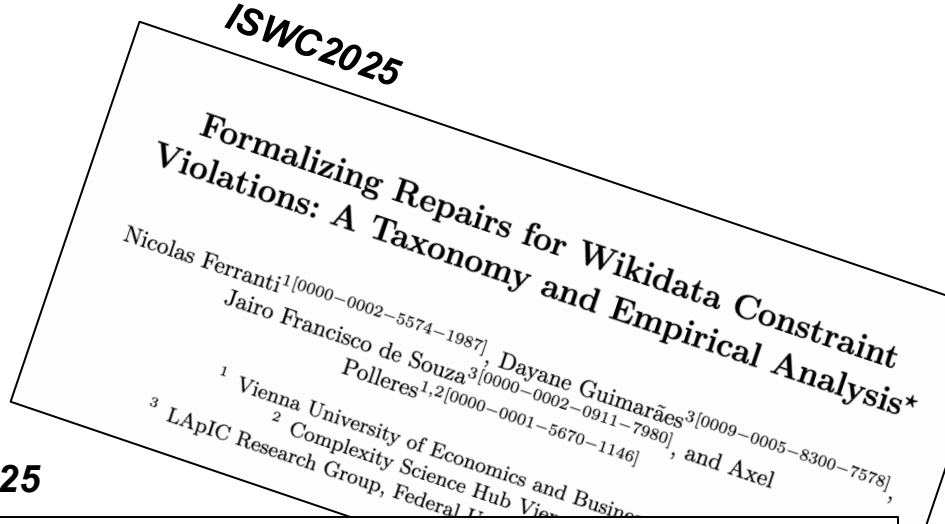
What can we do about it?

- Best of both worlds needs **Bilateral AI** approaches!
  - **Symbolic** methods to repair inconsistencies

Forthcoming…

- **Subsymbolic** methods to resolve inconsistencies
  - **LLM**s for resolving inconsistencies
  - **GNNs** for *learning* repairs (*from historical data*, from users' repairs)
  - Iterative **RAG** and **Agent-based** pipelines to fix and construct KGs from text

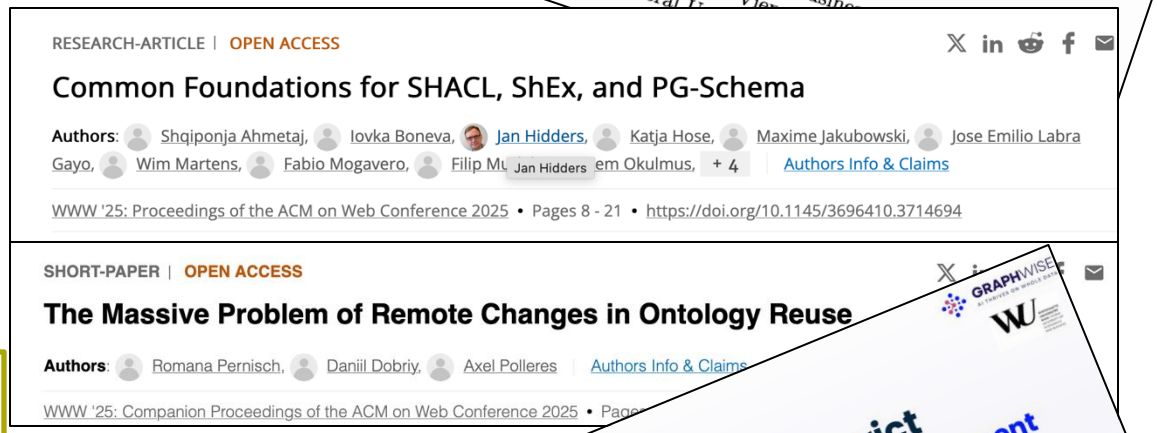
ISWC2025


Formalizing Repairs for Wikidata Constraint Violations: A Taxonomy and Empirical Analysis*

Nicolas Ferranti[1][0000−0002−5574−1987], Dayane Guimarães[3][0009−0005−8300−7578], Jairo Francisco de Souza[3][0000−0002−0911−7980], and Axel Polleres[1,2][0000−0001−5670−1146]

[1] Vienna University of Economics and Business
[2] Complexity Science Hub Vienna
[3] LApIC Research Group, Federal University of Juiz de Fora

Webconf2025


RESEARCH-ARTICLE | OPEN ACCESS

Common Foundations for SHACL, ShEx, and PG-Schema

Authors: Shqiponja Ahmetaj, Iovka Boneva, Jan Hidders, Katja Hose, Maxime Jakubowski, Jose Emilio Labra Gayo, Wim Martens, Fabio Mogavero, Filip Murlak, Magdalem Okulmus, + 4 Authors Info & Claims

WWW '25: Proceedings of the ACM on Web Conference 2025 • Pages 8 - 21 • https://doi.org/10.1145/3696410.3714694

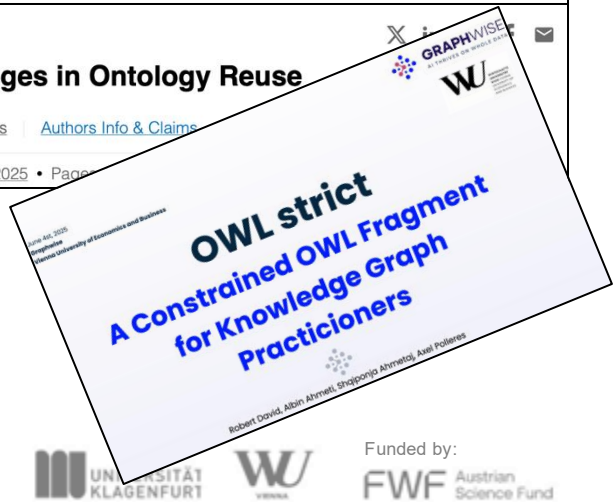
SHORT-PAPER | OPEN ACCESS

The Massive Problem of Remote Changes in Ontology Reuse

Authors: Romana Pernisch, Daniil Dobriy, Axel Polleres Authors Info & Claims

WWW '25: Companion Proceedings of the ACM on Web Conference 2025 • Pages

ESWC2025



Funded by: FWF Austrian Science Fund

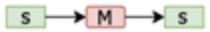
What's next?

Neurosymbolic AI Systems
Prof. Marta Sabou

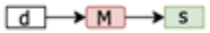
Ingredient1: We have a pretty good understanding of Engineering **Neuro-Symbolic Systems…**

Atomic-Patterns

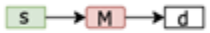
A1: s-M-s



A2: d-M-s

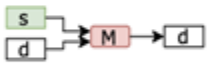


A3: s-M-d

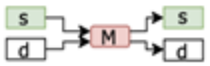


Fusion-Patterns

F1: d/s-M-d

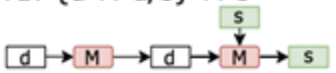


F2: d/s-M-s

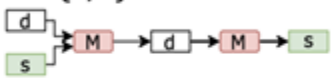


T-Patterns

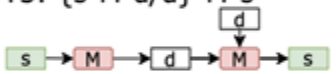
T1: {d-M-d/s}-M-s



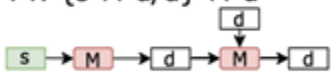
T2: {d/s}-M-d-M-s



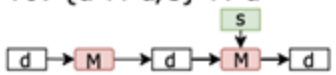
T3: {s-M-d/d}-M-s



T4: {s-M-d/d}-M-d

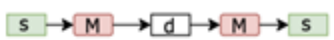


T6: {d-M-d/s}-M-d

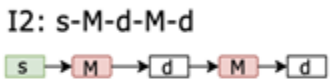


I-Patterns

I1: s-M-d-M-s

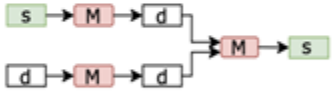


I2: s-M-d-M-d

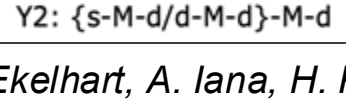


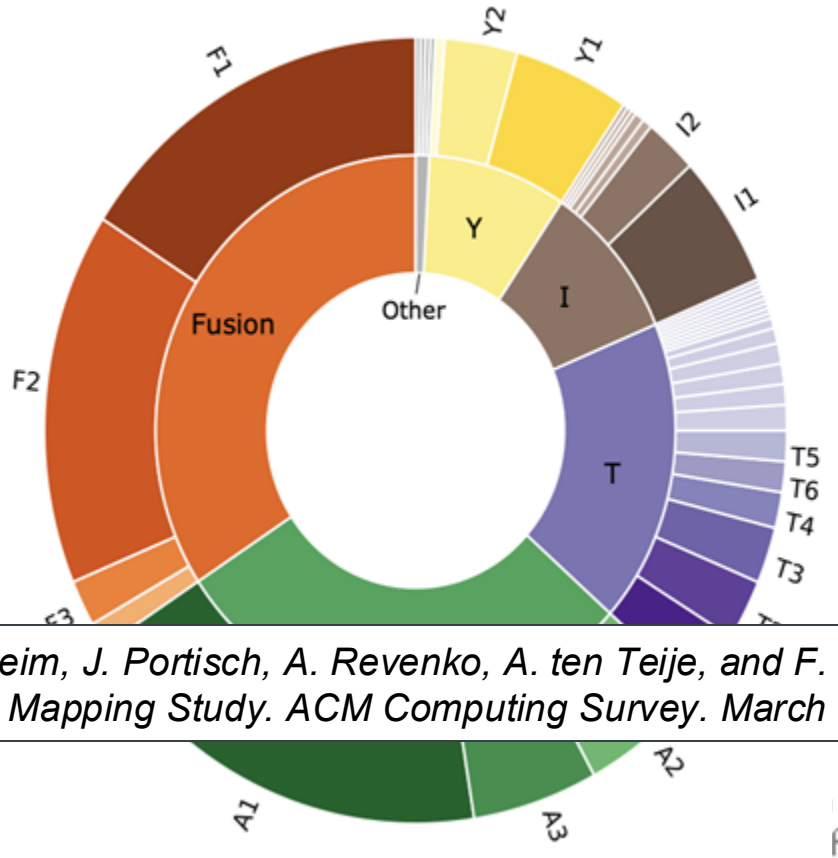
Y-Patterns

Y1: {s-M-d/d-M-d}-M-s



Y2: {s-M-d/d-M-d}-M-d





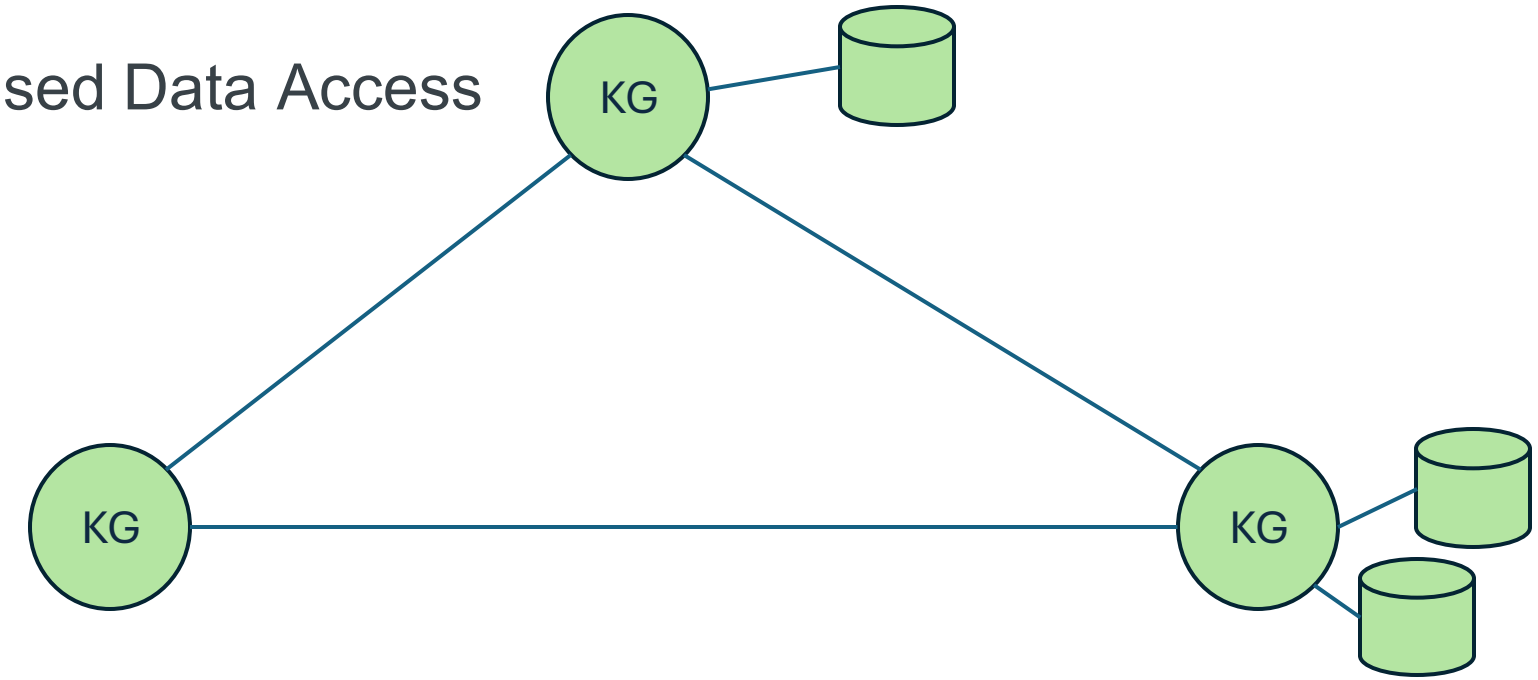
A. Breit, L. Waltersdorfer, F.J. Ekaputra, M. Sabou, A. Ekelhart, A. Iana, H. Paulheim, J. Portisch, A. Revenko, A. ten Teije, and F. van Harmelen. 2023. Combining Machine Learning and Semantic Web: A Systematic Mapping Study. ACM Computing Survey. March 2023.

by:
Austrian Science Fund

What's next? Where should this go?

Ingredient 2: Expertise on Symbolic Data Integration and Repair. Before the LLM hype, we long worked on ***purely symbolic*** solutions to integrating decentralised Knowledge

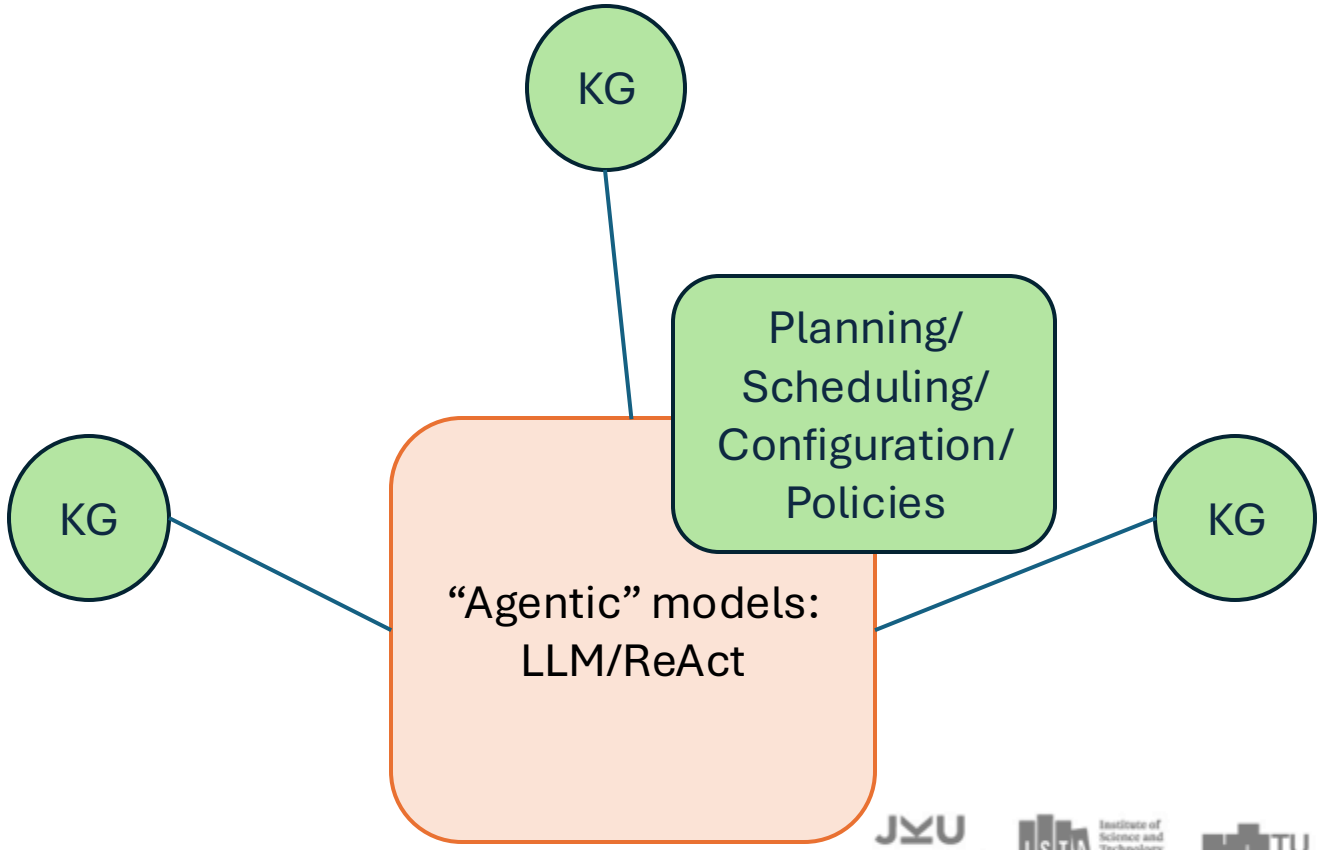
- Linked Data
- Ontology-based Data Access



Funded by: FWF Austrian Science Fund

What's next? Where should this go?

Ingredient3: *Combinations* of LLMs and decentral KGs via **RAG** and **Agent frameworks** are one example of such Neuro-symbolic architectures:



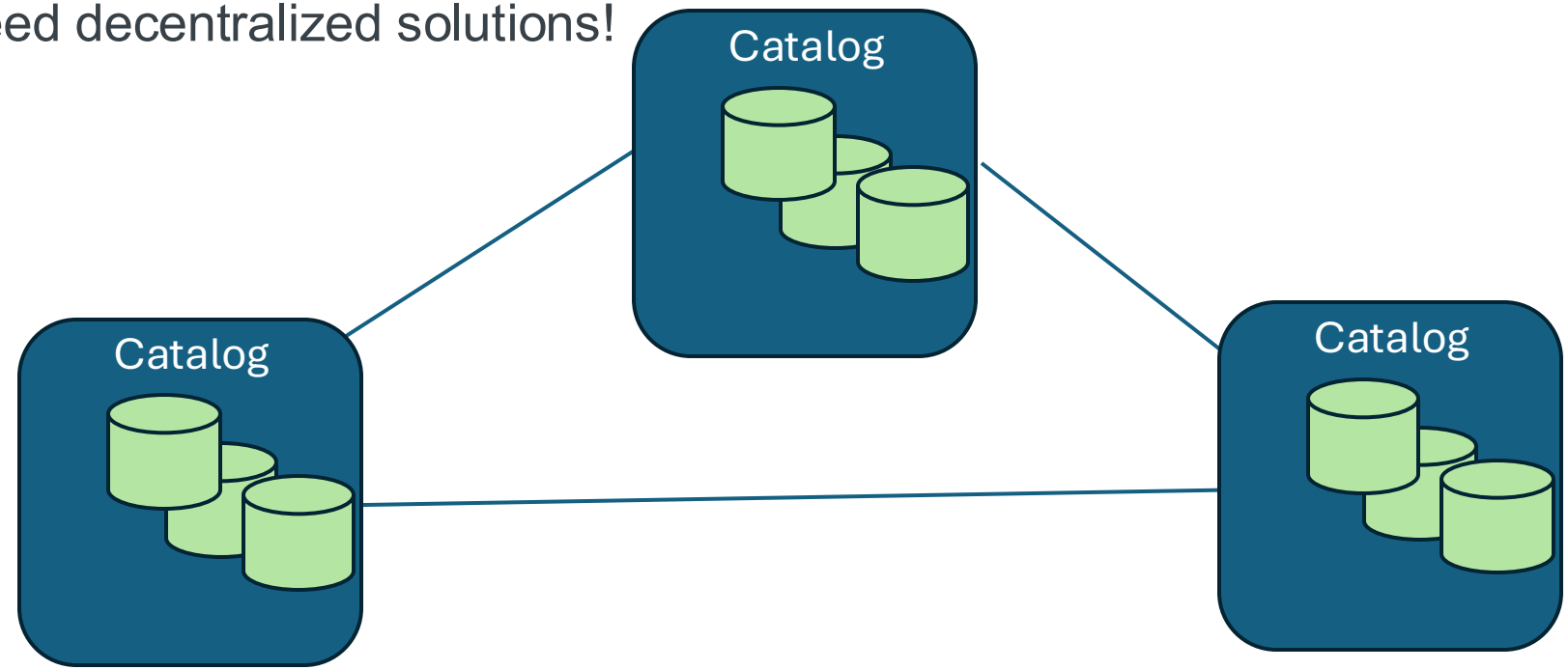
Funded by:
FWF Austrian Science Fund

What's next? Where should this go?

Caveat: Our data and knowledge is

- **distributed** and
- **not always sharable** with a central (API) model
- → we need decentralized solutions!

"Our knowledge will never go into a central LLM" – keynote Stefan Rohringer (Infineon) *this morning*



Ingredient 4: "Data Spaces"

Funded by: FWF Austrian Science Fund

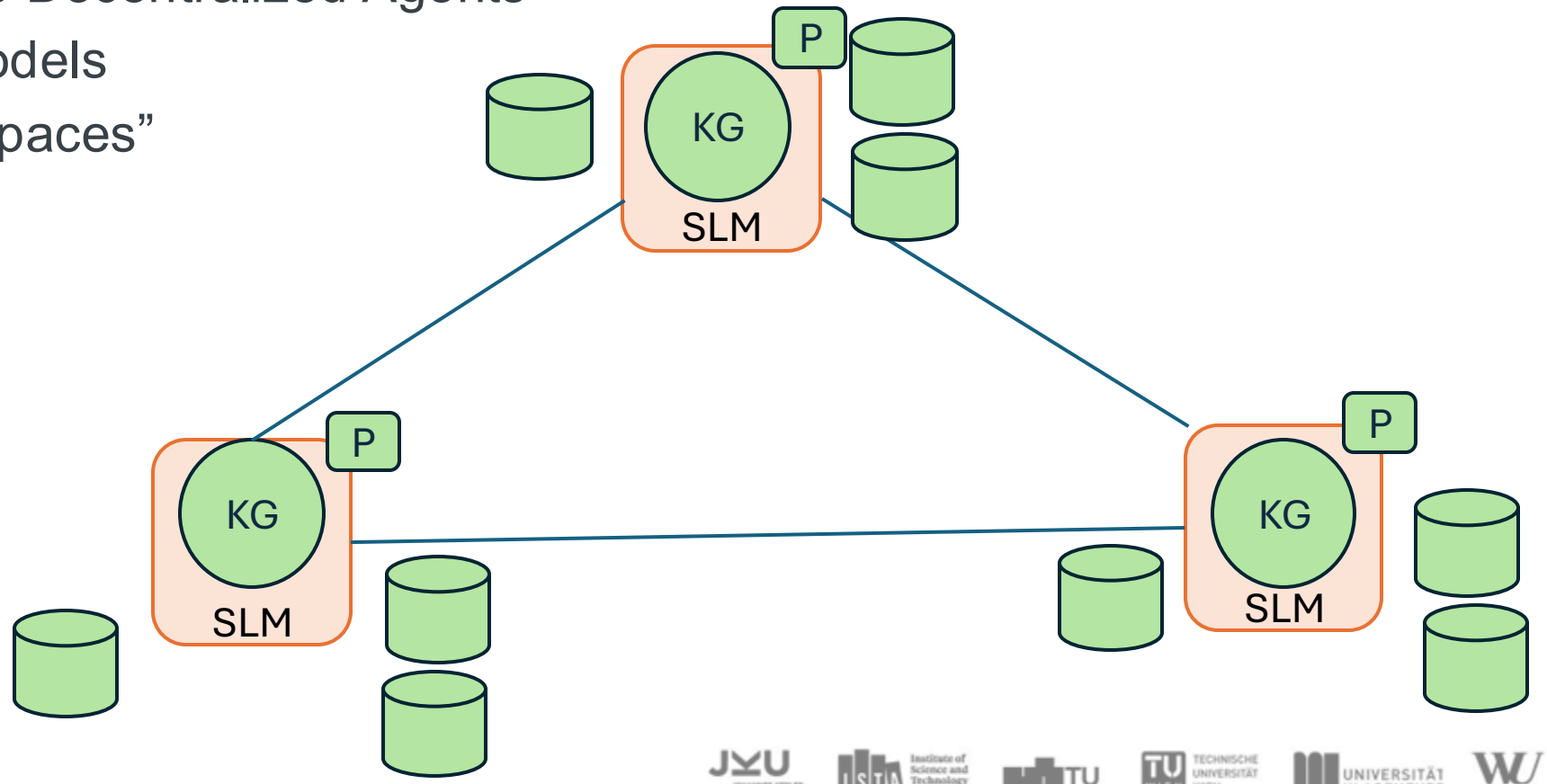
What's next? Where should this go? What is a solution we should jointly work towards?

Neuro-symbolic agents

Trustable Decentralized Agents

Small models

"Agent Spaces"



Funded by:
FWF Austrian Science Fund