

How Does Knowledge Evolve in Open Knowledge Graphs?

Axel Polleres

Vienna University of Economics and Business

wish I had more time to prepare for the talk...
... but I was hit by dynamic events

14:33 73%

← Route details

RJX 165 → Budapest-Keleti

12:28 12:29	Bludenz Bahnhof	Pl. 3
12:30 12:31		
13:01 13:01	St.Anton am Arlberg Bahnhof	Pl. 3
13:03 13:04		
13:25 13:38	Landeck-Zams Bahnhof	Pl. 2
13:27 13:39		
13:46 13:57	Ötztal Bahnhof	Pl. 2
13:48 13:58		
14:11 14:20	Innsbruck Hbf	Pl. 7
14:17 14:23		
14:41 14:45	Wörgl Hbf	Pl. 3
14:43 14:46		
14:51 14:53	Kufstein Bahnhof	Pl. 2
	⊗ Stop cancelled	
16:31 16:34	Bischofshofen Bahnhof	Pl. 3
16:33 16:35		
	⚠ Additional stop	
16:03 17:10	Salzburg Hbf	Pl. 4
16:07 17:14		
	⚠ Attention: Timetable changes are possible	
17:15 18:22	Linz/Donau Hbf	Pl. 7A-F
17:17 18:24		
	⚠ Attention: Timetable changes are possible	



wish I had more time to prepare for the talk...
... but I was hit by dynamic events

Instance data changing in a highly dynamic fashion:

Not only the route, but also temporal information:



Abfahrt		Zug train	Departure	16:41:00	OBB INFRA Bahnsteig platform
Zeit time	Erwartet estimated		Nach to		
14:43	17:03	OBB RJX 165	Budapest-Keleti	über Bischofshofen~Salzburg Hbf~Linz Hbf~ St.Pölten Hbf~Wien Meidling~Wien Hbf~ Zugteilung in Wien Hbf	3
14:43	17:03	OBB RJX 565	Flughafen Wien		3 A-C
15:19	17:14	OBB RJX 766	Innsbruck Hbf	über Bischofshofen~Salzburg Hbf~Linz Hbf~ St.Pölten Hbf~Wien Meidling~Wien Hbf	4 B-D
15:45	17:34	OBB RJX 869	Flughafen Wien	über Bischofshofen~Salzburg Hbf~Linz Hbf~ St.Pölten Hbf~Wien Meidling~Wien Hbf~ Zugteilung in Wien Hbf	3 B-D
16:43	17:40	OBB RJX 167	Bratislava hl.st.		3
16:43	17:40	OBB RJX 567	Flughafen Wien		3 A-C
16:46		OBB EC 89	Verona P. N.	Ausfall zwischen München Hbf und Innsbruck Hbf	
16:50		OBB REX3	Schwarzach-St.Veit	Leogang~Leogang-Steinberge~Saalfelden~ Zell am See~Bruck-Fusch~Taxenbach	1 D

Derzeit kein Zugverkehr zwischen Kufstein und Salzburg Hbf..
Grund: starker Schneefall

wish I had more time to prepare for the talk...
... but I was hit by dynamic events

Instance data changing in a highly dynamic fashion:

Not only the route, but also temporal information:

Zeit time	Erwartet estimated	Zug train	Nach to	platform
14:43	17:04	OBB RJX 165	Budapest-Keleti	3
14:43	17:04	OBB RJX 565	Flughafen Wien	3 A-C
15:19	17:14	OBB RJX 766	Innsbruck Hbf	4 B-D
15:45	17:34	OBB RJX 869	Flughafen Wien	3 B-D
16:43	17:40	OBB RJX 167	Bratislava hl.st.	3
16:43	17:40	OBB RJX 567	Flughafen Wien	3 A-C
16:46		OBB EC 89	Ausfall/Cancelled	
16:50		OBB REX3	Schwarzach-St. Veit	1 D

Information: SCOTTY mobil | Twitterert.
... between Kufstein and Salzburg Hbf



wish I had more time to prepare for the talk...
... but I was hit by dynamic events

Instance data changing in a highly dynamic fashion:

Not only the route, but also temporal information:

Abfahrt		Zug		Departure		16:42:38		ÖBB INFRA	
Zeit	Erwartet		train	Nach	to			Bahnsteig	platform
time	estimated								
14:43	17:05	ÖBB	RJX 165	Budapest-Keleti		Hält nicht in Kufstein.		3	
14:43	17:05	ÖBB	RJX 565	Flughafen Wien		Hält zusätzlich in Bischofshofen.		3 A-C	
15:19	17:14	ÖBB	RJX 766	Innsbruck Hbf		Zugtellung in Wien Hbf		4 B-D	
15:45	17:34	ÖBB	RJX 869	Flughafen Wien		über Bischofshofen~Salzburg Hbf~Linz Hbf~		3 B-D	
16:43	17:40	ÖBB	RJX 167	Bratislava hl.st.		St.Pölten Hbf~Wien Meidling~Wien Hbf		3	
16:43	17:40	ÖBB	RJX 567	Flughafen Wien		Hält nicht in Kufstein.		3 A-C	
16:46		ÖBB	EC 89	Verona P. N.		Hält zusätzlich in Bischofshofen.			
●16:50		ÖBB	REX3	Schwarzach-St. Veit		Zugtellung in Wien Hbf			
						Ausfall zwischen München Hbf und			
						Innsbruck Hbf			
						Kitzbühel Hahnenkamm~Kitzbühel~		1 D	
						St.Johann in Tirol~Fieberbrunn~Hochfilzen~			

Information: SCOTTY mobil | Twitterert.
No Train service between Kufstein and Salzburg Hbf



wish I had more time to prepare for the talk...
... but I was hit by dynamic events

Instance data changing in a highly dynamic fashion:

Not only the route, but also **temporal information**:

Zeit time	Erwartet estimated	Zug train	Nach to	16:45:55	ÖBB INFRA Bahnsteig platform
14:43	17:08	ÖBB RJX 165	Budapest-Keleti		3
14:43	17:08	ÖBB RJX 565	Flughafen Wien		3 A-C
15:19	17:14	ÖBB RJX 766	Innsbruck Hbf		4 B-D
15:45	17:34	ÖBB RJX 869	Flughafen Wien		3 B-D
16:43	17:44	ÖBB RJX 167	Bratislava hl.st.		3
16:43	17:44	ÖBB RJX 567	Flughafen Wien		3 A-C
16:46		ÖBB EC 89	Verona P. N.		
16:50		ÖBB REX3	Schwarzach-St. Veit		1 D



wish I had more time to prepare for the talk...
... but I was hit by dynamic events

Instance data changing in a highly dynamic fashion:

Dynamic constraints violated
(inconsistency via dynamics!):

16:29 Results

A Würgl

B Wien Hbf (U)

Dep now

Due to heavy snowfall no night trains can run today. Passengers affected should contact the train team or the ÖBB customer service at +43 (0)5-1717-9. We apologize for any inconvenience.

Storm warning
Due to the current storm damage in Austria, deviations and delays in train services are to be expected in the coming hours.. We ask you to delay non-urgent travels. As soon as we have any further information, we will pass it on to you.

Departure: Today, Saturday, 02.12.2023

14:43	RJX 165	18:32
16:53	3:49h, 0x Chg.	21:38

Journey suggestion according to current traffic.

! Cancellation of intermediate stops

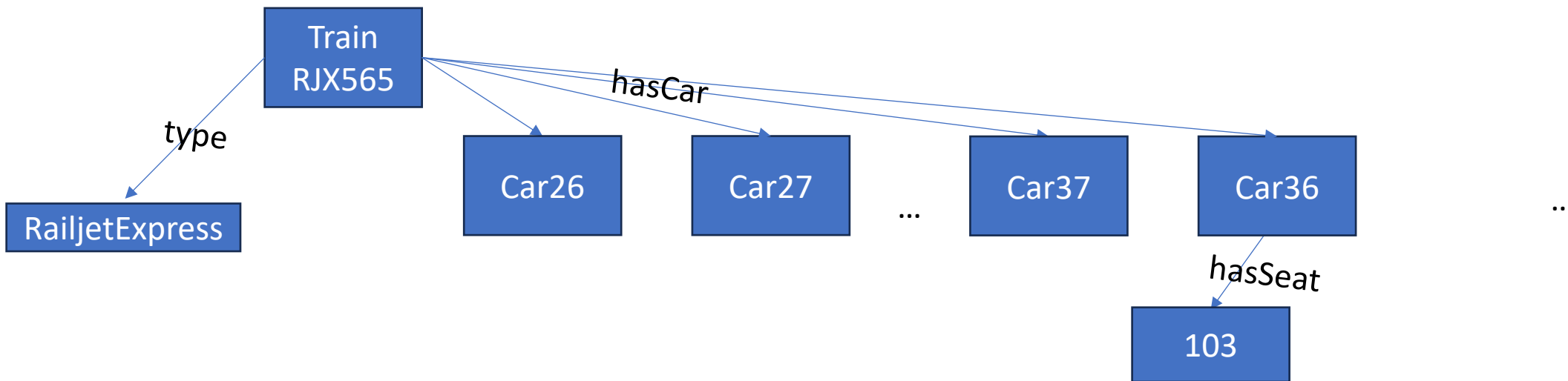
! 3 messages available.

14:43	RJX 565	18:32
16:53	3:49h, 0x Chg.	18:40

Journey suggestion according to current

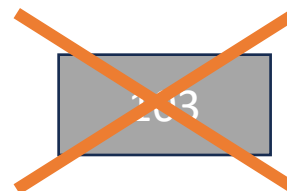
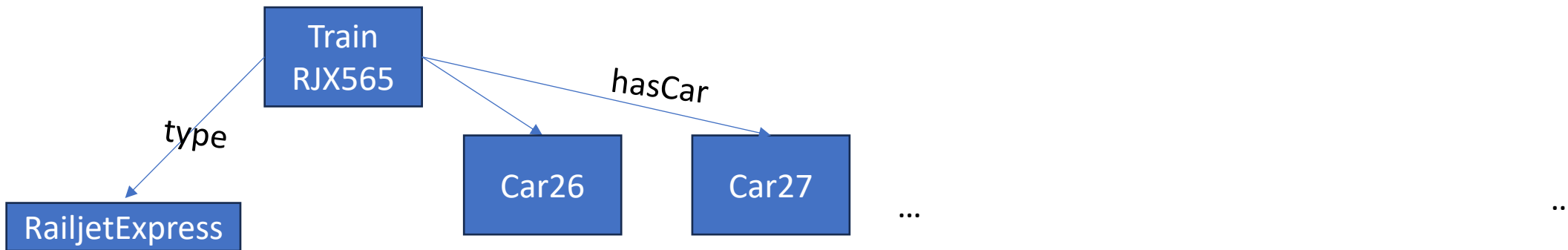
wish I had more time to prepare for the talk...
... but I was hit by dynamic events

Even seemingly static data no longer valid:

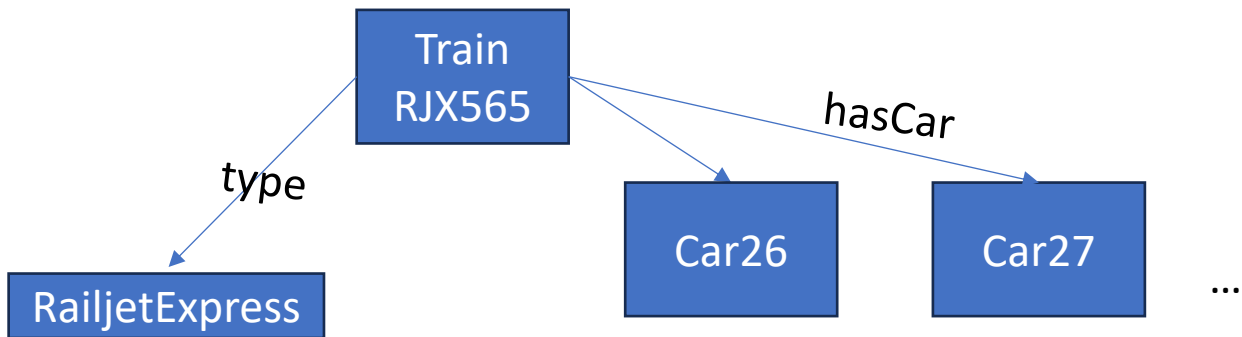


wish I had more time to prepare for the talk...
... but I was hit by dynamic events

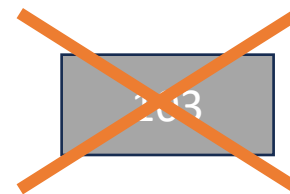
Even seemingly static data no longer valid...



wish I had more time to prepare for the talk...
... but I was hit by dynamic events



... working on the train as planned not possible :-S



wish I had more time to prepare for the talk...
... but I was hit by dynamic events...

...after a 5hrs delay arrival ...

Sa
2
DEZ

Wörgl Hbf > Wien Hbf
Ich

Berücksichtigte Ermäßigungen:
1x KlimaTicket Ö Classic

Willkommen in Wien Hbf

17:13 **RJX** **23:30**
~~14:43~~ ~~18:32~~
Bahnsteig 3

... just to realize the next morning, this:

Please do not reply to this email.

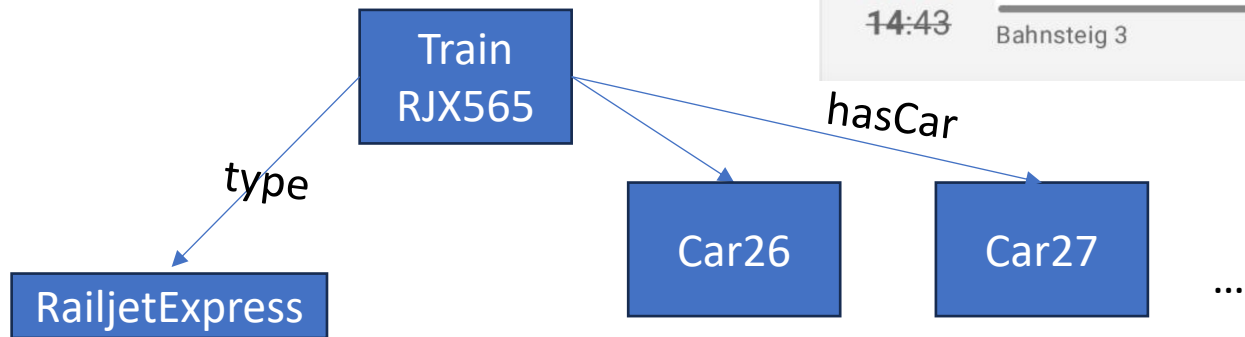
Austrian

Your Austrian flight has been cancelled

Dear Mr. Polleres,

Your Austrian flight OS405 from Vienna to Lyon on 03.12.2023 has been cancelled.

Please accept our sincere apologies.



...which perfectly motivates a talk
about dynamics and evolution in KGs!

Thanks to my co-authors and to the new TGDK journal!



Transactions on Graph Data and Knowledge (TGDK)

GENERAL EDITORIAL BOARD SELECTION/REVIEW AUTHOR INSTRUCTIONS EDITOR INSTRUCTIONS RECENTLY PUBLISHED

Transactions on Graph Data and Knowledge is a Diamond Open Access journal that publishes research contributions relating to the use of graphs for data and knowledge management.

Publications

All documents published in this journal are available open access on DROPS:



Moreover, all papers are indexed in dblp: [TGDK @ dblp](#)

Aims and Scope

Transactions on Graph Data and Knowledge (TGDK) is an Open Access journal that publishes original research articles and survey articles on graph-based abstractions for data and knowledge, and the techniques that such abstractions enable with respect to integration, querying, reasoning and learning. The scope of the journal thus intersects with areas such as Graph Algorithms, Graph Databases, Graph Representation Learning, Knowledge Graphs, Knowledge Representation, Linked Data and the Semantic Web. Also in-scope for the journal is research investigating graph-based abstractions of data and knowledge in the context of Data Integration, Data Science, Information Extraction, Information Retrieval, Machine Learning, Natural Language Processing, and the Web.

The journal is Open Access without fees for readers nor for authors (also known as *Diamond Open Access*).

Open Access Policy

TGDK articles are peer-reviewed and published according to the principle of OpenAccess, i.e., they are available online and free of charge. The authors retain their copyright.

QUICK LINKS

How Does Knowledge Evolve in Open Knowledge Graphs?

Axel Polleres ✉ 🏠

Vienna University of Economics and Business, Austria
Complexity Science Hub Vienna, Austria

Romana Pernisch¹ ✉ 🏠

Vrije Universiteit Amsterdam, the Netherlands
Discovery Lab, Elsevier, the Netherlands

Angela Bonifati ✉ 🏠

Lyon 1 University, CNRS LIRIS, France
IUF, France

Daniele Dell'Aglio ✉ 🏠

Aalborg University, Denmark

Daniil Dobriy ✉ 🏠

Vienna University of Economics and Business, Austria

Stefania Dumbrava ✉ 🏠

ENSIIE, France
SAMOVAR, IP Paris, France

Lorena Etcheverry ✉

Universidad de la República, Uruguay

Nicolas Ferranti ✉

Vienna University of Economics and Business, Austria

Katja Hose ✉ 🏠

TU Wien, Austria

Ernesto Jiménez-Ruiz ✉ 🏠

City, University of London, UK
SIRIUS, University of Oslo, Norway

Matteo Lissandrini ✉

Aalborg University, Denmark

Ansgar Scherp ✉ 🏠

Ulm University, Germany

Riccardo Tommasini ✉ 🏠

INSA Lyon, CNRS LIRIS, France

Johannes Wachs ✉

Corvinus University of Budapest, Hungary
Centre for Economic and Regional Studies, Hungary
Complexity Science Hub Vienna, Austria

¹ Corresponding author

What did I want to talk about?

- Open Knowledge Graphs?
- Perspectives on evolution: dimensions of time and temporality
- Observability of evolution
- Metrics for evolution
- How does evolution affect downstream tasks and resp. techniques?
- Recent own work...

How Does Knowledge Evolve in Open Knowledge Graphs?

Axel Polleres ✉ 🌐
Vienna University of Economics and Business, Austria
Complexity Science Hub Vienna, Austria

Romana Pernisch¹ ✉ 🌐
Vrije Universiteit Amsterdam, the Netherlands
Discovery Lab, Elsevier, the Netherlands

Angela Bonifati ✉ 🌐
Lyon 1 University, CNRS LIRIS, France
IUF, France

Daniele Dell'Aglio ✉ 🌐
Aalborg University, Denmark

Daniil Dobriy ✉ 🌐
Vienna University of Economics and Business, Austria

Stefania Dumbrava ✉ 🌐
ENSIE, France
SAMOVAR, IP Paris, France

Lorena Etcheverry ✉ 🌐
Universidad de la República, Uruguay

Nicolas Ferranti ✉ 🌐
Vienna University of Economics and Business, Austria

Katja Hose ✉ 🌐
TU Wien, Austria

Ernesto Jiménez-Ruiz ✉ 🌐
City, University of London, UK
SIRIUS, University of Oslo, Norway

Matteo Lissandrini ✉ 🌐
Aalborg University, Denmark

Ansgar Scherp ✉ 🌐
Ulm University, Germany

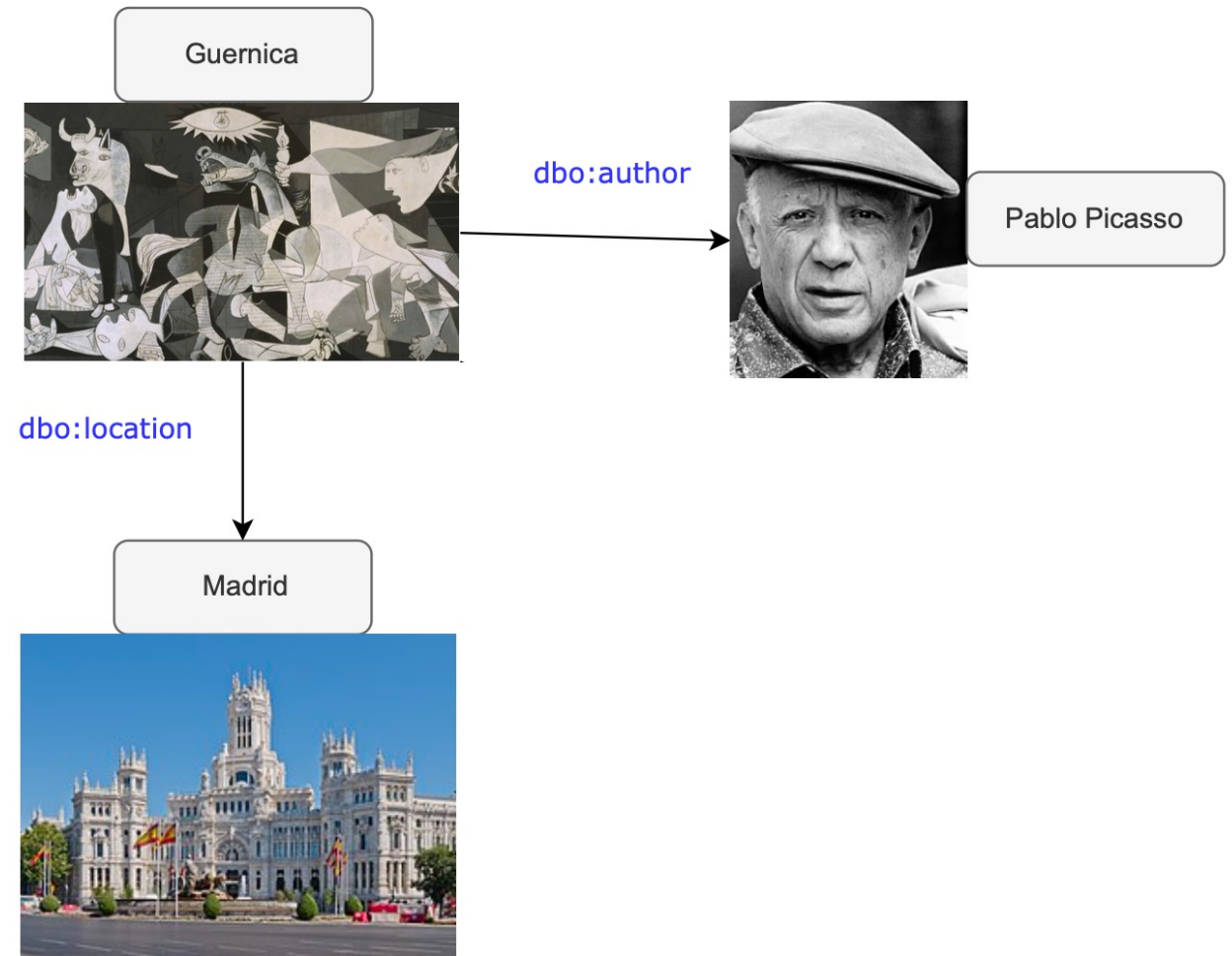
Riccardo Tommasini ✉ 🌐
INSA Lyon, CNRS LIRIS, France

Johannes Wachs ✉ 🌐
Corvinus University of Budapest, Hungary
Centre for Economic and Regional Studies, Hungary
Complexity Science Hub Vienna, Austria

¹ Corresponding author

Open, collaboratively edited KGs

- E.g.
 - DBPedia
 - Wikidata
- Maintain collaboratively edited, curated, reusable knowledge
- Serve as a backbone for various applications!



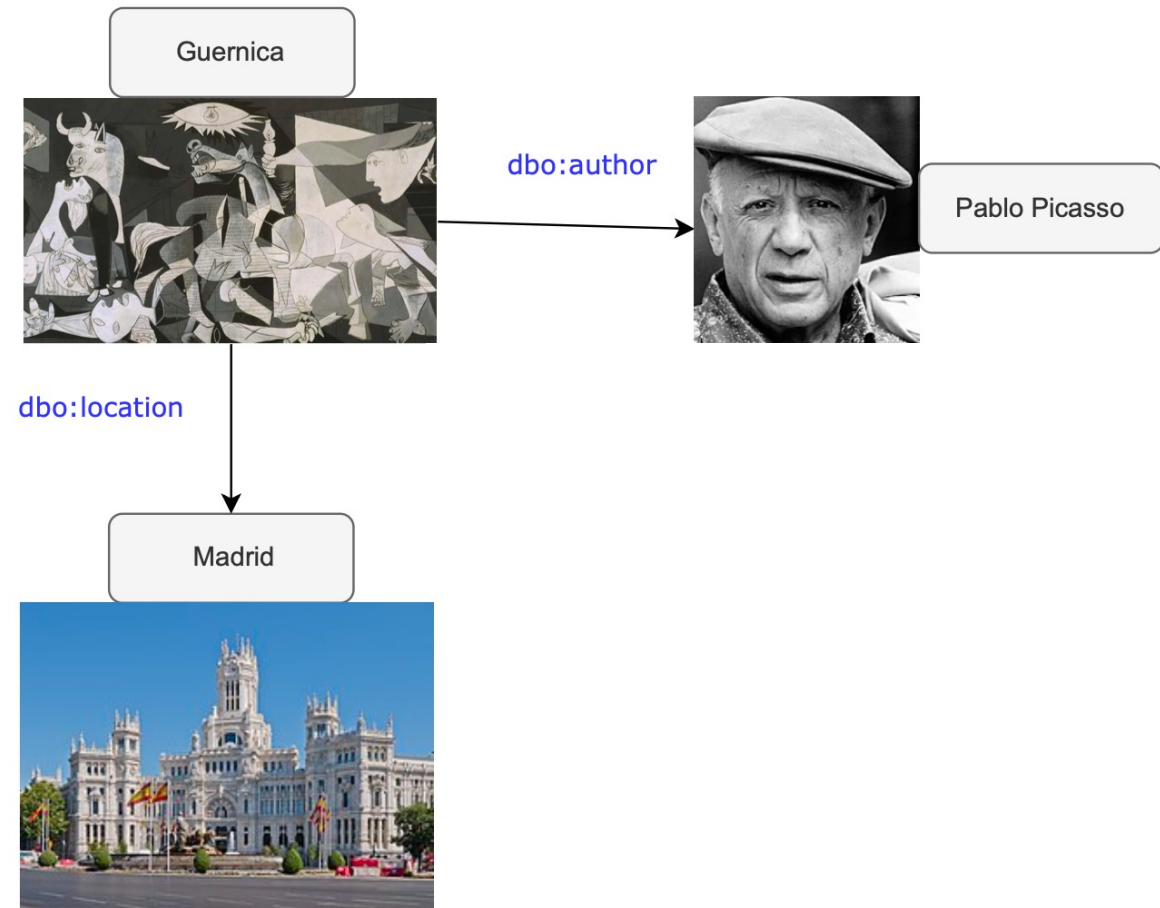
Main questions:

- *Which publicly accessible, open KGs are observable in a manner that would allow a longitudinal analysis of their evolution and how?*
 - *What dimensions does evolution have at all?*
- *Do we have the right metrics to analyse KGs' evolution?*
- *Do we have the right techniques to process evolving KGs?*

What dimensions does evolution have at all?

- Temporal KGs: Time as Data (“valid time”)

- Dates/Timestamps

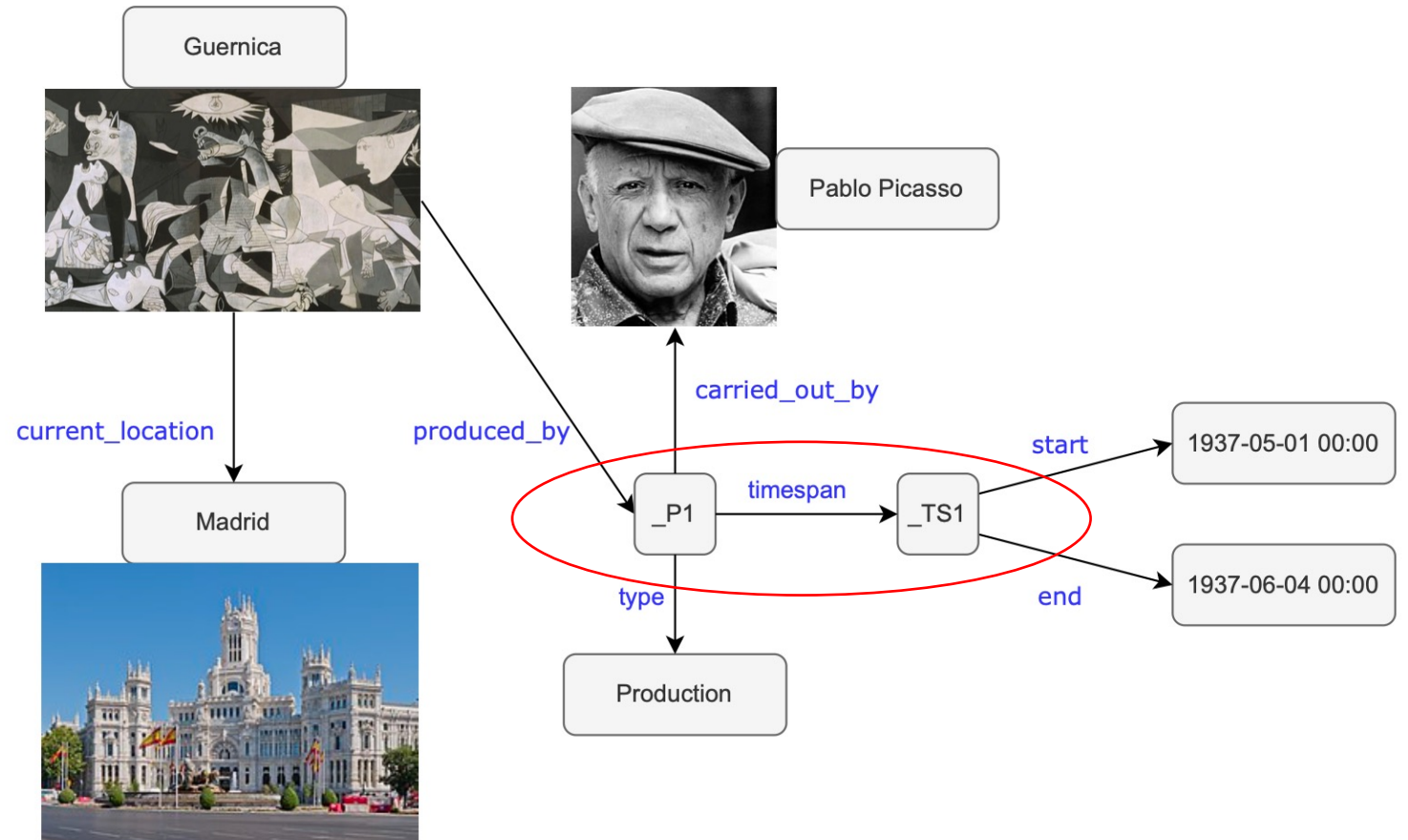


What dimensions does evolution have at all?

- Temporal KGs: Time as Data (“valid time”)

- Dates/Timestamps
- Intervals
- (start/end events)

Challenge: Needs reification!



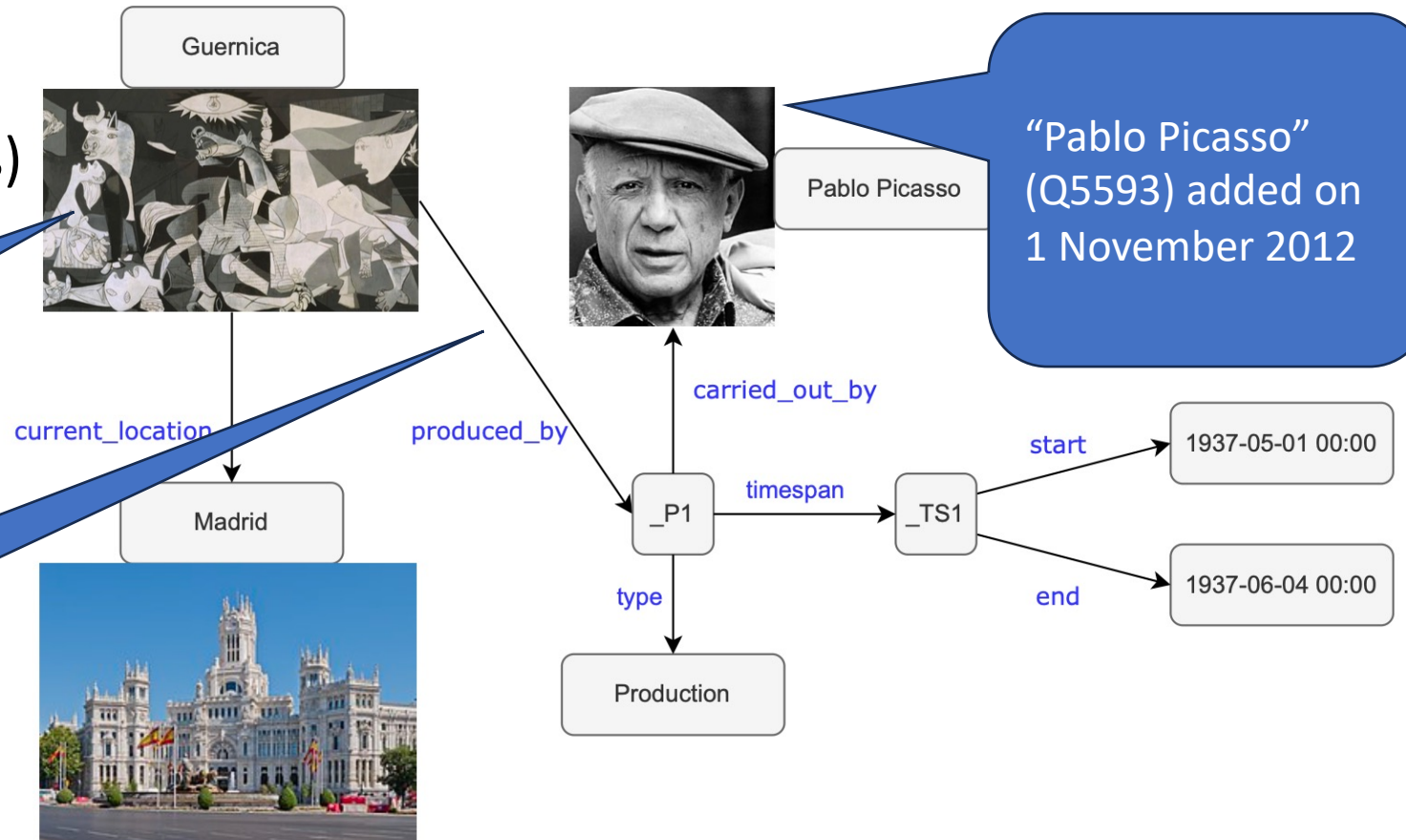
What dimensions does evolution have at all?

- **Time-varying KGs:** Time as meta-data/log data (“transaction time”)

- Edit events(Dates/Timestamps)

“Guernica”
(Q175036) was
created on 28
November 2012 in
Wikidata

The statement that Picasso
created Guernica was only
created in Wikidata in March
2013



What dimensions does evolution have at all?

- Time-varying KGs: Time as meta-data/log data (“transaction time”)

Revision history of "Guernica" (Q175036)

[? Help](#)

[View logs for this item \(view abuse log\)](#)

- Edit events(Dates/Timestamps)

Filter revisions

Diff selection: Mark the radio buttons of the revisions to compare and hit enter or the button at the bottom.

Legend: **(cur)** = difference with latest revision, **(prev)** = difference with preceding revision, **m** = minor edit.

(latest | earliest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

- (cur | prev) 16:54, 15 October 2014 Dexbot (talk | contribs) .. (19,851 bytes) (+66) .. (Updated Item: Bot: setting proper label for sh)
- (cur | prev) 11:00, 6 October 2014 Yoosef Pooranvary (talk | contribs) .. (19,785 bytes) (+13) .. (Page moved from [fawiki:گرنیکا] to [fawiki:گرنیکا (نقاشی)])
- (cur | prev) 01:52, 30 September 2014 Dexbot (talk | contribs) .. (19,772 bytes) (-66) .. (Removed Persian alias: گرنیکا)
- (cur | prev) 23:34, 17 September 2014 OC Ripper (talk | contribs) .. (19,838 bytes) (+8,775) .. (Added link to [shwiki]: Guernica (Picasso))
- (cur | prev) 16:35, 7 August 2014 Coyau (talk | contribs) .. (11,063 bytes) (+253) .. (Created claim: genre (P136): history painting (Q742333))
- (cur | prev) 09:11, 20 May 2014 Dexbot (talk | contribs) .. (10,810 bytes) (+25) .. (Updated Item: Bot: setting proper label for war)
- (cur | prev) 03:00, 19 May 2014 AkkakkBot (talk | contribs) .. (10,785 bytes) (+38) .. (Updated Item: - set pa label to गुरनीका (task 5))
- (cur | prev) 23:56, 11 May 2014 AkkakkBot (talk | contribs) .. (10,747 bytes) (-28) .. (Updated Item: - remove redundant simple label (task 7))

- Metadata also contains

additional provenance information

- ... yet, often not part of the RDF/Graph model itself!

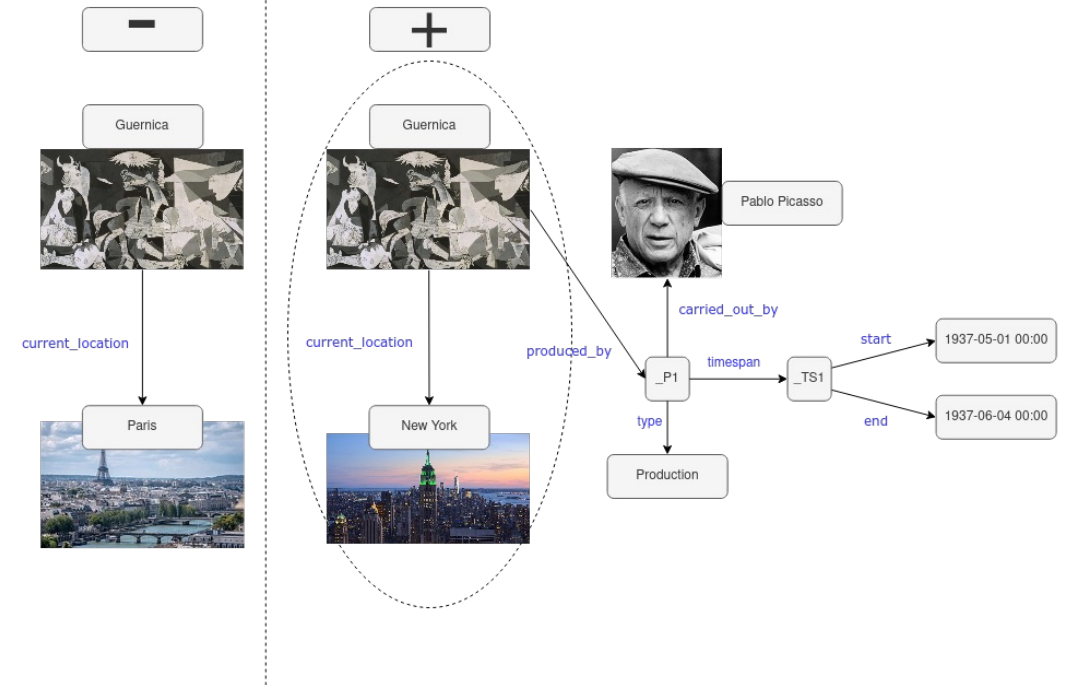
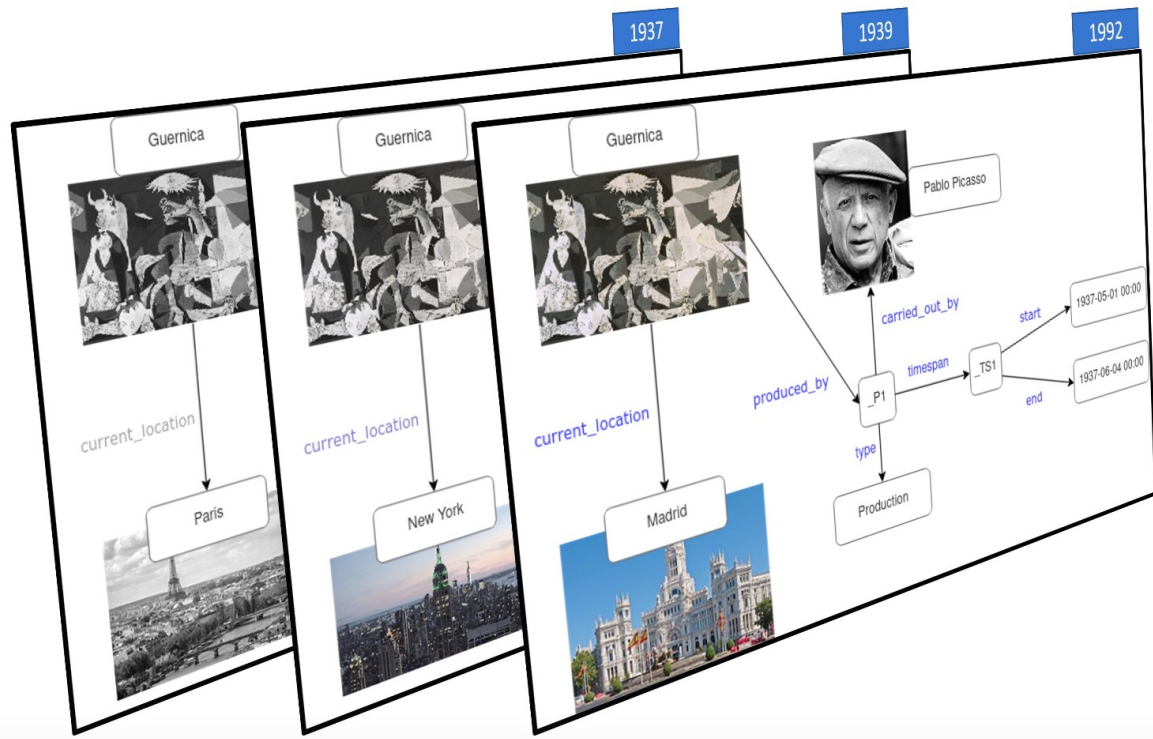
What dimensions does evolution have at all?

- Time-varying KGs: Granularity of dynamicity (observability):

- “Versioned KG”
 (“snapshot”) of particular materialisations

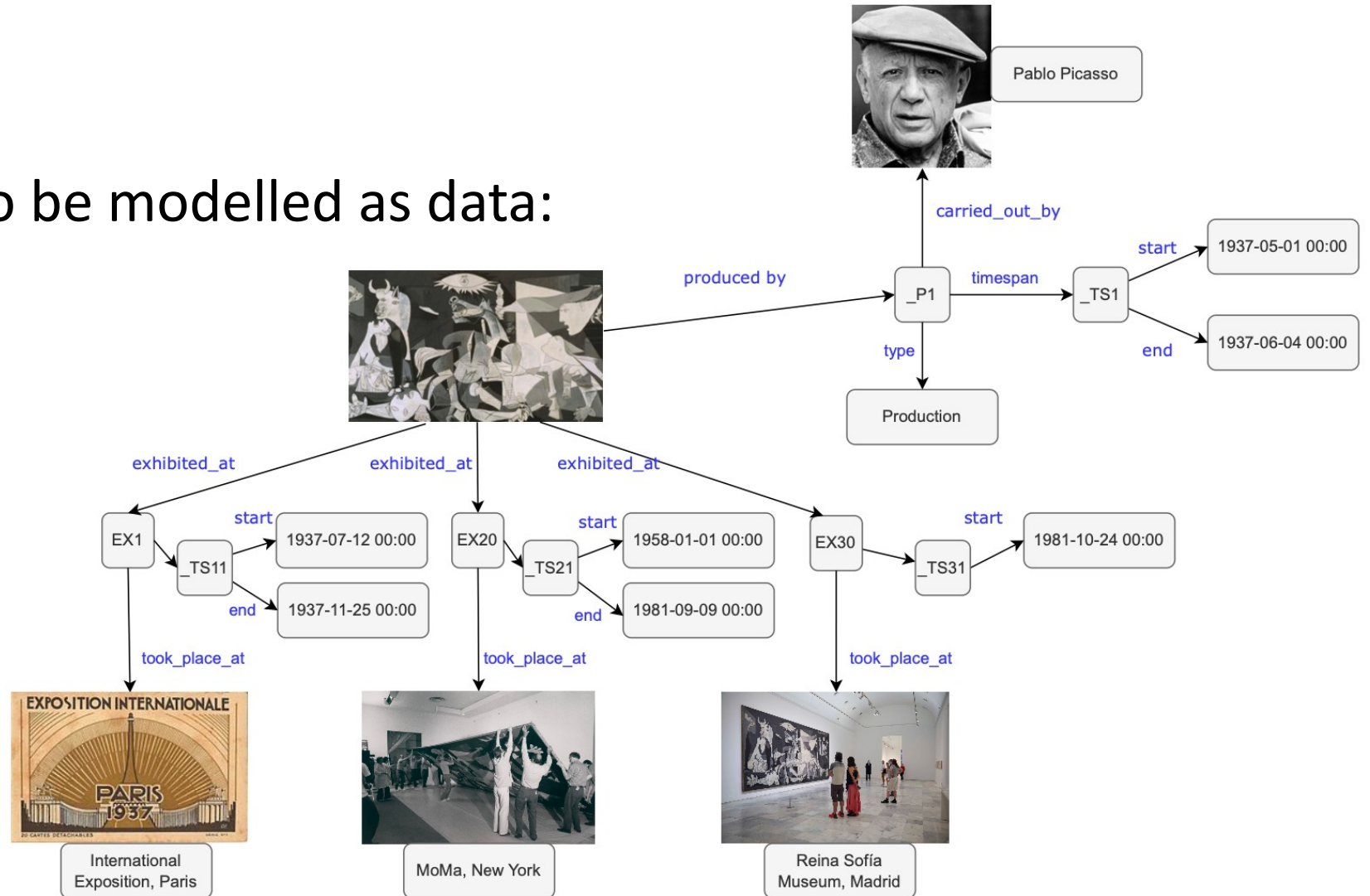
vs.

- “Dynamic KG”
 down to single Edit events

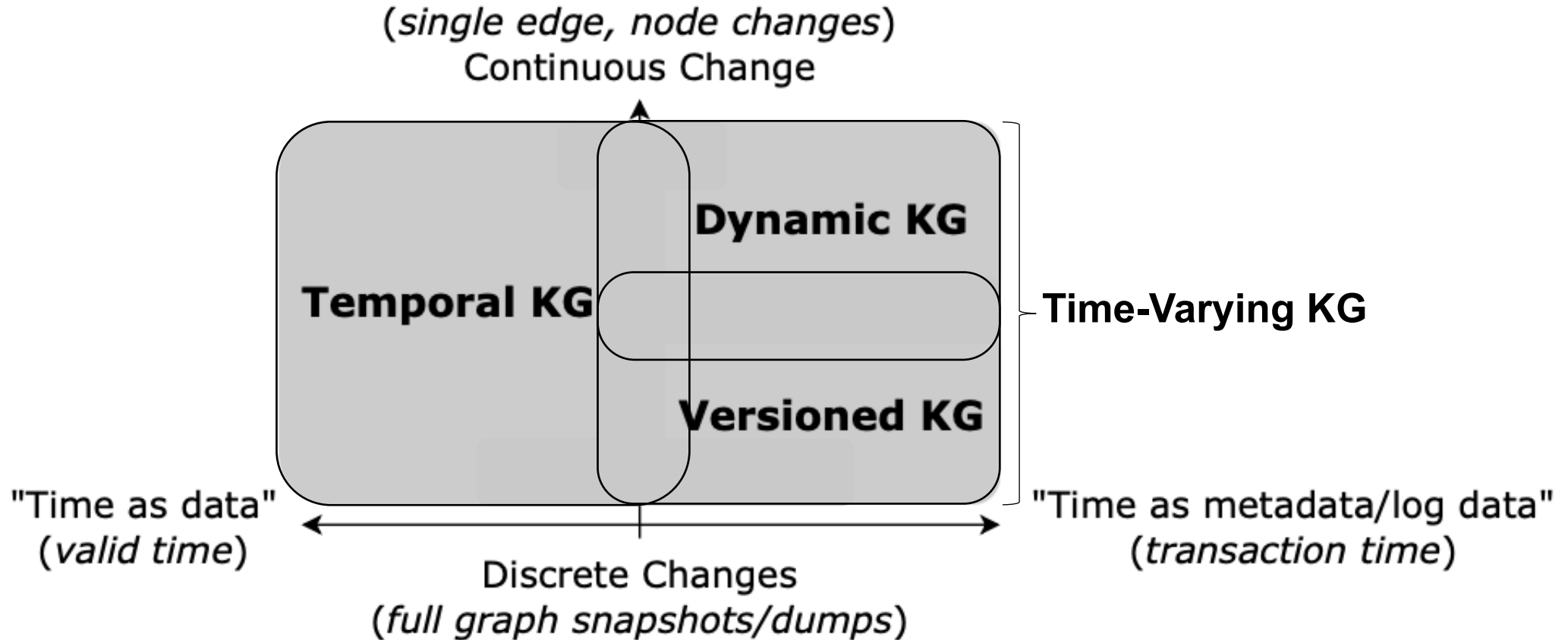


What dimensions does evolution have at all?

- Lines are blurry...
- location could also be modelled as data:



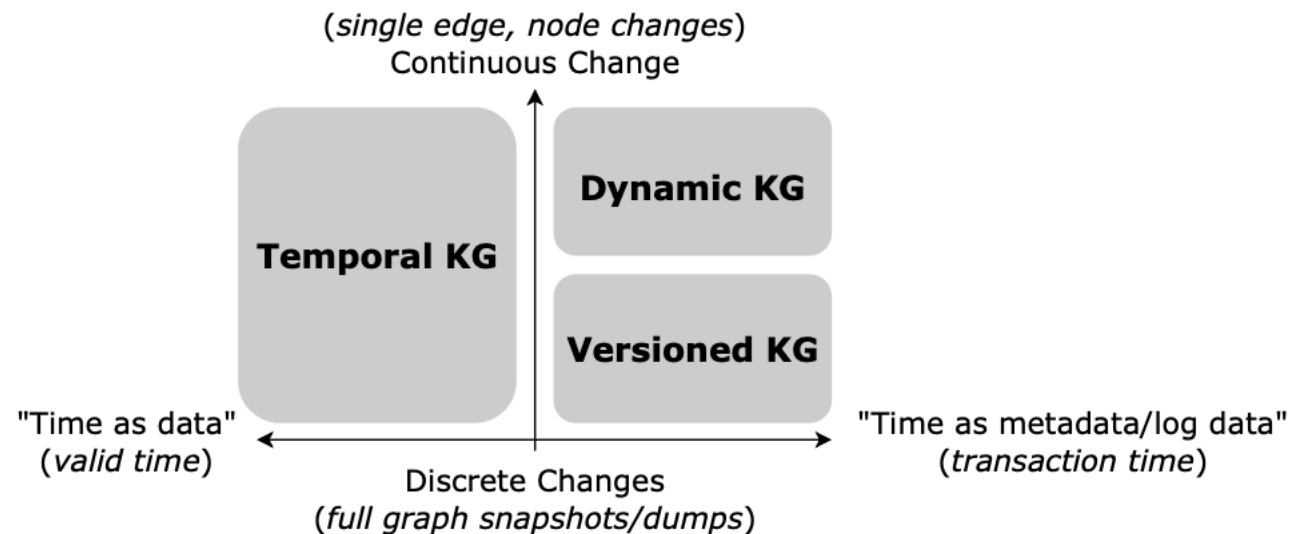
What dimensions does evolution have at all?



additional

What dimensions does evolution have at all?

- Instance evolution
- Schema evolution
- Collaboration evolution
- Structural evolution
- Dynamics (change frequencies, etc.)
- Timeliness (recency of temporal information, delays)
- Monotonicity (“growth” vs. “deletions”)



Availability of Dbpedia and Wikidata:

	Level	Queryable	Collaborative	Formats	Protocol	Metadata	Temporality	Timeliness
Wikidata	V	Yes	Yes	NT, TTL, HDT, JSON	HTTP, SPARQL	schema.org	No	2-3 Days
	S	Yes	Yes	NT, TTL, JSON	HTTP, SPARQL	schema.org	No	2-3 Days
	CL	Yes	Yes	JSON	SSE	No	Event TS	Seconds
DBpedia	V	Yes	Partial	NT	HTTP, SPARQL	No	No	Quarterly
	S	Yes	Yes	RDF	HTTP, SPARQL	No	No	Daily
	CL	Yes	Yes	RDF	HTTP	No	Graph TS	Daily

Availability of Open KG Versions (V), Schema (S), and Change logs (CL), find more in the paper!

Note: additional Caveat:

- Historic versions are hard to maintain and host:

<https://www.rdfhdt.org/datasets/>

Important Note (12 April 2022): We are experiencing some **technical problems** on our “gaia” server, so unfortunately **some datasets could be unavailable** (e.g. Wikidata). We hope to resolve this issue soon, thanks for your understanding!

Dataset	Size	Triples	Details	Provenance
Latest Wikidata (3rd march 2021) NEW	53GB uncomp		<h2>What is the “DBpedia Snapshot” Release?</h2> <p>Historically, this release has been associated with many names: “DBpedia Core”, “EN DBpedia”, and – most confusingly – just “DBpedia”. In fact, it is a combination of –</p>	
Latest Wikidata (9th march 2020)	50GB uncomp		<ul style="list-style-type: none">• EN Wikipedia data – A small, but very useful, subset (~ 1 Billion triples or 14%) of the <u>whole DBpedia extraction using the DBpedia Information Extraction Framework (DIEF)</u>, comprising structured information extracted from the English Wikipedia plus some enrichments from other Wikipedia language editions, notably multilingual abstracts in ar, ca, cs, de, el, eo, es, eu, fr, ga, id, it, ja, ko, nl, pl, pt, sv, uk, ru, zh .	
DBPedia 2016-10 English			<ul style="list-style-type: none">• Links – 62 million community-contributed cross-references and owl:sameAs links to other linked data sets on the Linked Open Data (LOD) Cloud that allow to effectively find and retrieve further information from the largest, decentral, change-sensitive knowledge graph on earth that has formed around DBpedia since 2007.	

<https://www.dbpedia.org/resources/snapshot-release/>

Another approach (for dbpedia):

- Dbpedia “Wayback machine”

<https://wayback.cluster.ai.wu.ac.at/>



The screenshot shows the DBpedia Wayback Machine interface. At the top, there is a logo for DBpedia Wayback Machine. Below the logo, a red banner displays the URL <http://dbpedia.org/resource/Lyon> at 2013-12-04T07:33:00Z. Below the banner, a white bar indicates the raw data format: N-Quads N-Triples N3/Turtle JSON XML Trig. The main content is a table with four columns: Subject, Predicate, Object, and Revision. The table contains seven rows of RDF triples, each with a corresponding revision URL.

Subject	Predicate	Object	Revision
http://en.dbpedia.org/resource/Lyon	http://en.dbpedia.org/property/aprHighC	"16.3"^^<http://www.w3.org/2001/XMLSchema#double>	http://data.wu.ac.at/wayback/dbpedia/Lyon/revision/id/584036615
http://en.dbpedia.org/resource/Lyon	http://en.dbpedia.org/property/aprLowC	"6.5"^^<http://www.w3.org/2001/XMLSchema#double>	http://data.wu.ac.at/wayback/dbpedia/Lyon/revision/id/584036615
http://en.dbpedia.org/resource/Lyon	http://en.dbpedia.org/property/aprPercentsun	"47"^^<http://www.w3.org/2001/XMLSchema#integer>	http://data.wu.ac.at/wayback/dbpedia/Lyon/revision/id/584036615
http://en.dbpedia.org/resource/Lyon	http://en.dbpedia.org/property/aprPrecipitationDays	"9"^^<http://www.w3.org/2001/XMLSchema#integer>	http://data.wu.ac.at/wayback/dbpedia/Lyon/revision/id/584036615
http://en.dbpedia.org/resource/Lyon	http://en.dbpedia.org/property/aprPrecipitationMm	"74.9"^^<http://www.w3.org/2001/XMLSchema#double>	http://data.wu.ac.at/wayback/dbpedia/Lyon/revision/id/584036615
http://en.dbpedia.org/resource/Lyon	http://en.dbpedia.org/property/aprRecordHighC	"30.1"^^<http://www.w3.org/2001/XMLSchema#double>	http://data.wu.ac.at/wayback/dbpedia/Lyon/revision/id/584036615
http://en.dbpedia.org/resource/Lyon	http://en.dbpedia.org/property/aprRecordLowC	"-4.4"^^<http://www.w3.org/2001/XMLSchema#double>	http://data.wu.ac.at/wayback/dbpedia/Lyon/revision/id/584036615

The DBpedia Wayback Machine * SEMANTiCS2015

Javier D. Fernández
Vienna University of
Economics and Business,
Vienna, Austria
javier.fernandez@wu.ac.at

Patrik Schneider
Vienna University of
Economics and Business,
Vienna, Austria

patrik.schneider@wu.ac.at

Jürgen Umbrich
Vienna University of
Economics and Business,
Vienna, Austria

juergen.umbrich@wu.ac.at


Challenge: keep these services running...

Another Open Question:

- How representative are big Open Collaborative KGs like Dbpedia & Wikidata at all?

... What about Enterprise KGs?


Challenge: not accessible 😞



Bachelor Thesis

SMW Cloud: A Corpus of Domain-Specific Knowledge Graphs from Semantic MediaWikis

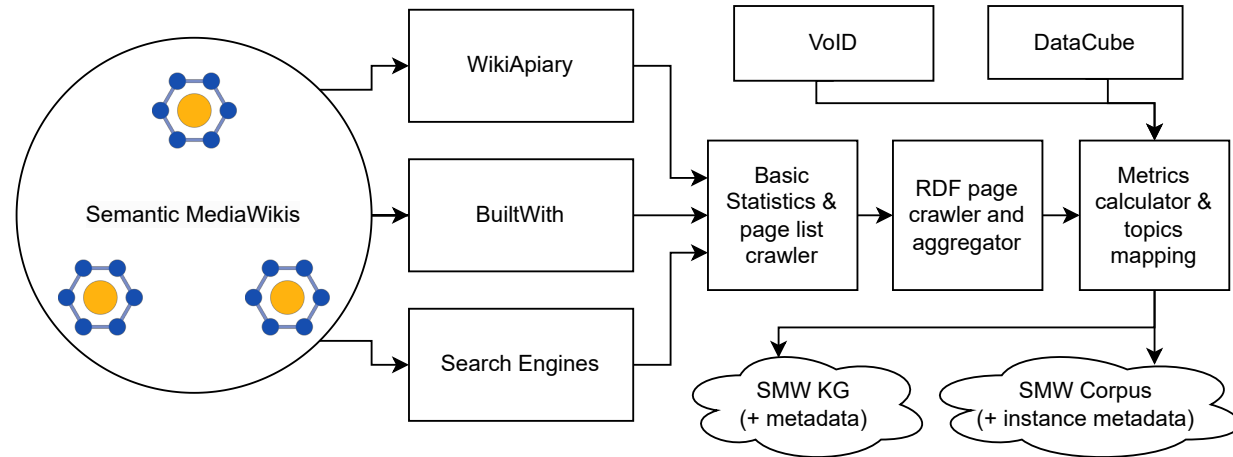
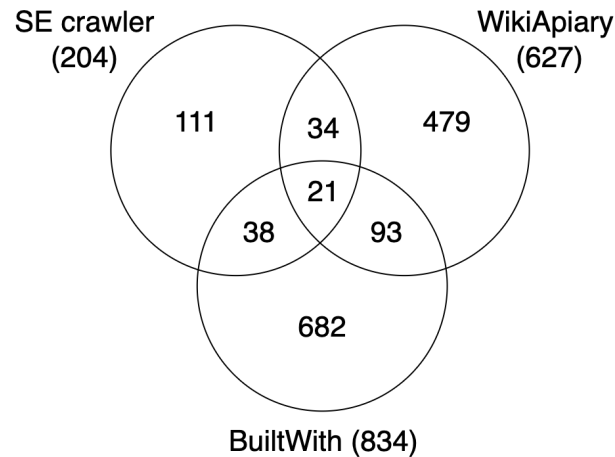
Daniil Dobriy





How many Semantic MediaWikis?

SMW Cloud (1458 wikis)



Dataset	#Triples	#Subjects	#Predicates	#Objects	#Literals
LODStats [10]	192,230,648	Not reported	49,916	Not reported	90,261,655
SMW Cloud	236,505,705	24,010,566	52,670	66,052,823	160,108,216
Wikidata 2021 [23]	17,662,800,665	1,625,057,179	38,867	Not reported	Not reported
LOD-a-lot [13]	28,362,198,927	3,214,347,198	1,168,932	3,178,409,386	1,302,285,394

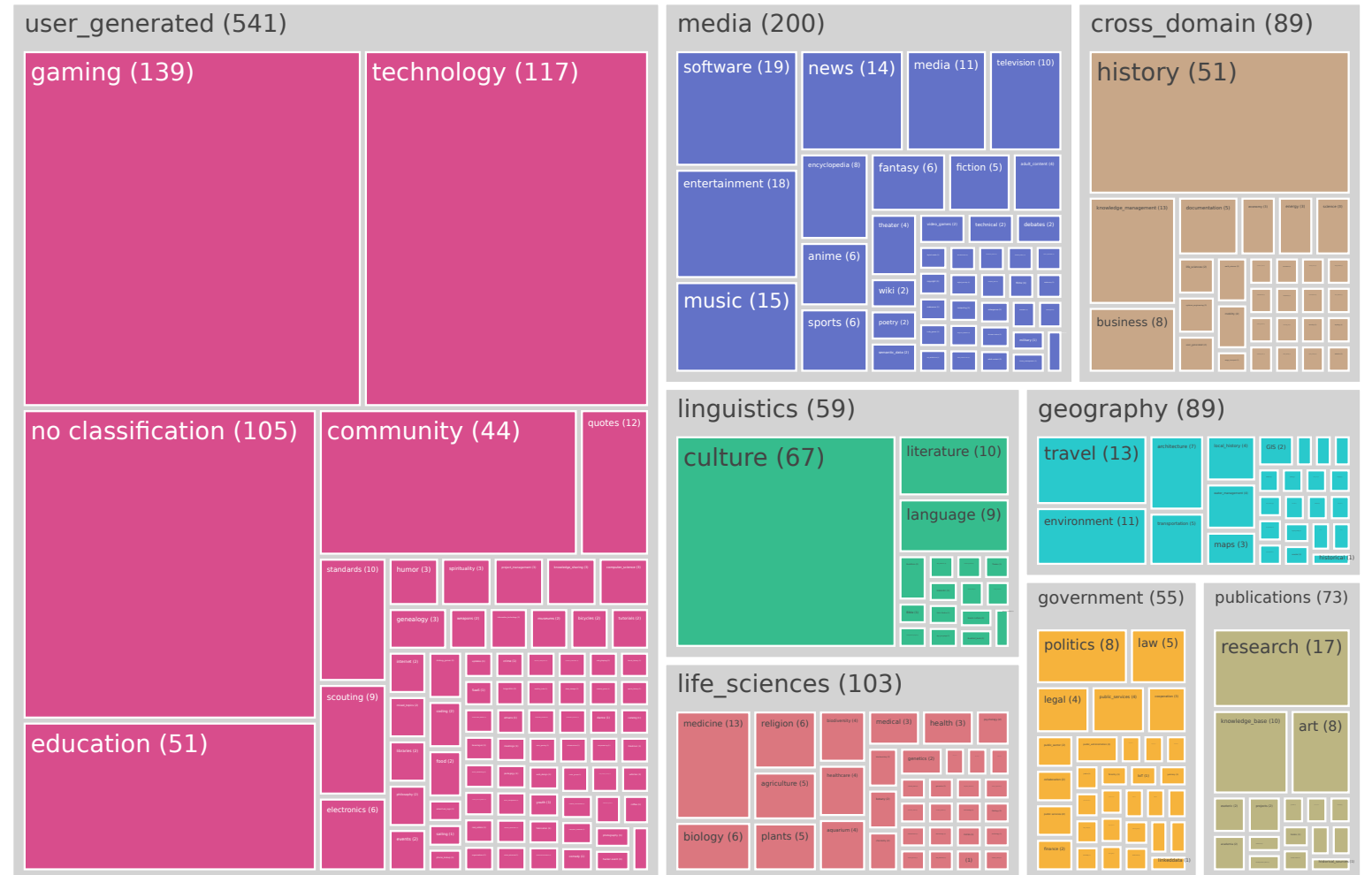
Crawled RDF data available at semantic-data.cluster.ai.wu.ac.at/smwcloud/

Currently ongoing work/next steps:

- also crawl historic data (Semantic MediaWiki edit history)
- also crawl Wikiba.se instances!

SMW Cloud:

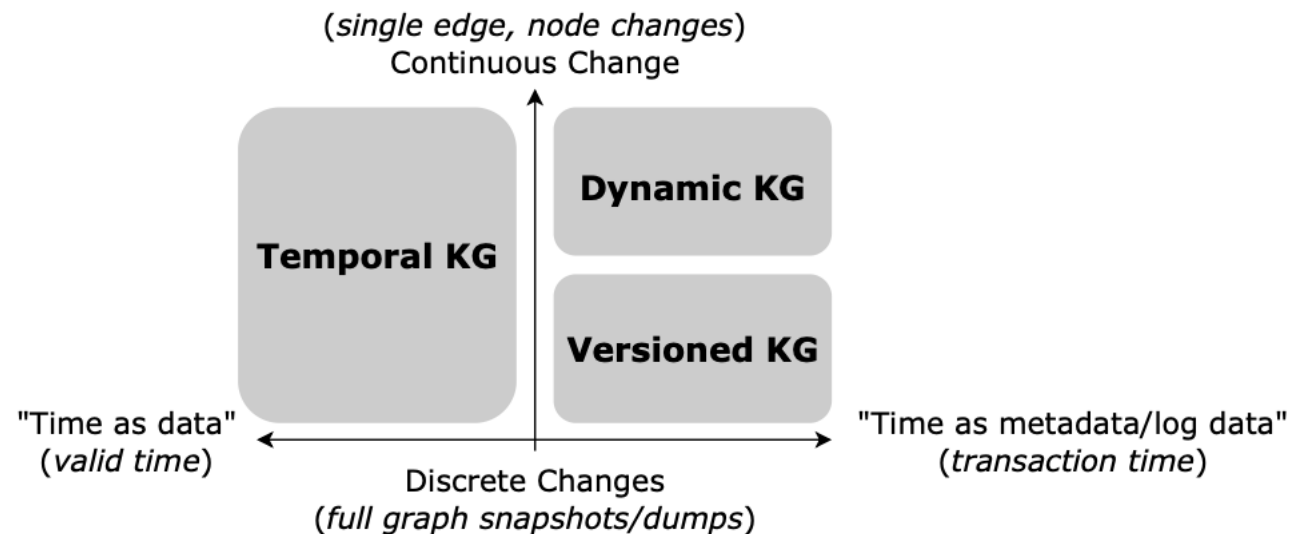
- Very different KGs than Dbpedia and Wikidata
- Small, narrow scheme
- ...very likely, very different evolution!



additional

What dimensions does evolution have at all?

- Instance evolution
- Schema evolution
- **Collaboration evolution**
- Structural evolution
- Dynamics (change frequencies, etc.)
- Timeliness (recency of temporal information, delays)
- Monotonicity (“growth” vs. “deletions”)



Underlying Collaborative KG-Creation Processes

KG	Expert-driven	Crowd-sourced	Resource-dependent	Community-driven	Bot-assisted
Wikidata [244]		✓		✓	✓
DBpedia [148]	✓		✓		(✓)

- **Question:** How can we observe and analyse KG collaboration models?

Following:

Piscopo and Simperl. Who Models the World? Collaborative Ontology Creation and User Roles in Wikidata.

Possible further directions to analyse collaboration in more detail:

- **Question:** *How does the schema evolve in relation to the data?*
- **Question:** *How is the use of the schema related to specific user communities?*

Possible further directions to analyse collaboration in more detail:



Extract aggregated log for analysis

Users, vocabulary (classes/properties), entities:

index,	username,	userID,	vocabulary,	entity
1,	"David",	1234,	P31:Q128207,	Q2
2,	"David",	1234,	P2067,	Q2
3,	"David",	u1234,	P138,	Q2
4,	"David",	u1234,	P2579,	Q3
...				

Evolution of Community-Driven
Schemas within Wikidata

Maria Carla [0000-0001-8004-0464], Sofia Baroncini¹[0000-0002-5636-8328], Margherita
[0000-0002-1690-4093], Mario Scrocca³[0000-0002-8235-7331], Zuzanna
and Axel Polleres⁵[0000-0001-5670-1146]

Master Thesis

- *Challenge:*

- *Again: User/collaboration **data not "readily available"***

**Towards analysing the evolution
of community-driven (sub-)schemas
within Wikidata**

Nicola Pascal Krenn

http://polleres.net/supervised_theses/Nicola_Krenn_MSc_2023.pdf

From static metrics to dynamic metrics

- Basic (static) **Graph metrics**, e.g.

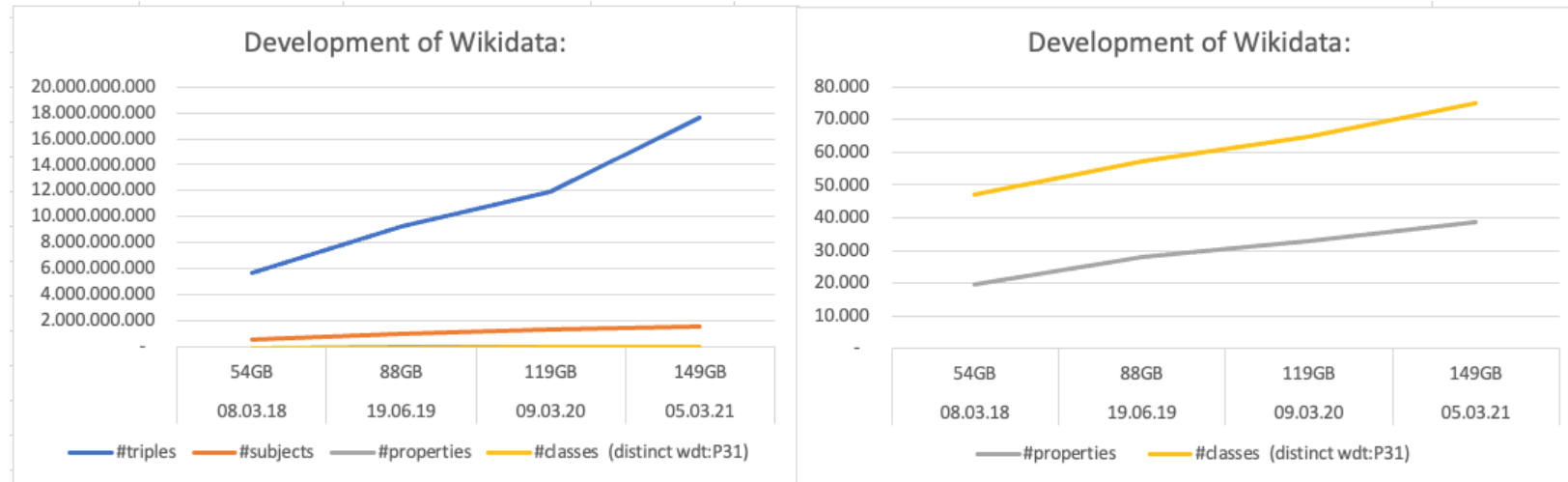
	Metric	Description	Used/Defined in
Graph	Absolute depth	$d_a =$ sum over the cardinality of each path in a set of paths in graph	IsA graph [11, 91, 143, 252]
	Average depth	$d_a / paths $	IsA graph [11, 91, 143, 252], graph [37, 71], OWL schema [73, 71]
	Maximal depth	longest path	IsA graph [11, 91, 143], graph [37]
	Number of paths	$ paths $	DAG [143, 252, 254]
	Tengledness	$\frac{n_G}{t}$, $n_G =$ cardinality of G, $t =$ cardinality of the set of nodes with more than one ingoing IsA arc in G	IsA graph [11, 91, 143]
	Degree Distribution	mean-square deviation of the degree of graph nodes	graph [37, 67, 143]

From static metrics to dynamic metrics

- Basic (static) *Knowledge Graph metrics*, e.g.

Knowledge Graph	Primitives	Entities	number of entities, classes and instances	graph [37, 143], IsA graph [91], OWL [215], DAG [252]
		Properties	number of unique properties or relations	OWL schema [174], OWL [229, 231, 233], DAG [254]
		Classes	$ C $ = number of classes (concepts)	OWL [229, 233], DAG [252, 254]
		Instances	$ I $ = number of individuals	OWL [229, 231, 233]
		Object properties	P_o = number of object properties (non-inheritance)	Schema [143], OWL [215, 231]
Knowledge Graph	T-Box/Schema	Depth of Inheritance Tree		Tree [174], OWL [73, 207, 229], DAG [252]
		Property Class Ratio	$\frac{ P }{ C }$	OWL [231, 174, 73], DAG [252]
		Inheritance Richness	$\frac{ H }{ C }$, H = inheritance relations	OWL [71, 73, 207, 229], Schema [143]
		Attribute Richness	$\frac{ P_d }{ C }$, P_d = datatype properties	OWL [71, 229], Schema [143]
		Class Property Ratio	$\frac{ C }{ P }$	Onto [11, 91, 143]
A-Box/Data		Average Population	$\frac{ I }{ C }$	OWL [73, 229], Onto [91]
		Cohesion	number of connected components	OWL [71, 229]
		Average Class Connectivity	$mean((c1, p, c2))$ where $c1$ and $c2$ are instances of classes	OWL [207, 229]

Challenge: What do these metrics tell us over time?



- **Bottomline/Challenges:**

- These metrics are not sufficient to track ***patterns of evolution...***
- ***We need to track changes on a finer granularity level***
- ***We need new metrics (from other fields):***
 - *Time series analyses (change frequencies, seasonality)*
 - *Network science (dynamics of networks)*
 - etc.*

More Open question(s)

- How does consistency evolve over time (and why is this important)?
- *Challenge:*
 - *Our “classical” tools (OWL, SHACL?) are not really useful here directly*

Reasoning+Constraints in Wikidata: OWL?

- Challenges:
- Wikidata does not use OWL!
- Big Open KGs are all quite inconsistent!

RDFS/OWL property	Equivalence established through	Wikidata property
rdf:type	equivalent property (P1628)	instance of (P31)
rdfs:subClassOf	equivalent property (P1628)	subclass of (P279)
rdfs:subPropertyOf	equivalent property (P1628)	subproperty of (P1647)
rdfs:subPropertyOf	equivalent property (P1628)	external subproperty (P2236)
Inverse rdfs:subPropertyOf	equivalent property (P1628)	external superproperty (P2235)
rdfs:range	equivalent property (P1628)	expressed via property constraint (P2302)
rdfs:domain	equivalent property (P1628)	expressed via property constraint (P2302)
rdfs:label	documented as matching ¹²	rdfs:label
rdfs:comment	documented as matching ¹²	schema:description
rdf:first	documented as matching ¹²	expressed via series ordinal (P1545)
rdf:rest	documented as matching ¹²	expressed via series ordinal (P1545)
rdfs:member	documented as matching ¹²	part of (P361)
Inverse rdfs:member	inverse property (P1696) of part of (P361)	has part (P527)
owl:equivalentProperty	equivalent property (P1628)	equivalent property (P1628)
owl:equivalentClass	equivalent property (P1628)	equivalent class (P1709)
owl:inverseOf	equivalent property (P1628)	inverse property (P1696)
owl:differentFrom	equivalent property (P1628)	different from (P1889)
owl:unionOf	equivalence intended ¹³	union of (P2737)
owl:disjointUnionOf	equivalence intended ¹³	disjoint union of (P2738)
owl:onProperty	no documented equivalence	possible candidates: property constraint (P2302)
owl:sameAs	no documented equivalence	possible candidates: exact match (P2888), said to be the same as (P460)
owl:disjointWith	no documented equivalence	N/A
owl:propertyDisjointWith	no documented equivalence	N/A
owl:propertyChainAxiom	no documented equivalence	N/A
owl:assertionProperty	no documented equivalence	N/A

An Analysis of Links in Wikidata

Armin Haller¹[0000-0003-3425-0780], Axel Polleres²[0000-0001-5670-1146], Daniil Dobriy²[0000-0001-5242-302X], Nicolas Ferranti²[0000-0002-5574-1987], and Sergio J. Rodríguez Méndez¹[0000-0001-7203-8399]

Reasoning+Constraints in Wikidata: SHACL?

- Challenges:
- Wikidata does not use SHACL either!
 - Formalization
 - Analysis of violations over time!

Formalizing and Validating Wikidata's Property Constraints using SHACL and SPARQL

Nicolas Ferranti ^{a,*}, Jairo Francisco De Souza ^b, Shqiponja Ahmetaj ^c and Axel Polleres



country of citizenship (P27)

the object is a country that recognizes the subject as its citizen

Constraints

property constraint

value-type constraint

class

former or current state
country in a fiction work
nation

contemporary constraint (Q25796498)

type of constraint for Wikidata properties: used to specify that the subject and the object have to coincide or coexist at some point of h

↳ In more languages

Configure

Language	Label	Description	Also known as
English	contemporary constraint	type of constraint for Wikidata properties: used to specify that the subject and the object have to coincide or coexist at some point of history	

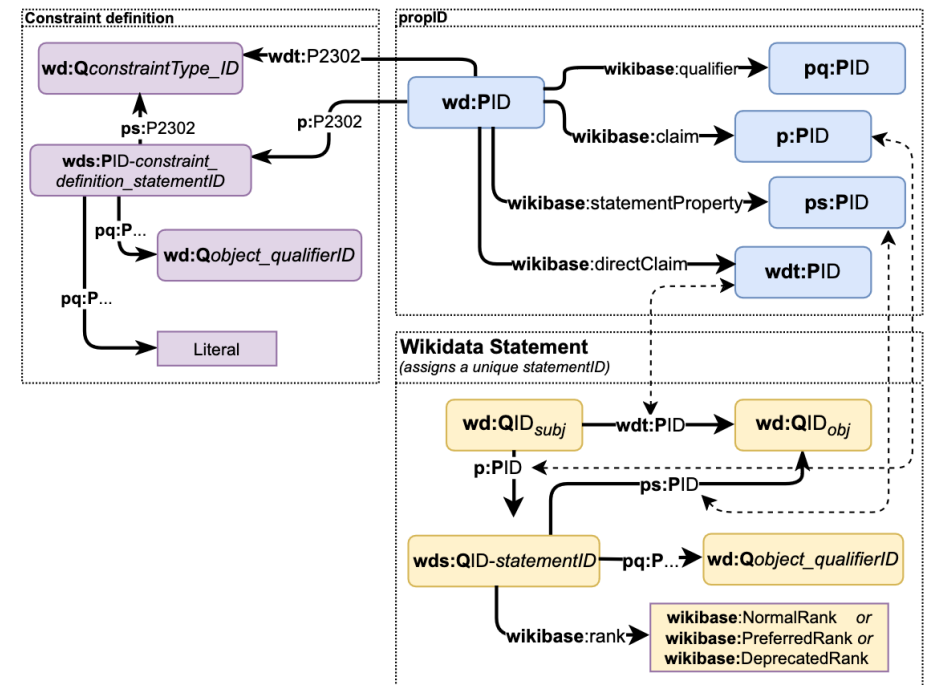
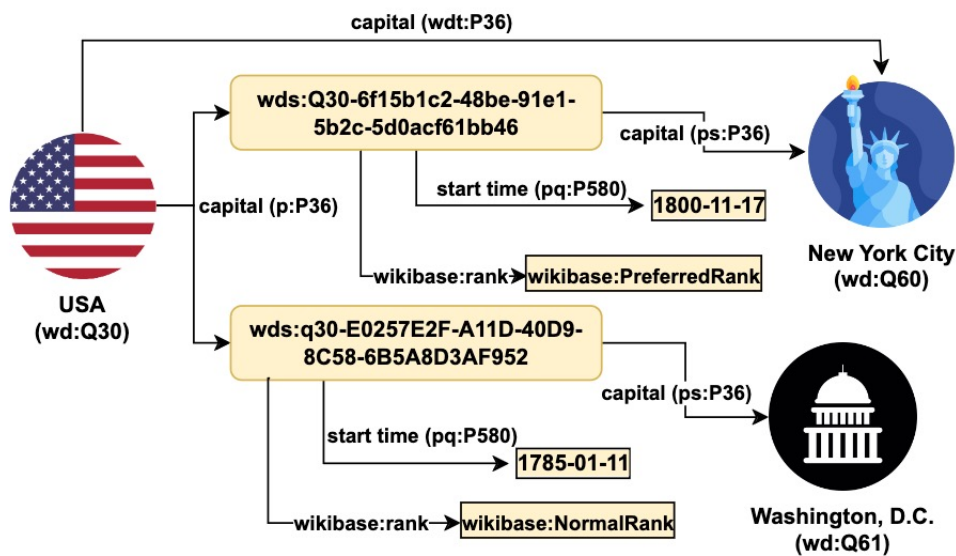
fictional country
city-state
political territorial entity
sovereign state
empire
citizenship
colony
federated state
instance of

relation

<https://www.semantic-web-journal.net/content/formalizing-and-validating-wikidatas-property-constraints-using-shacl-and-sparql>

Reasoning+Constraints in Wikidata: SHACL?

- Challenges:
- Wikidata does not use SHACL either!
 - Formalization - *what makes it challenging?* ---> Reification galore! ;-)



Main questions:

- *Which publicly accessible, open KGs are observable in a manner that would allow a longitudinal analysis of their evolution and how?*
- *Do we have the right metrics to analyse KGs' evolution?*
- ***Do we have the right techniques to process evolving KGs?***

Do we have the right techniques to process evolving KGs?

- What else will you find in our paper? Survey of ...
 - **Storage techniques** for *evolving KGs*
 - **Reasoning & Querying techniques** for *evolving KGs*
 - **Learning & Embeddings** for *evolving KGs*
- **Challenges:**
 - *Again: How do we make these methods scale to large-scale, evolving, collaborative KGs?*
 - *E.g. How to reason and query over evolving KGs?*
 - Need to extend our techniques to deal with reification?
 - What's the "right" reification?
 - Labelled property Graphs?
 - RDF-*?
 - Wikidata's proprietary reification mechanism?
 - How to scale and modularize existing techniques over highly reified KGs?



Some take home messages:

- There's a lot to learn about the dynamics of (Open) KGs!
 - Understanding the evolution of knowledge (graphs) is a hot topic!

Yet, there are some major challenges:

- **Data Availability:**
 - (fine-grained) data about their evolution it not available (Streams!)
 - We need more long-tail data!
 - It's hard to to sustain efforts to sustain data about evolution!
- **Metrics:**
 - We need metrics and technniques to analyse KG dynamics and evolution:
 - New metrics, look into other fields!
 - Adaptions and extensions of existing metrics
- **Techniques: Storage/Querying/Learning**
 - How to scale and modularize existing techniques over highly reified KGs?
 - Dynamic Embeddings/Model Dynamics?
- P.S.: Submit to TGDK 😊

... would have loved to discuss these challenges with you in person 😞

