How Does Knowledge Evolve in Open Knowledge Graphs?

*Axel Polleres, **Daniil Dobriy**, Nicolas Ferranti*
(authors present here at ISWC)

How Does Knowledge Evolve in Open Knowledge Graphs?

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

Romana Pernisch[1]
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

Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 7, pp. 7:1–7:60
Transactions on Graph Data and Knowledge
Schloss Dagstuhl – Leibniz-Zentrum für Informatik

Disclaimer: Our paper is/was a call for ACTION!

Main questions raised (not answered!) in the paper:

- *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?*

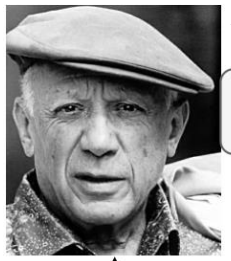
- *Do we have the right techniques to process evolving KGs?*

What dimensions does KG evolution have at all?

- **Temporal KGs**: Time as Data (“valid time”, Dates/Timestamps, Intervals, start/end events)
- **Time-varying KGs**: Time as meta-data/log data (“transaction time”, Edit events(Dates/Timestamps)

***Bottomline**:*
Tracing is time in published KGs is not standardized:

- *Not observable/ recorded for many KGs*
- *Modeled within or outside of the KG*
- *Using different forms of reification*



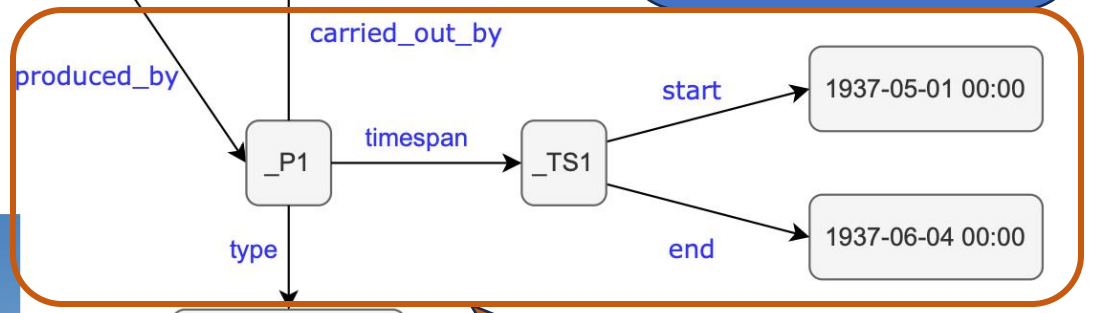
“Guernica” (Q175036) was created on 28 November 2012 in Wikidata

“Pablo Picasso” (Q5593) added on 1 November 2012

current_location

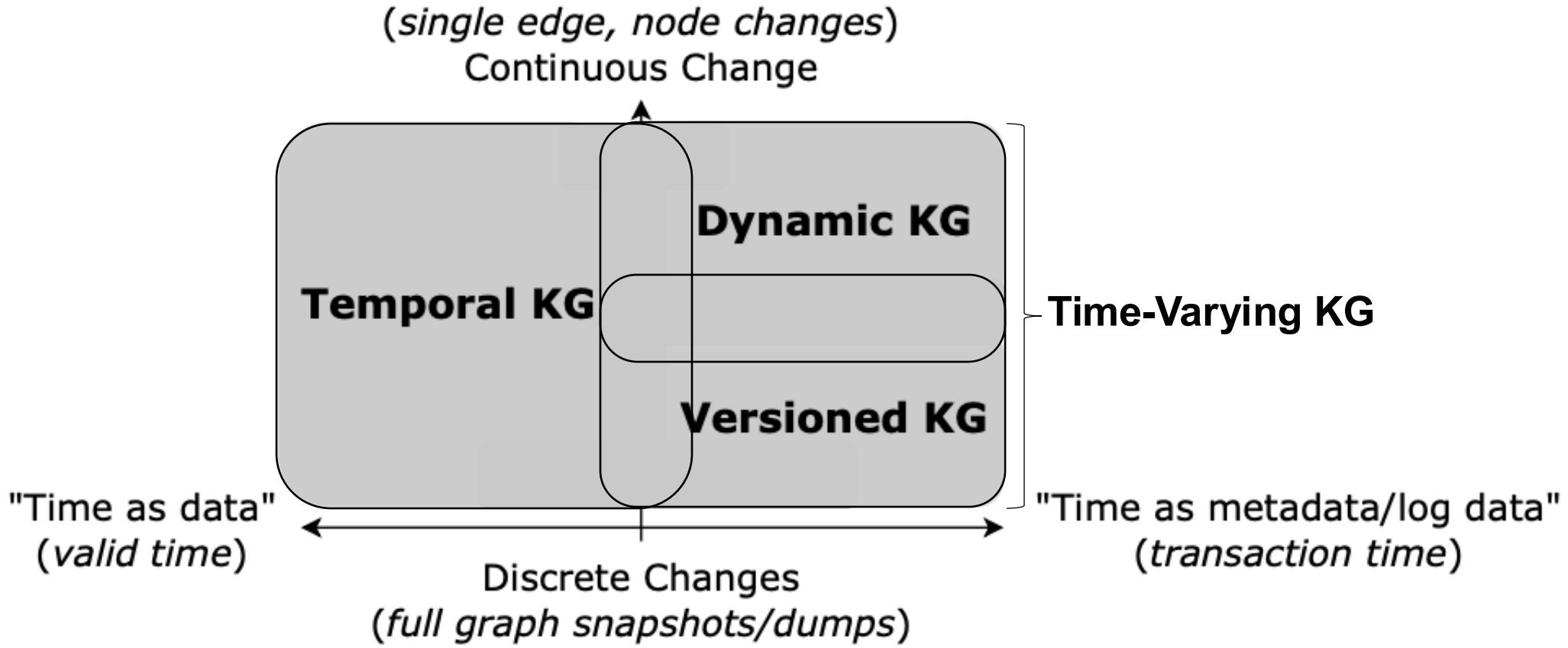
Madrid

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



“Guernica” (Q175036) was painted between May and June 1937

What dimensions does evolution have at all?

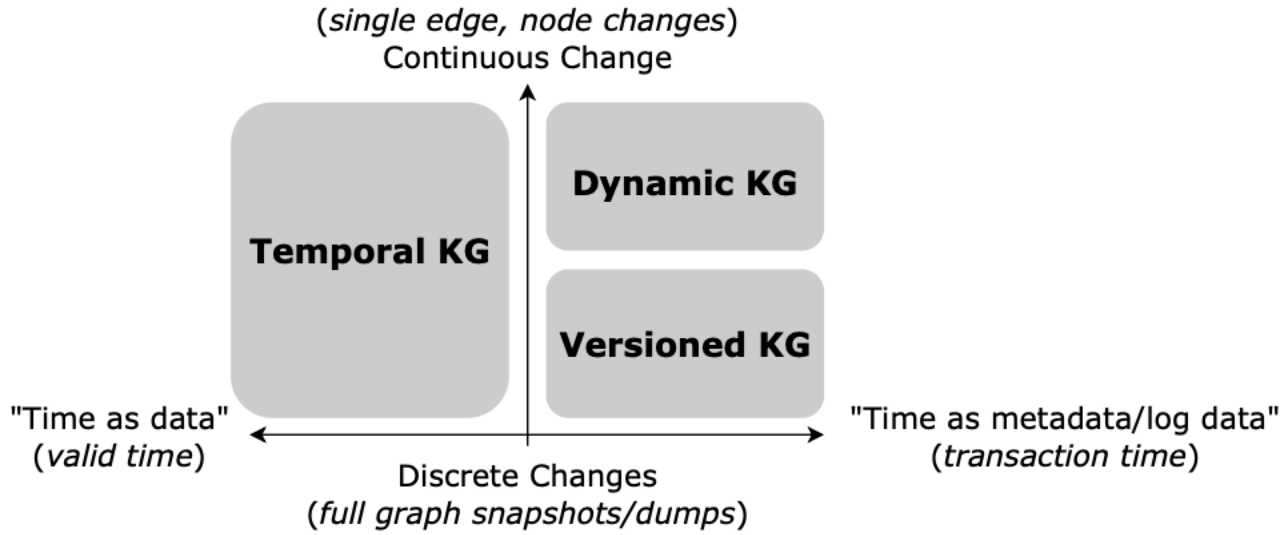


additional

What dimensions does evolution have at all?

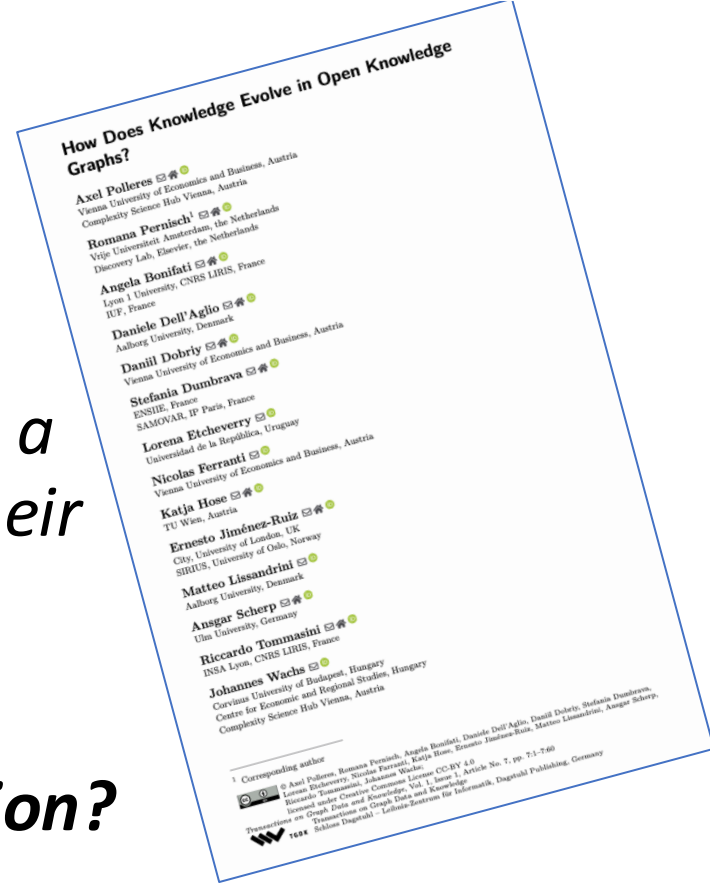
- Instance evolution
- Schema evolution
- Collaboration evolution

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



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?*


How Does Knowledge Evolve in Open Knowledge Graphs?

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

Romana Pernisch[1]
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

Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 7, pp. 7:1-7:60
Transactions on Graph Data and Knowledge
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

From static metrics to dynamic metrics

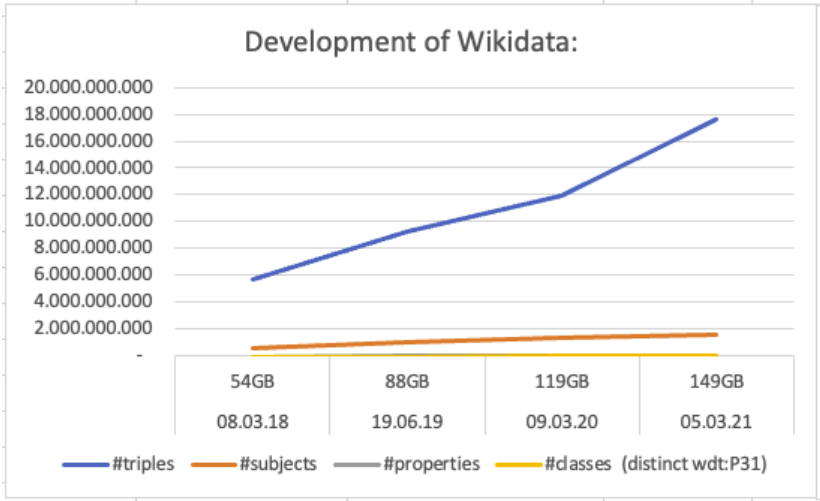
- Basic (static) (***Knowledge) Graph metrics*** "over time"…

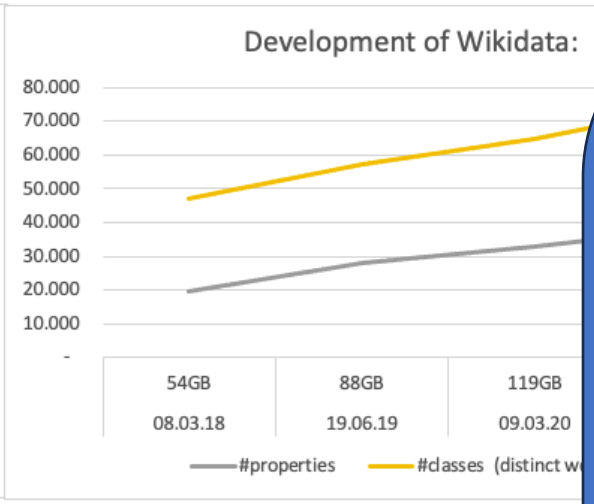
	Metric	Description	Used/Defined in
Graph	Absolute depth	d_a = sum over the cardinality of each path in a set of paths in graph	IsA graph [11, 91, 143, 252]
	Average depth	d_a / $\|paths\|$	IsA graph [11, 91, 143, 252], graph [37, 71], OWL schema [73, 71]
	Maximal depth	longest path	IsA graph [11, 91, 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]

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]
	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]

… are not ENOUGH to understand evolution!

Challenge: What do these metrics tell us over time?





- Bottomline/Challenges:
  - ***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.
  - ***Metrics to track the evolution of consistency in KGs?***

(shameless self-advertisement ;-)) Our work towards such metrics since the TGDK paper:

- How "remote" ontology updates affect reasoning closure for linked data? WebConf2025:

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 1254 - 1258 • https://doi.org/10.1145/3701716.3715478

- How constraint repairs evolve over time? Check our main conf. paper!

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, Austria
[2] Complexity Science Hub Vienna, Austria
[3] LApIC Research Group, Federal University of Juiz de Fora, Brazil

ISWC 2025 NARA JAPAN

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 here in our paper? Survey of …
  - **Storage techniques** for *evolving KGs*
  - **Reasoning & Querying techniques** for *evolving KGs*
  - **Learning & Embeddings** for *evolving KGs*

- *Challenges:*
  - *How do we make these methods scale to large-scale, evolving, collaborative KGs?*
  - *E.g.*
    - How to reason and query over evolving KGs?
    - How to scale and modularize existing techniques over highly reified KGs?

How Does Knowledge Evolve in Open Knowledge Graphs?

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

Romana Pernisch[1]
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

Transactions on Graph Data and Knowledge, Vol. 1, Issue 1, Article No. 7, pp. 7:1–7:60
Transactions on Graph Data and Knowledge
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany