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
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:
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:
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:
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**:
Instance data changing in a highly dynamic fashion:

Dynamic constraints violated (inconsistency via dynamics!):
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Even seemingly static data no longer valid: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Even seemingly static data no longer valid...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 :-(

Train RJX565

hasCar

Car26

Car27

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

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

... 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!
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...

<table>
<thead>
<tr>
<th>What did I want to talk about?</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Open Knowledge Graphs?</td>
</tr>
<tr>
<td>- Perspectives on evolution: dimensions of time and temporality</td>
</tr>
<tr>
<td>- Observability of evolution</td>
</tr>
<tr>
<td>- Metrics for evolution</td>
</tr>
<tr>
<td>- How does evolution affect downstream tasks and resp. techniques?</td>
</tr>
<tr>
<td>- Recent own work...</td>
</tr>
</tbody>
</table>

How Does Knowledge Evolve in Open Knowledge Graphs?

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

Romana Pernisch
Vrije Universiteit Amsterdam, the Netherlands
Discover Lab, Electica, the Netherlands

Angela Bonfieti
Lyon 1 University, CNRS LIRIS, France
IUF, France

Daniele Dell’Aglie
Aalborg University, Denmark

Danill Dobryi
Vienna University of Economics and Business, Austria

Stefania Dumbraeva
ENSIEE, France
SAMOVAR, IP Paris, France

Lorena Etchevery
Universidad de la República, Uruguay

Nicolas Ferranti
Vienna University of Economics and Business, Austria

Katja Hase
TU Wien, Austria

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

Matteo Liessandri
Aalborg University, Denmark

Asgar Scherp
Ulm University, Germany

Riccardo Tommasini
INSA Lyon, CNRS LIRIS, France

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

1 Corresponding author

© Axel Polleres, Romana Pernisch, Angela Bonfieti, Daniele Dell’Aglie, Danill Dobryi, Stefania Dumbraeva,
Matteo Liessandri, Nicolas Sarradj, Katja Hase, Ernesto Jiménez-Ruiz, Matteo Liessandri, Asgar Scherp,
Asgar Scherp, Lorena Etchevery, Johannes Wachs

Licensed under Creative Commons License CC-BY 4.0

Proceedings on Graph Data and Knowledge
1648 - Online Informal - Institute für Informatik, Informatik Publishing, Germany
Open, collaborively 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

"Pablo Picasso" (Q5593) added on 1 November 2012
What dimensions does evolution have at all?

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

  Revision history of "Guernica" (Q175036)

  View logs for this item (view abuse log)

  ▼ 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  Yoosf Poornvany (talk | contribs) . . (19,785 bytes) (+13) . . (Page moved from [fawiki] to [fawiki])
  • (cur | prev) 01:52, 30 September 2014  Dexbot (talk | contribs) . . . (19,772 bytes) (-66) . . (Removed Persian alias: " Asians")
  • (cur | prev) 23:34, 17 September 2014  OC Ripper (talk | contribs) . . . (19,838 bytes) (+6,775) . . (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 [2742338])
  • (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  AkkaakkBot (talk | contribs) . . (10,785 bytes) (+36) . . (Updated Item: - set pa label to [task 5])
  • (cur | prev) 23:56, 11 May 2014  AkkaakkBot (talk | contribs) . . (10,747 bytes) (-28) . . (Updated Item: - remove redundant simple label [task 7])

• Edit events(Dates/Timestamps)

• 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") 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?

- Instance evolution
- Schema evolution
- Collaboration models

(Time-varying KG)

(Dynamic KG)

(Temporal KG)

-Versioned KG

"Time as data" (valid time)

Discrete Changes (full graph snapshots/dumps)

"Time as metadata/log data" (transaction time)
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:

<table>
<thead>
<tr>
<th>Level</th>
<th>Queryable</th>
<th>Collaborative</th>
<th>Formats</th>
<th>Protocol</th>
<th>Metadata</th>
<th>Temporality</th>
<th>Timeliness</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>Yes</td>
<td>Yes</td>
<td>NT, TTL, HDT, JSON</td>
<td>HTTP, SPARQL</td>
<td>schema.org</td>
<td>No</td>
<td>2-3 Days</td>
</tr>
<tr>
<td>S</td>
<td>Yes</td>
<td>Yes</td>
<td>NT, TTL, JSON</td>
<td>HTTP, SPARQL</td>
<td>schema.org</td>
<td>No</td>
<td>2-3 Days</td>
</tr>
<tr>
<td>CL</td>
<td>Yes</td>
<td>Yes</td>
<td>JSON</td>
<td>SSE</td>
<td>No</td>
<td>Event TS</td>
<td>Seconds</td>
</tr>
</tbody>
</table>

### DBpedia

<table>
<thead>
<tr>
<th>Level</th>
<th>Queryable</th>
<th>Collaborative</th>
<th>Formats</th>
<th>Protocol</th>
<th>Metadata</th>
<th>Temporality</th>
<th>Timeliness</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>Yes</td>
<td>Partial</td>
<td>NT</td>
<td>HTTP, SPARQL</td>
<td>No</td>
<td>No</td>
<td>Quarterly</td>
</tr>
<tr>
<td>S</td>
<td>Yes</td>
<td>Yes</td>
<td>RDF</td>
<td>HTTP, SPARQL</td>
<td>No</td>
<td>No</td>
<td>Daily</td>
</tr>
<tr>
<td>CL</td>
<td>Yes</td>
<td>Yes</td>
<td>RDF</td>
<td>HTTP</td>
<td>No</td>
<td>Graph TS</td>
<td>Daily</td>
</tr>
</tbody>
</table>

*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!

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th>Details</th>
<th>Provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latest Wikidata (3rd march 2021)</td>
<td>53GB uncomp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latest Wikidata (9th march 2020)</td>
<td>50GB uncomp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBPedia 2016-10 English</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What is the “DBpedia Snapshot” Release?

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 —

- EN Wikipedia data — A small, but very useful, subset (~ 1 Billion triples or 14%) of the whole DBpedia extraction using the DBpedia Information Extraction Framework (DIEF), 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.

- 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 DBpedia Wayback Machine * SEMANTiCS2015

Javier D. Fernández  Patrik Schneider  Jürgen Umbrich  
Vienna University of Economics and Business, Vienna University of Economics and Business, Vienna University of Economics and Business, Vienna, Austria Vienna, Austria Vienna, Austria  
javier.fernandez@wu.ac.at patrik.schneider@wu.ac.at 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

• Powered by MediaWiki software.
• We know of 60527 currently active wikis.
  (It’s a lot 😊)
How many **Semantic MediaWikis**?

SMW Cloud (1458 wikis)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Triples</th>
<th>#Subjects</th>
<th>#Predicates</th>
<th>#Objects</th>
<th>#Literals</th>
</tr>
</thead>
<tbody>
<tr>
<td>LODStats [1]</td>
<td>192,230,648</td>
<td>Not reported</td>
<td>49,916</td>
<td>Not reported</td>
<td>90,261,655</td>
</tr>
<tr>
<td>SMW Cloud</td>
<td>236,505,705</td>
<td>24,010,566</td>
<td>52,670</td>
<td>66,052,823</td>
<td>160,108,216</td>
</tr>
<tr>
<td>Wikidata 2021</td>
<td>17,662,800,665</td>
<td>1,625,057,179</td>
<td>38,867</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
</tbody>
</table>

Crawled RDF data available at [semantic-data.cluster.ai.wu.ac.at/smwcloud/](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!
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

<table>
<thead>
<tr>
<th>KG</th>
<th>Expert-driven</th>
<th>Crowd-sourced</th>
<th>Resource-dependent</th>
<th>Community-driven</th>
<th>Bot-assisted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikidata [244]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DBpedia [148]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓ (✓)</td>
</tr>
</tbody>
</table>

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

Following:
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:

- **Challenge:**
  - Again: User/collaboration **data not “readily available”**

**Extract aggregated log for analysis**

<table>
<thead>
<tr>
<th>index</th>
<th>username</th>
<th>userID</th>
<th>vocabulary</th>
<th>entity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&quot;David&quot;</td>
<td>1234</td>
<td>P31:Q128207</td>
<td>Q2</td>
</tr>
<tr>
<td>2</td>
<td>&quot;David&quot;</td>
<td>1234</td>
<td>P2067</td>
<td>Q2</td>
</tr>
<tr>
<td>3</td>
<td>&quot;David&quot;</td>
<td>u1234</td>
<td>P138</td>
<td>Q2</td>
</tr>
<tr>
<td>4</td>
<td>&quot;David&quot;</td>
<td>u1234</td>
<td>P2579</td>
<td>Q3</td>
</tr>
</tbody>
</table>

**Position paper, Wikidata Workshop 2022**

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

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

Nicola Pascal Krenn

http://polleres.net/supervised_theses/Nicola_Krenn_MSc_2023.pdf
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?
From static metrics to dynamic metrics

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

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Used/Defined in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute depth</td>
<td>$d_a = \text{sum over the cardinality of each path in a set of paths in graph}$</td>
<td>IsA graph [11, 91, 143, 252], graph [37, 71], OWL schema [73, 71]</td>
</tr>
<tr>
<td>Average depth</td>
<td>$d_a /</td>
<td>paths</td>
</tr>
<tr>
<td>Maximal depth</td>
<td>longest path</td>
<td>IsA graph [11, 91, 143], graph [37]</td>
</tr>
<tr>
<td>Number of paths</td>
<td>$</td>
<td>paths</td>
</tr>
<tr>
<td>Tangledness</td>
<td>$\frac{n_G}{t}$, $n_G = \text{cardinality of G}$, $t = \text{cardinality of the set of nodes with more than one ingoing IsA arc in G}$</td>
<td>IsA graph [11, 91, 143]</td>
</tr>
<tr>
<td>Degree Distribution</td>
<td>mean-square deviation of the degree of graph nodes</td>
<td>graph [37, 67, 143]</td>
</tr>
</tbody>
</table>
From static metrics to dynamic metrics

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

<table>
<thead>
<tr>
<th>Primitives</th>
<th>Entities</th>
<th>number of entities, classes and instances</th>
<th>graph [37, 143], IsA graph [91], OWL [215], DAG [252]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Properties</td>
<td>number of unique properties or relations</td>
<td>OWL schema [174], OWL [229, 231, 233], DAG [254]</td>
</tr>
<tr>
<td></td>
<td>Classes</td>
<td>$</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>Instances</td>
<td>$</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>Object properties</td>
<td>$P_o = \text{number of object properties (non-inheritance)}$</td>
<td>Schema [143], OWL [215, 231]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Knowledge Graph</th>
<th>Depth of Inheritance Tree</th>
<th>Tree [174], OWL [73, 207, 229], DAG [252]</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-Box/Schema</td>
<td>Property Class Ratio</td>
<td>$\frac{</td>
</tr>
<tr>
<td></td>
<td>Inheritance Richness</td>
<td>$\frac{</td>
</tr>
<tr>
<td></td>
<td>Attribute Richness</td>
<td>$\frac{</td>
</tr>
<tr>
<td></td>
<td>Class Property Ratio</td>
<td>$\frac{</td>
</tr>
<tr>
<td>A-Box/Data</td>
<td>Average Population</td>
<td>$\frac{</td>
</tr>
<tr>
<td></td>
<td>Cohesion</td>
<td>number of connected components</td>
</tr>
<tr>
<td></td>
<td>Average Class Connectivity</td>
<td>$mean(</td>
</tr>
</tbody>
</table>
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!

<table>
<thead>
<tr>
<th>RDFS/OWL property</th>
<th>Equivalence established through Wikidata property</th>
<th>Wikidata property</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdf:type</td>
<td>equivalent property (P1628)</td>
<td>instance of (P31)</td>
</tr>
<tr>
<td>rdfs:subClassOf</td>
<td>equivalent property (P1628)</td>
<td>subclass of (P279)</td>
</tr>
<tr>
<td>rdfs:subPropertyOf</td>
<td>equivalent property (P1628)</td>
<td>subproperty of (P1647)</td>
</tr>
<tr>
<td>rdfs:subPropertyOf</td>
<td>equivalent property (P1628)</td>
<td>external subproperty (P2236)</td>
</tr>
<tr>
<td>inverse rdfs:subPropertyOf</td>
<td>equivalent property (P1628)</td>
<td>external superproperty (P2235)</td>
</tr>
<tr>
<td>rdfs:range</td>
<td>equivalent property (P1628)</td>
<td>expressed via property constraint (P2302)</td>
</tr>
<tr>
<td>rdfs:domain</td>
<td>equivalent property (P1628)</td>
<td>expressed via property constraint (P2302)</td>
</tr>
<tr>
<td>rdfs:label</td>
<td>documented as matching[2]</td>
<td>rdfs:label</td>
</tr>
<tr>
<td>rdfs:comment</td>
<td>documented as matching[2]</td>
<td>schema:description</td>
</tr>
<tr>
<td>rdf:first</td>
<td>documented as matching[2]</td>
<td>expressed via series ordinal (P1545)</td>
</tr>
<tr>
<td>rdf:rest</td>
<td>documented as matching[2]</td>
<td>expressed via series ordinal (P1545)</td>
</tr>
<tr>
<td>rdfs:member</td>
<td>documented as matching[2]</td>
<td>rdfs:member</td>
</tr>
<tr>
<td>inverse rdfs:member</td>
<td>inverse property (P1696) of part of (P361)</td>
<td>part of (P361)</td>
</tr>
<tr>
<td>owl:equivalentProperty</td>
<td>equivalent property (P1628)</td>
<td>equivalent property (P1628)</td>
</tr>
<tr>
<td>owl:equivalentClass</td>
<td>equivalent property (P1628)</td>
<td>equivalent class (P1709)</td>
</tr>
<tr>
<td>owl:inverseOf</td>
<td>equivalent property (P1628)</td>
<td>inverse property (P1696)</td>
</tr>
<tr>
<td>owl:disjointFrom</td>
<td>equivalent property (P1628)</td>
<td>different from (P1889)</td>
</tr>
<tr>
<td>owl:unionOf</td>
<td>equivalence intended[3]</td>
<td>union of (P2737)</td>
</tr>
<tr>
<td>owl:disjointUnionOf</td>
<td>equivalence intended[3]</td>
<td>disjoint union of (P2738)</td>
</tr>
<tr>
<td>owl:onProperty</td>
<td>no documented equivalence</td>
<td>possible candidates: property constraint (P2302)</td>
</tr>
<tr>
<td>owl:sameAs</td>
<td>no documented equivalence</td>
<td>possible candidates: exact match (P2888), said to be the same as (P460)</td>
</tr>
<tr>
<td>owl:disjointWith</td>
<td>no documented equivalence</td>
<td>N/A</td>
</tr>
<tr>
<td>owl:propertyDisjointWith</td>
<td>no documented equivalence</td>
<td>N/A</td>
</tr>
<tr>
<td>owl:propertyChainAxiom</td>
<td>no documented equivalence</td>
<td>N/A</td>
</tr>
<tr>
<td>owl:assertionProperty</td>
<td>no documented equivalence</td>
<td>N/A</td>
</tr>
</tbody>
</table>
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, b, Jairo Francisco De Souza b, Shqiponja Ahmetaj c and Axel Polleres

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 lost 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 techniques 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 😊