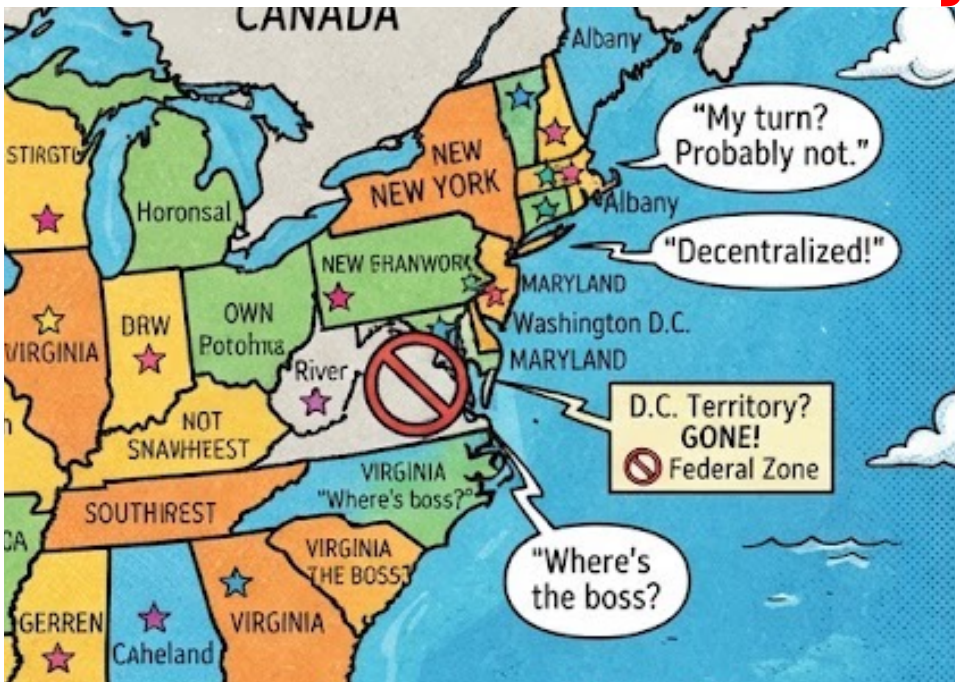
# Let this sink in:

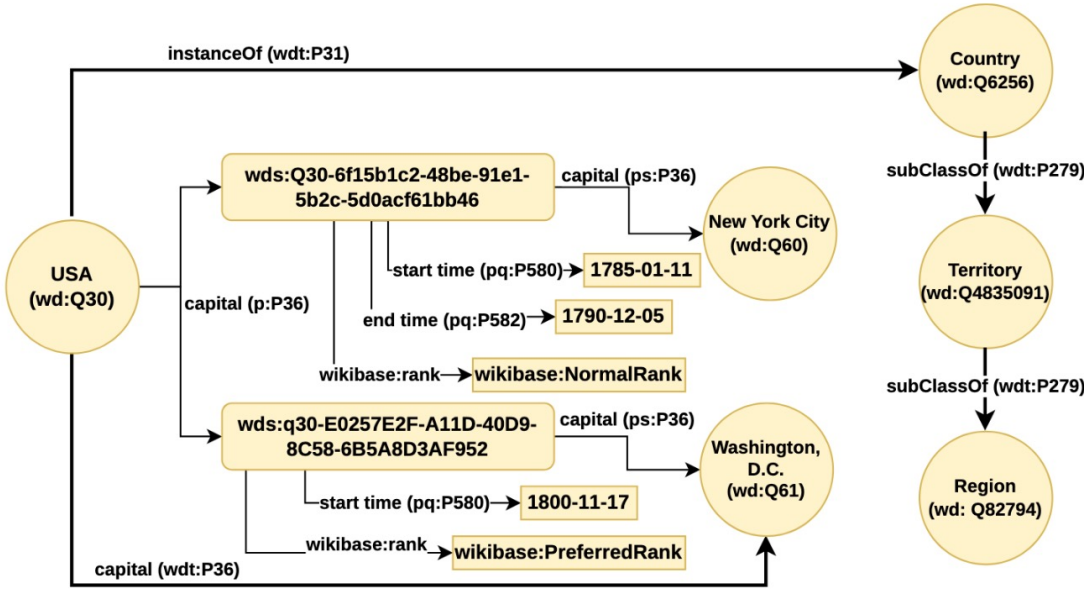
WIKIDATA

ISWC BARI 2026

According to Wikidata,

**the USA did not have a proper capital between Nov 2018 and May 2019** ... for **179 days(!)**




Learn why at

**Poster 7**: Nicolas Ferranti, Axel Polleres - Reconstructing Wikidata Statement Lifecycles for Temporal Maintenance Analysis

BIlAI

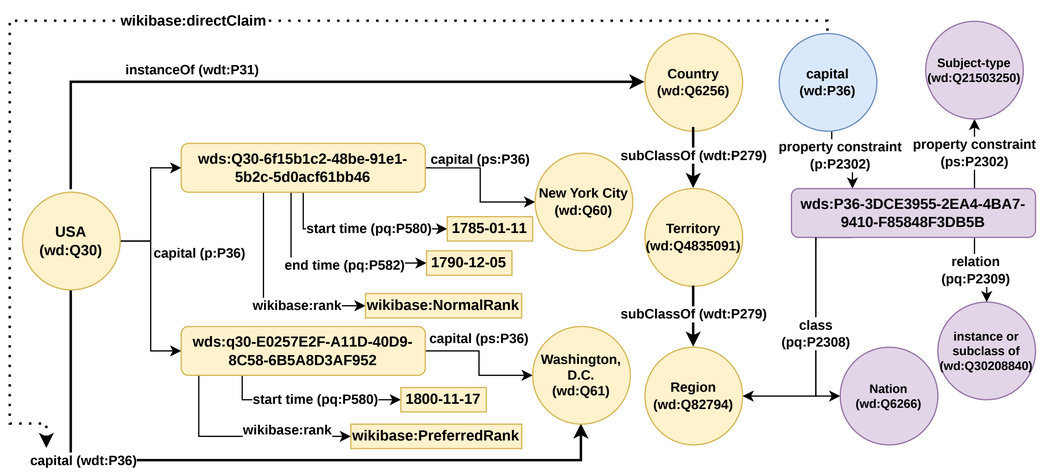
# Reconstructing Wikidata Statement Lifecycles for Temporal Maintenance Analysis

Nicolas Ferranti, Axel Polleres

Institute for Data, Process and Knowledge Management, Wirtschaftsuniversität Wien

## 1- Knowledge Evolution on Wikidata

The Knowledge on Wikidata is constantly evolving. Data, schema, and constraints evolve at the same time.




**Problem:** (1) Static analysis often overlooks specific changes between snapshots; (2) Temporal consistency is essential when querying, validating, and consuming evolving data, schemas, and constraints in downstream applications, including ML pipelines.

**Solution:** Rebuild Statement lifecycles for fine-grained data analysis [1].

## 2- From Page Histories to Lifecycle Records

**Page histories:**

```
<title>Q42</title>
<ns>0</ns>
<id>138</id>
<revision>
  <id>112</id>
  <timestamp>2012-10-29T18:18:48Z</timestamp>
  <contributor><ip>127.0.0.1</ip></contributor>
  <comment>Import</comment>
  <origin>112</origin>
  <model>wikibase-item</model>
  <format>application/json</format>
  <text bytes="174" sha1="8jgn7adhtq3zz4luz6i99e1p4bmg1vk" xml:space="preserve">
    { "id": "Q42", "labels": { "en": "Douglas Adams" }, "claims": [] }
  </text>
  <sha1>8jgn7adhtq3zz4luz6i99e1p4bmg1vk</sha1>
</revision>
```

**Lifecycle records:**

Table 1:Statement lifecycle records for Douglas Adams `wd:Q42`.

| $s$ | $p$ | $o$ | $t_c$ | $u_c$ | $t_d$ | $u_d$ |
|---|---|---|---|---|---|---|
| wd:Q42 | schema:description | "An English writer and humorist"@en | 2012-10-29 22:03:26 | Denny | 2013-01-07 13:07:27 | Tamzin |
| wd:Q42 | wdt:P21 | wd:Q36794 | 2013-02-06 00:19:16 | PiRSquared17 | 2013-02-06 00:19:25 | PiRSquared17 |
| wd:Q42 | wdt:P50 | wd:Q25169 | 2013-02-15 05:07:10 | Legoktm | 2013-02-15 05:08:19 | Legoktm |
| wd:Q42 | wdt:P227 | "119033364" | 2013-03-11 15:50:26 | Bello | 2013-04-15 20:19:12 | FallingGravity |
| wd:Q42 | wdt:P101 | wd:Q1 | 2013-07-01 16:16:01 | Addshore | 2013-07-01 16:16:52 | Addshore |

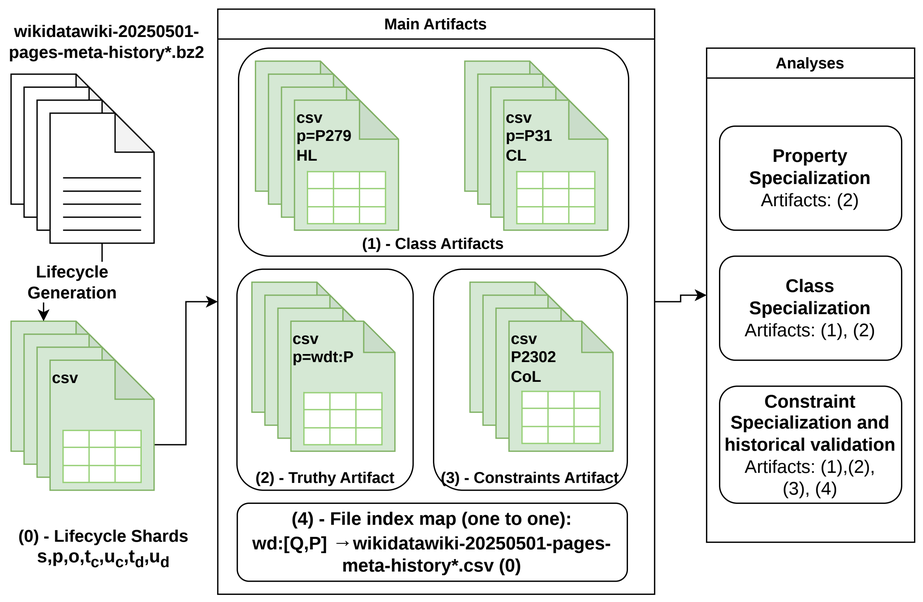
$t_c/u_c$: creation time/user; $t_d/u_d$: deletion time/user.

## Contact Information

- {nicolas.ferranti, axel.polleres}@wu.ac.at

This research was funded in part by the Austrian Science Fund (FWF) [10.55776/COE12].

## Pipeline




## Property-Centered Specialization

- **Editor–property matrix:** $e_{u,p}$ counts edits by user $u$ on property $p$.
- **Property share:** $S(u,p) = \frac{e_{u,p}}{\sum_v e_{v,p}}$, **User focus:** $F(u,p) = \frac{e_{u,p}}{\sum_q e_{u,q}}$
- **Dominance:** User $u$ dominates property $p$ if $\sum_v e_{v,p} \geq 50$, $S(u,p) \geq 0.75$, and $F(u,p) \geq 0.05$. Threshold sensitivity details are in the paper.
- **Specialist types:** Single-property (1 dominated property) or portfolio ($\geq 2$ dominated properties).
- **Results:** 528 dominant pairs across 412 editors: 333 single-property and 79 portfolio specialists.

**Finding:** Portfolio specialists show greater editing focus (34.7% vs. 17.8%) and maintain semantically related properties (e.g., astronomical metadata, dictionary and anime/manga identifiers).

## Historical Constraint Validation

Assess whether reported Wikidata repairs [2] were supported by constraints active at the time.

- **Method:** Compare repaired statement lifecycles with historical constraint definitions.
- **Evaluation:** Six constraint types, comparing the reported constraint ID and any historical constraint of the same type and property.

| Constraint | Repairs | Reported ID no overlap | Any ID no overlap |
|---|---|---|---|
| Conflicts-with | 496k | 36% | 23% |
| Inverse | 2.18M | 8% | 8% |
| Item-requires-statement | 16.2M | 4% | 4% |
| One-of | 15.1k | 55% | 55% |
| Symmetric | 928k | 4% | 4% |
| Value-requires-statement | 15.1M | 8% | 8% |

Percentages indicate repairs without temporal overlap with the corresponding historical constraints.

**Finding:** Between 4% and 55% of reported repairs lack temporal overlap with relevant historical constraints; some predate the constraints themselves.

## References

[1] M. Hofer, M. M. Töpfer, C. Rost, and E. Rahm. Dbpedia-tkg: Capturing wikipedia's evolution as temporal knowledge graphs. In *The Semantic Web - 22nd European Semantic Web Conference, ESWC 2025*, volume 15719 of *Lecture Notes in Computer Science*, pages 262–279. Springer, 2025.

[2] T. Pellissier Tanon, C. Bourgaux, and F. Suchanek. Learning how to correct a knowledge base from the edit history. In *The World Wide Web Conference*, pages 1465–1475, 2019.