Contexts in Knowledge Graphs

G. Falquet

Source:

Aljalbout S. (2021). Contextual knowledge representation and reasoning on knowledge graphs. PhD dissertation. Université de Genève.

Contextual information in not limited to validity

- 2010-2015 : livesIn(bob, geneva) -- validity context
- Wikipedia : inventor(bob, airplane) -- provenance
- WW2 : constructionDate(berlinWall, 1960) -- causality

What is Provenance?

Defined in the PROV-O ontology (https://www.w3.org/TR/prov-o/)

An prov:Entity is a physical, digital, conceptual, or other kind of thing with some fixed aspects; entities may be real or imaginary.

An prov:Activity is something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities.

An prov:Agent is something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent's activity.

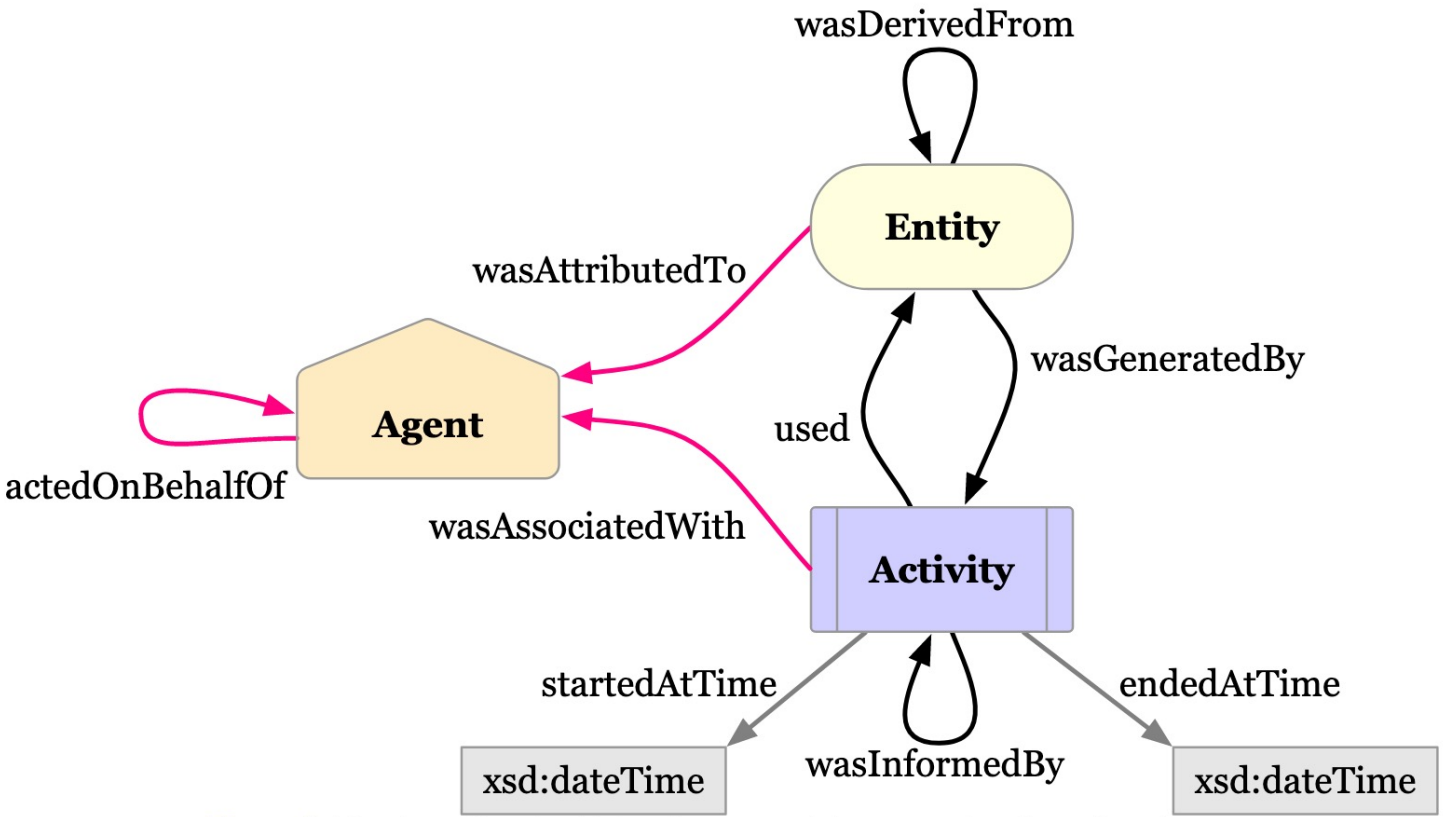
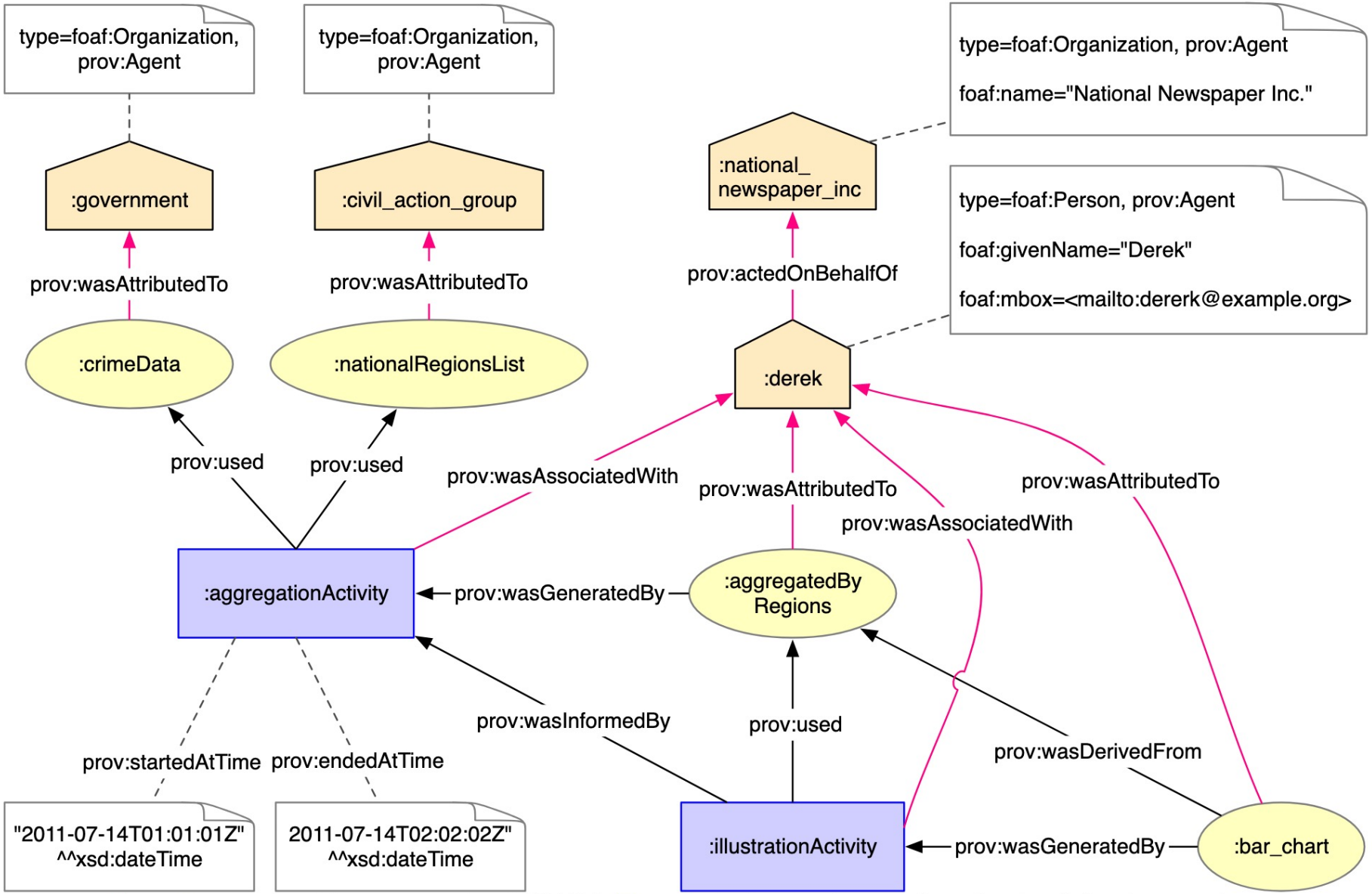
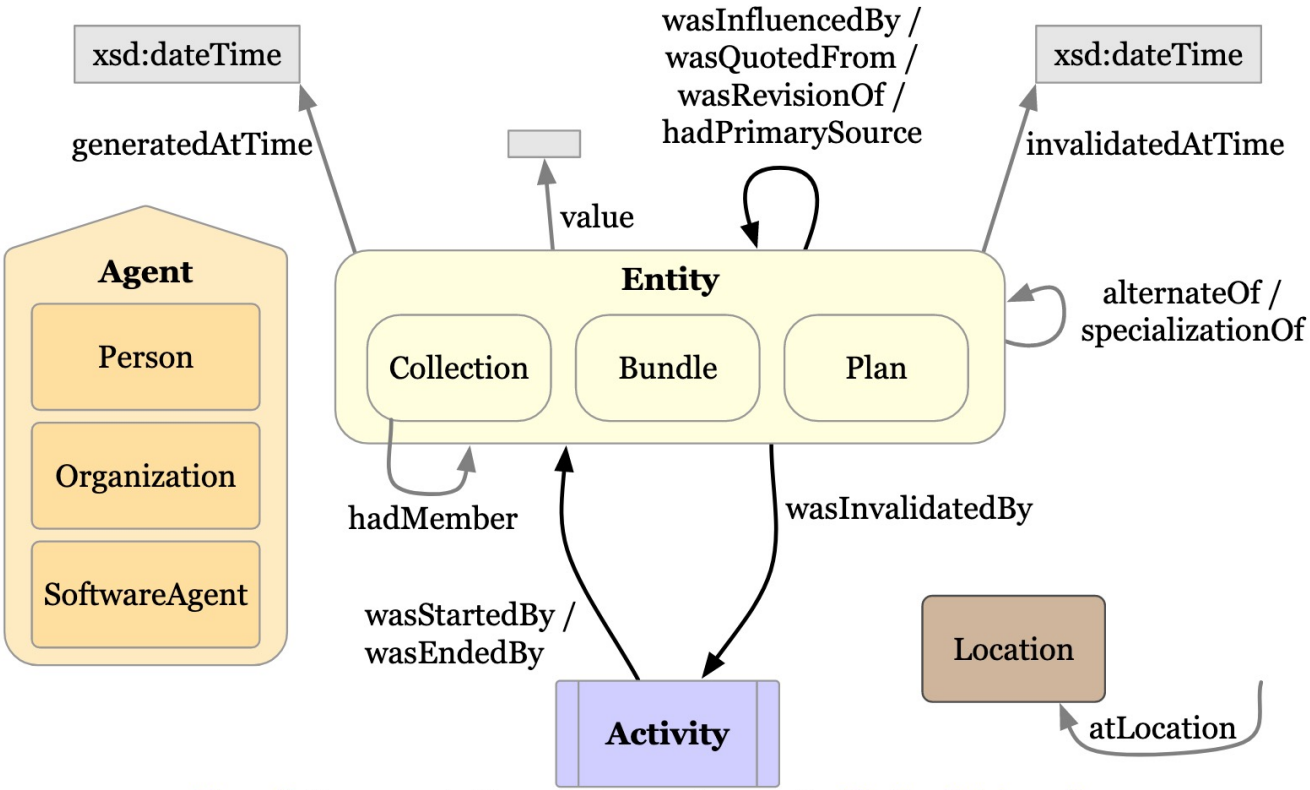


Figure 1. The three Starting Point classes and the properties that relate them. The diagrams in this document depict Entities as yellow ovals, Activities as blue rectangles, and Agents as orange pentagons. The responsibility properties are shown in pink.


type=foaf:Organization, prov:Agent
type=foaf:Organization, prov:Agent
type=foaf:Organization, prov:Agent
foaf:name="National Newspaper Inc."
:government
:civil_action_group
:national_ newspaper_inc
type=foaf:Person, prov:Agent
foaf:givenName="Derek"
foaf:mbox=<mailto:dererk@example.org>
prov:wasAttributedTo
prov:wasAttributedTo
prov:actedOnBehalfOf
:crimeData
:nationalRegionsList
:derek
prov:used
prov:used
prov:wasAssociatedWith
prov:wasAttributedTo
prov:wasAttributedTo
prov:wasAssociatedWith
:aggregationActivity
prov:wasGeneratedBy
:aggregatedBy Regions
prov:wasInformedBy
prov:used
prov:wasDerivedFrom
prov:startedAtTime
prov:endedAtTime
"2011-07-14T01:01:01Z" ^^xsd:dateTime
2011-07-14T02:02:02Z" ^^xsd:dateTime
:illustrationActivity
prov:wasGeneratedBy
:bar_chart

Expanded terms



Provenance and reasoning

[e1] (wasAttributedTo: P1) : p(a, b)

[e2] (wasAttributedTo: P2) : p(b, c)

[e3] (wasAttributedTo: P3) : transitive(p)

??? : p(a, c)

for example

[e1] (wasAttributedTo: P1) : p(a, b)
[e2] (wasAttributedTo: P2) : p(b, c)
[e3] (wasAttributedTo: P3) : transitive(p)

(wasGeneratedBy: infer1) : p(a, c)

InferenceActivity(infer1)
used(infer1, e1), used(infer1, e2), used(infer1, e3),
wasAssociatedWith(infer1, XYZ-reasoner)

Requirement: operations on contexts

- For provenances
- For validity contexts

[2010-2020]: worksFor(a, k)
[2017-2022]: worksFor(b, k)

[2010-2020] ∩ [2017-2022] : colleague(a, b)

- For confidence / probability contexts

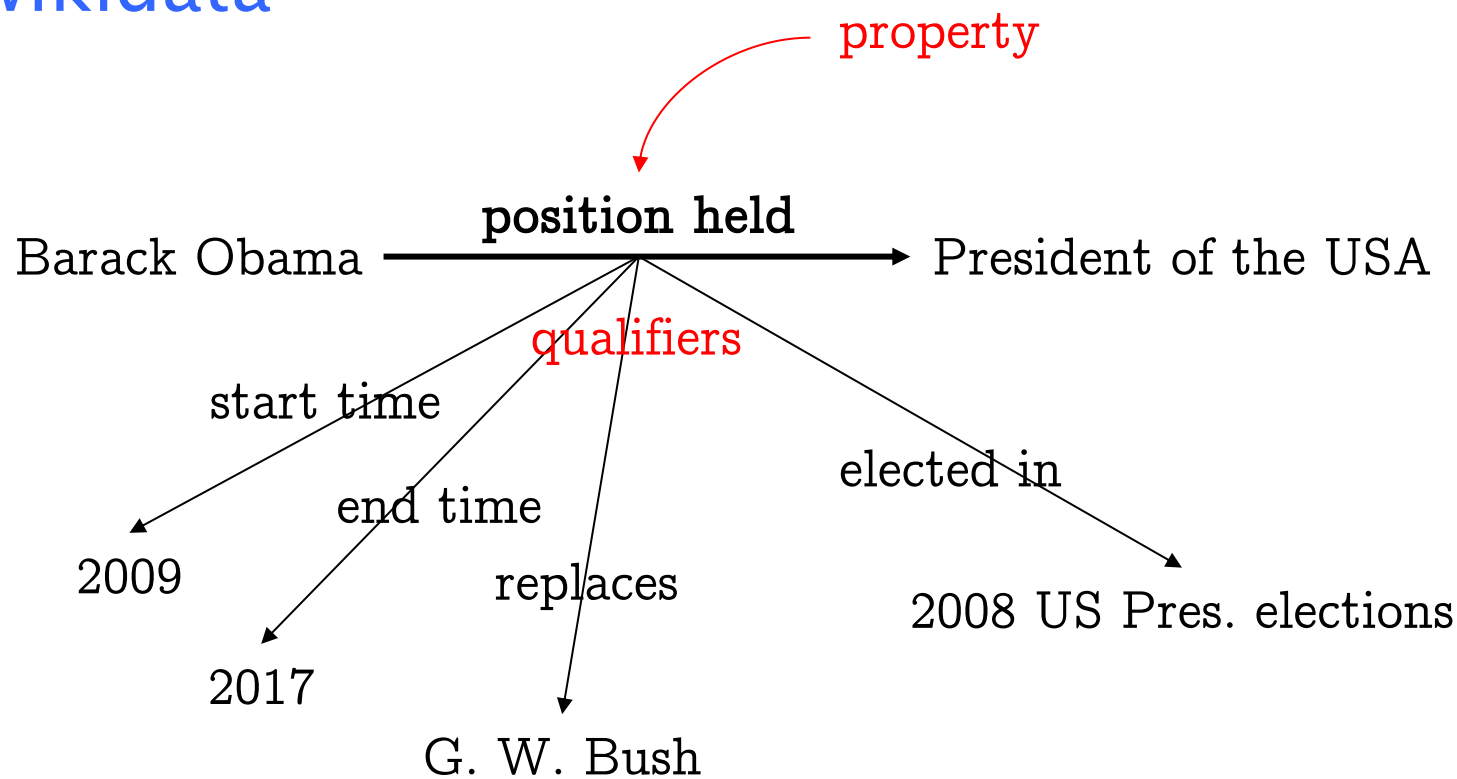
0.7 : Toxic(t)
0.8 : disposedOfIn(t, a)

0.56 : PollutedArea(a)

Contexts in knowledge graphs

- Large graphs
- No strict schema or logical formalism
- Need for contextualization
- Contextual information generally attached to the graph edges

in wikidata


property
position held
Barack Obama
President of the USA
qualifiers
start time
2009
end time
2017
replaces
G. W. Bush
elected in
2008 US Pres. elections

Wikidata in RDF

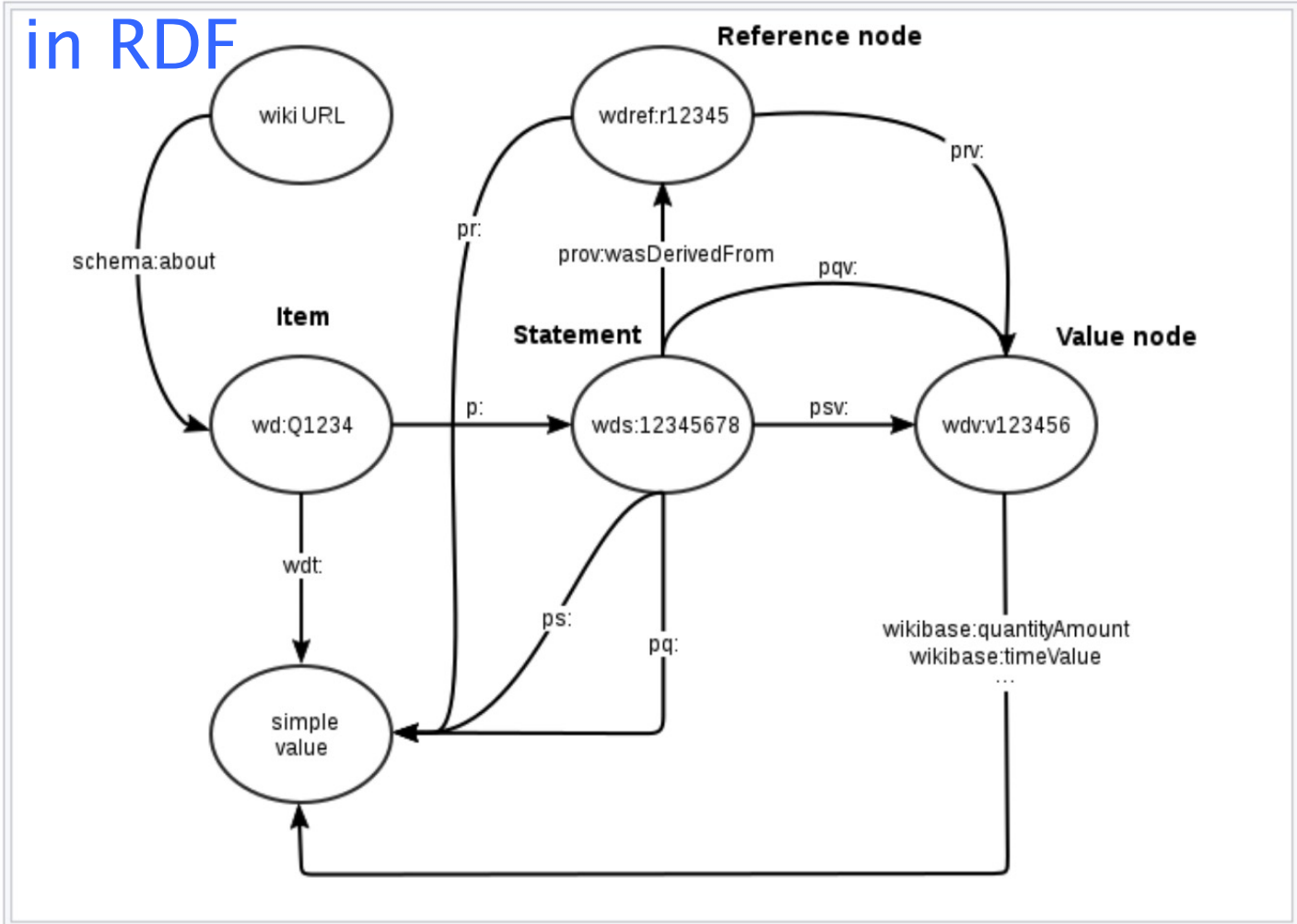


Fig. 5.3 Wikidata data model, taken from https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format

Analysis of the qualifiers

- Validity
- Causality
- Sequence
- Provenance
- Annotation

Validity qualifiers in wikidata

Qualifier name	ID	Wikidata classification	Data type	Prominence
valid in period	P1264	restrictive qualifier	Item	781 020
start period	P3415	restrictive qualifier	Item	376
end period	P3416	restrictive qualifier	Item	180
start time	P580	unqualified	Point in time	4 284 404(*)
end time	P582	unqualified	Point in time	2 311 341(*)
valid in place	P3005	unqualified	Point in time	42 896
country	P17	depicts	Item	71 824
applies to part	P518	restrictive qualifier	Item	800 949
applies to jurisdiction	P1001	restrictive qualifier	Item	1 187 489
applies to name of item	P5168	restrictive qualifier	Monolingual text	6 312
applies to people	P6001	restrictive qualifier	Item	171
applies to name of value	P8338	restrictive qualifier	Monolingual text	736

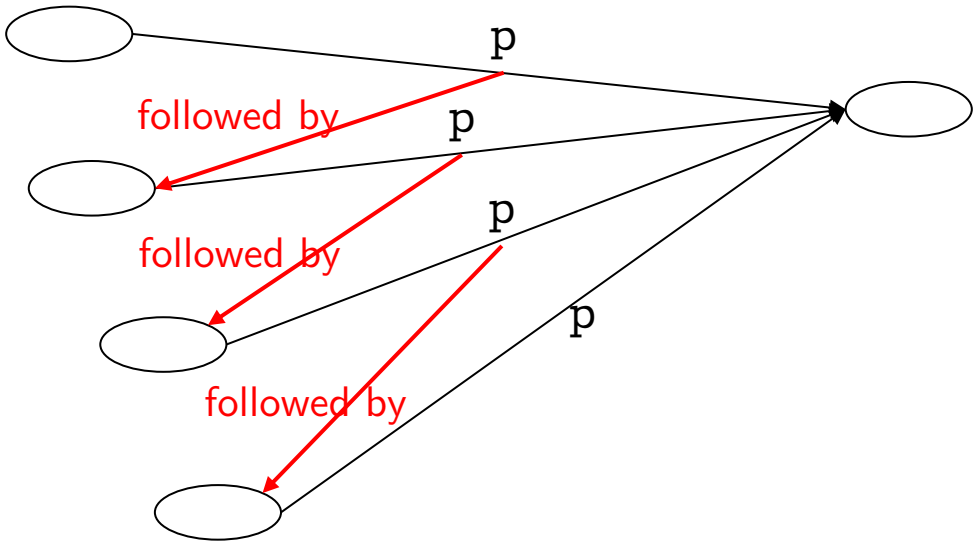
Table 6.1 The validity context qualifiers
(*) these qualifiers are taken from our 114GB (almost complete) downloaded version of the Wikidata 2020-11-09 dump on 22 nov 2020

causality

(Chicago head government Carter Harisson)[
start time : 1893,
end time : 28 october 1893,
end cause : death in the office,
]

sequences

same predicate – object, sequence of subjects



Expressing contextual rules on KG

- take into account several types of contexts

- may be very complex
  - large number of qualifiers
  - different ways to express the same context
    - start time + end time $\leftrightarrow$ start time + duration
  - express conditions on contexts
    - time interval 1 overlaps time interval 2
  - express context construction
    - time interval intersection, provenance combination, ...

Model with Many sorted logic

Many-sorted logic

- first order logic with "typed" (sorted) constants and variables,
- functions and predicates have profiles (arity)

Model with Many sorted logic

Model of a KG G

- A many-sorted vocabulary Σ with a set of sorts S that contains the sort resource ,
- a set F of functions symbols,
- a set of constants C that comprises all the resource denotations that appear as subject, object, or qualifier value in a statement,
- a set of predicate symbols P that is made of all the resource denotations that appear in predicate position in some statement of G.
- Each $p \in P$ has the same arity : resource $\times$resource $\times$s1 $\times \cdots \times$sk

- For each statement $(s,p,o,\{q_i: v_i\})$ of G

  an atom

  $$p(s,o,t_1,\ldots,t_k)$$

  where each t_i is a term of sort s_i, $i = 1,k$.

  typically the sorts are

  {resource, validity, sequence, causality, provenance, annotation}

- A theory Spec over Σ that describes the properties that the (interpretations of the) functions in F are required to satisfy.

Example

```
spouse(Douglas Adams,Jane Belson,
        addTime(timeInterval(25/11/1991,11/5/2001),empty),
        addEndCause(death of person,empty),
        empty,
        empty,
        addPublisher(NNDB,addRetrieved(7/12/2013),empty))
)
```

Algebraic specification of the functions

```
Module ValidityTimeInterval {

import bool
import validityInstantTime

sort timeInterval
sort duration as natural

\\ Generators
op undefined                                          → timeInterval
op interval          instantTime  * instantTime       → timeInterval
op interval          instantTime  * duration          → timeInterval

op equal             timeInterval  * timeInterval     → bool
op disjoint          timeInterval  * timeInterval     → bool
op inside            instantTime  *  timeInterval     → bool
op incl              timeInterval  *  timeInterval    → bool
op testIntersect    timeInterval  * timeInterval      → bool

op union             timeInterval  * timeInterval     → timeInterval
op intersection      timeInterval  * timeInterval     → timeInterval



op startTime         timeInterval                     → instantTime
op endTime           timeInterval                     → instantTime
```

Rule expressions

subclass transitivity

Statement(x,subclass of,y,V1,C1,S1,A1,P1) $\wedge$
Statement(y,subclass of,z,V2,C2,S2,A2,P2)
$\wedge$ (testIntersect(V1,V2))
$\rightarrow$
Statement(x,subclass of,z,inter(V1,V2), union(C1,C2),null, null, union(P1,P2))

Rule expressions

previous – next consistency

Statement(x, p, y, V1, C1, S1, A1, P1) ∧ hasPrevious(S1) = true
→
Statement(previous(S1), p, y,
setTime(V1,interval(null,startTime(extractTime(V1))),
null,
seqWithNext(x),
null,
,P1
)