

Semantic web technologies and principles (in Open Science)

Filip Kovacevic
28.03.2022

- Semantic technologies and principles
- RDF & Ontologies
- Linked Open Data
- RDF Triple stores & SPARQL
- Versioning RDF datasets

Motivation for the Semantic Web

“A new form of web content that is **meaningful to computers** will unleash a revolution of new possibilities”

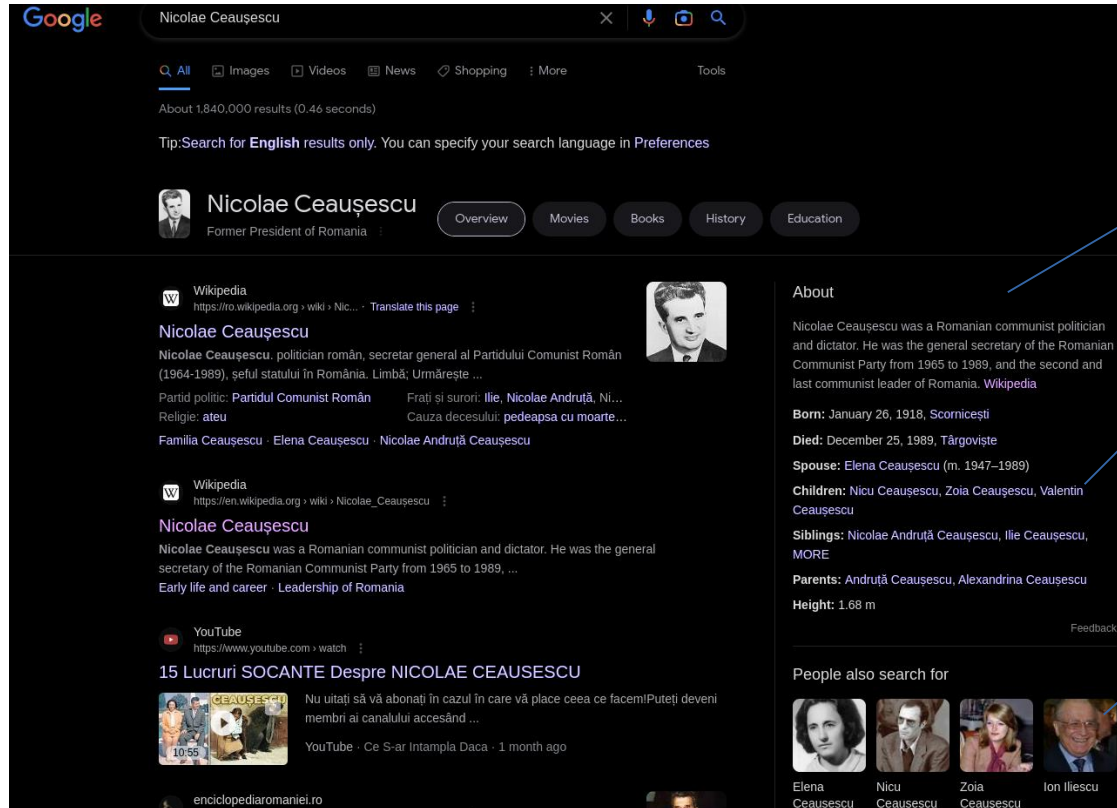
- Tim Berners-Lee

Definition of *semantic*:
of or relating to **meaning** in
language
(merriam-webster dictionary)

Motivation for the Semantic Web

- Extension of the WWW
- Common data formats and exchange protocols
- Common vocabulary
- Re-use data across applications
- Linked data
- Actionable links

Motivating examples (1/3) - Google



Google Nicolae Ceaușescu

About 1,840,000 results (0.46 seconds)

Tip: Search for **English** results only. You can specify your search language in Preferences

Nicolae Ceaușescu
Former President of Romania

Overview Movies Books History Education

Wikipedia
https://ro.wikipedia.org › wiki › Nic... · Translate this page

Nicolae Ceaușescu
Nicolae Ceaușescu, politician român, secretar general al Partidului Comunist Român (1964-1989), șeful statului în România. Limbă: Urmărește ...
Partid politic: **Partidul Comunist Român** Frați și surori: **Ilie, Nicolae Andruță, Ni...**
Religie: **ateu** Cauza decesului: **pedeapsa cu moarte...**
Familia Ceaușescu · Elena Ceaușescu · Nicolae Andruță Ceaușescu

Wikipedia
https://en.wikipedia.org › wiki › Nicolae_Ceaușescu

Nicolae Ceaușescu
Nicolae Ceaușescu was a Romanian communist politician and dictator. He was the general secretary of the Romanian Communist Party from 1965 to 1989, ...
Early life and career · Leadership of Romania

YouTube
https://www.youtube.com › watch

15 Lucruri SOCANTE Despre NICOLAE CEAUȘESCU
Nu uitați să vă abonați în cazul în care vă place ceea ce facem! Puteți deveni membri ai canalului accesând ...
YouTube · Ce S-ar Intampla Daca · 1 month ago

enciclopediaromaniei.ro

About
Nicolae Ceaușescu was a Romanian communist politician and dictator. He was the general secretary of the Romanian Communist Party from 1965 to 1989, and the second and last communist leader of Romania. [Wikipedia](#)
Born: January 26, 1918, Scornicești
Died: December 25, 1989, Târgoviște
Spouse: [Elena Ceaușescu](#) (m. 1947–1989)
Children: [Nicu Ceaușescu](#), [Zoia Ceaușescu](#), [Valentin Ceaușescu](#)
Siblings: [Nicolae Andruță Ceaușescu](#), [Ilie Ceaușescu](#), [MORE](#)
Parents: [Andruță Ceaușescu](#), [Alexandrina Ceaușescu](#)
Height: 1.68 m

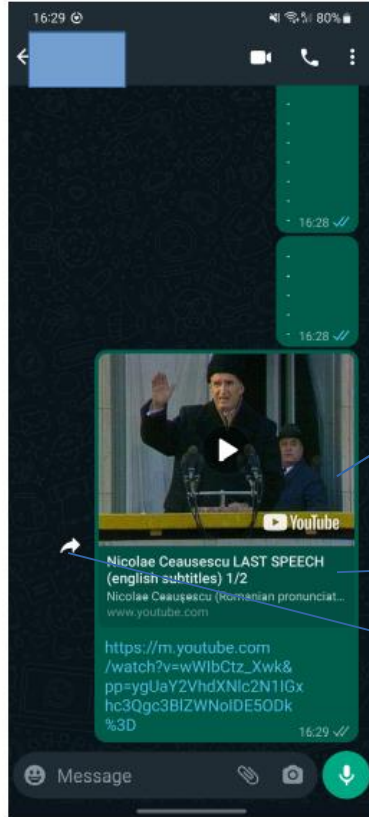
People also search for
[Elena Ceaușescu](#) [Nicu Ceaușescu](#) [Zoia Ceaușescu](#) [Ion Iliescu](#)

Knowledge Box

Links to related people (Linked Data)

Pictures of related people (Metadata)

Motivating examples (2/3) - Whatsapp



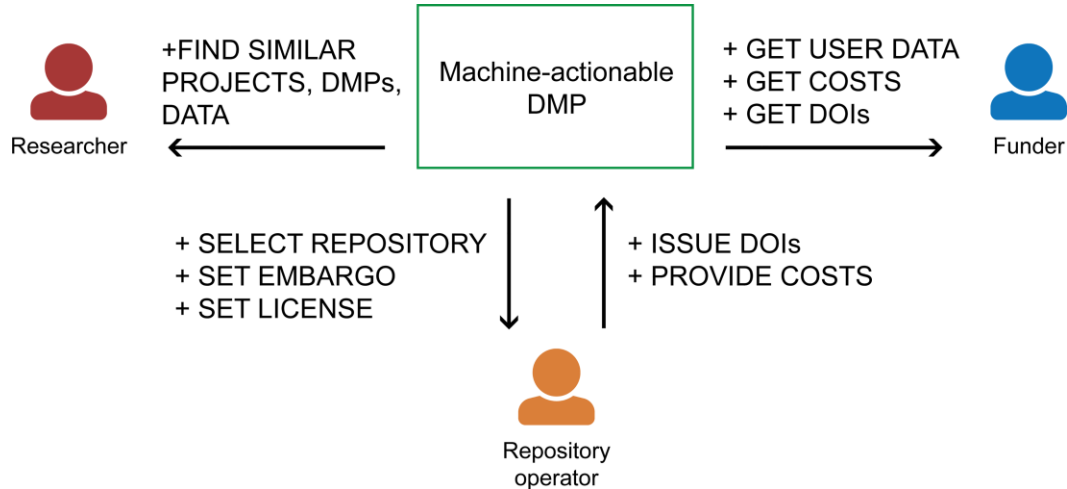
Playable (Actionable)
video

Metadata (title,
description)

Shareable (Actionable)
video

Motivating examples (3/3)

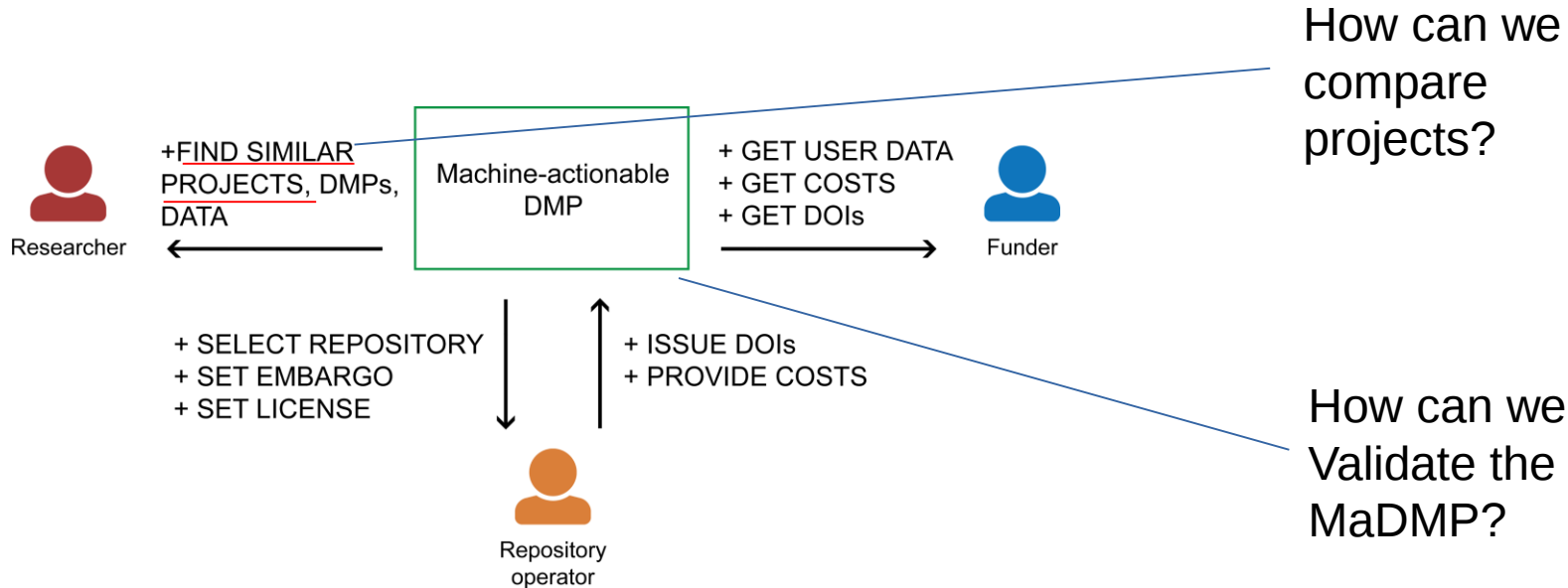
– machine-actionable Data management plans



[2] Ten principles for machine-actionable data management plans

Motivating examples (3/3)

– machine-actionable Data management plans



[2] Ten principles for machine-actionable data management plans

E.g against the DMP common standard?

How do we enhance the WWW
so that the data shared across
it becomes meaningful,
actionable and FAIR?

One efficient way is to use
Semantic Web technologies and
principles

Semantic web – main principles

Principle ^[3]	Example
Everything can be identified by URI's	Especially nowadays: persistent URIs (e.g. DOIs)
Resources and links can have types	RDF and ontologies
Partial information is tolerated	Resources that cease to exist should not stop the web from working
There is no need for absolute truth	But: We need trust (provenance)!
Evolution is supported	Timestamped-based versioning (see later)
Minimalist design	light-weight ontologies (see later)

Resource Description Framework (RDF)

- Model for data interchange on the web
- W3C standard
- Framework for describing and linking resources
- Assigning types to resources
- Resources and their relations/links represented in a graph

Subject

predicat
e

objec
t

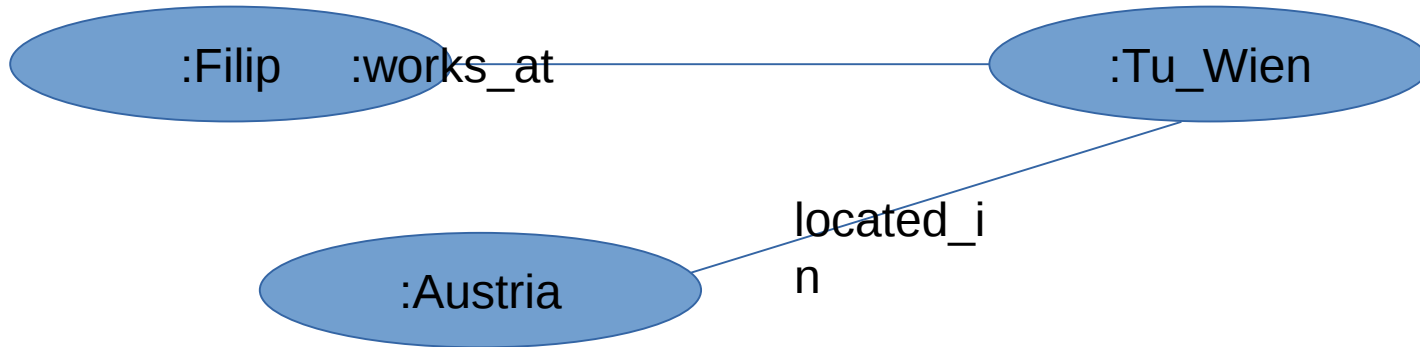
[<http://example.com/Filip>](http://example.com/Filip) [<http://example.com#works_at>](http://example.com#works_at) [<http://example.com/TU_Wien>](http://example.com/TU_Wien)

Internationalized URIs
(=IRIs)

RDF – Graph representation



RDF – Graph representation



Subject

predicate

object

[<http://example.com/Filip>](http://example.com/Filip) [<http://example.com#works_at>](http://example.com#works_at) [<http://example.com/TU_Wien>](http://example.com/TU_Wien)
 [<http://example.com/TU_WIEN>](http://example.com/TU_WIEN) [<http://example.com#located_in>](http://example.com#located_in) [<http://example.com/Austria>](http://example.com/Austria)

RDF practical example (ntriples format)

@prefix og: <http://ogp.me/ns#> .

@prefix video: <http://ogp.me/ns/video#> .

@prefix youtube: <https://www.youtube.com/> .

@prefix caucescu_video: <https://www.youtube.com/watch?v=wWlBctz_Xwk&ab_channel=AvocatulPoporului> .

caucescu_video a

og:video .

caucescu_video og:title

"Nicolae Ceausescu LAST SPEECH (english subtitles) 1/2" .

caucescu_video og:description

"This video features the last speech of Nicolae Ceausescu" .

caucescu_video og:type

video:video .

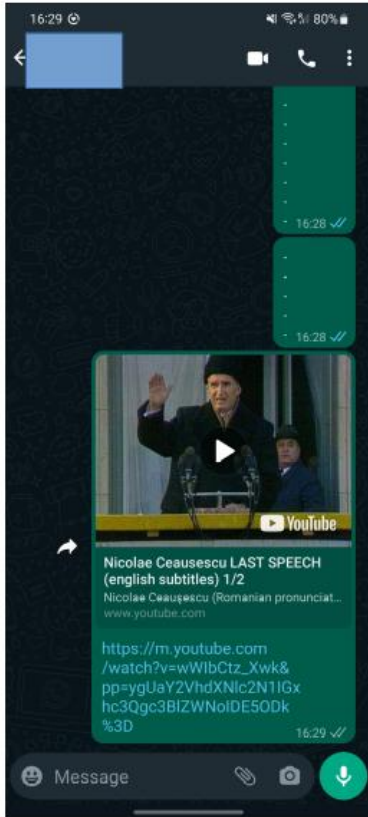
caucescu_video og:url

https://www.youtube.com/watch?v=wWlBctz_Xwk&ab_channel=AvocatulPoporului .

caucescu_video og:image

"https://someImageUrl" .

RDF practical example (xml format)



```
<meta property="og:video"
content="https://www.youtube.com/watch?v=wWlbCtz_Xwk&ab_channel=Avocatu
<meta property="og:description" content="Nicolae Ceausescu LAST
SPEECH (english subtitles) 1/2" />
<meta property="og:video:type" content="video/mp4" />
<meta property="og:image" content="http://someImage.jpg" />
<meta property="og:video:width" content="400" />
<meta property="og:video:height" content="300" />
```

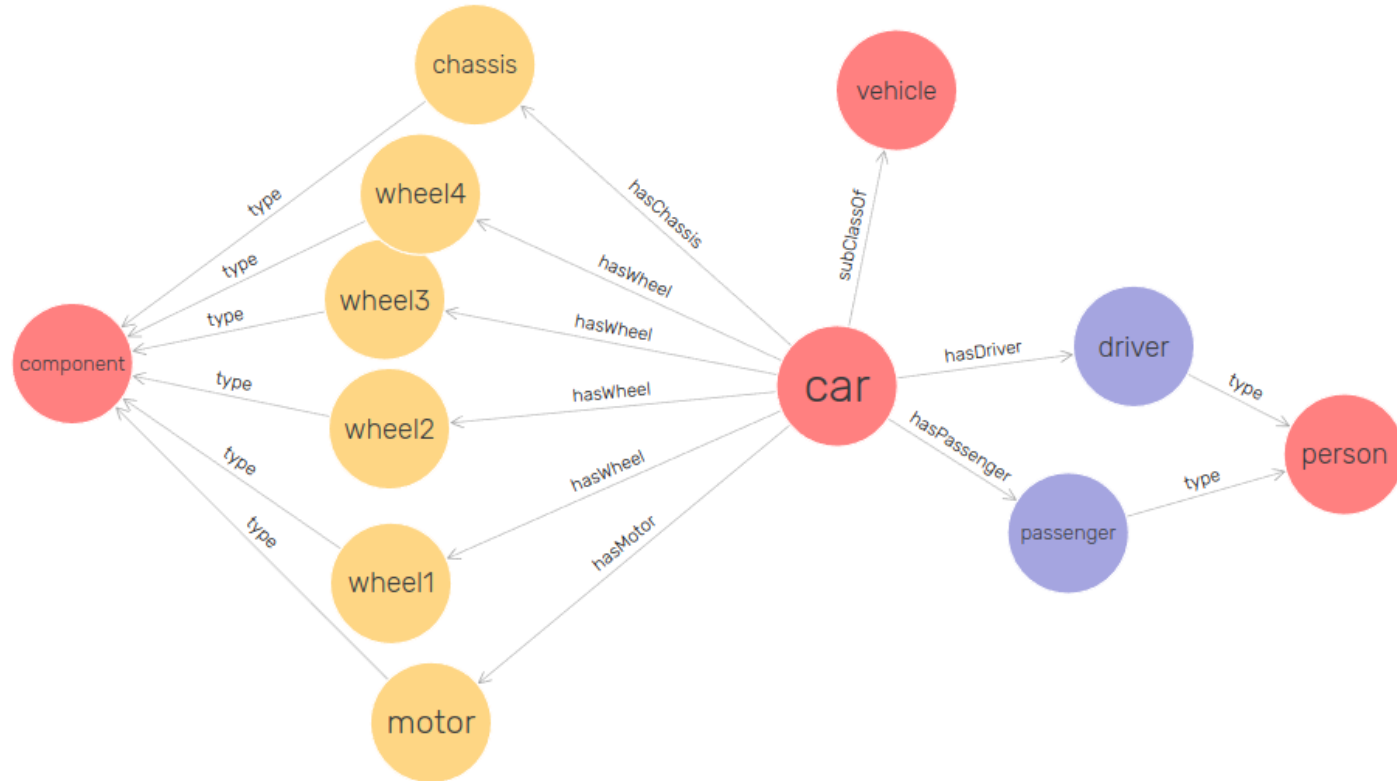
Ontologies

- Definitions
 - “Specification of a conceptualization” (Gruber, 93)
 - a representation, formal naming, and definition of the categories, properties, and relations between the concepts, data, and entities that substantiate one, many, or all domains of discourse [5]

Ontologies - Example

- A car is a vehicle that has four wheels, a chassis, a motor and other components. Moreover, a car can have a driver and other passengers.

Ontologies - Conceptualization



Ontologies - Specification

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://example.com/> .
@prefix vehicle: <http://example.com/vehicle/> .
@prefix car: <http://example.com/car/> .
@prefix component: <http://example.com/component/> .
@prefix person: <http://example.com/person/> .
```

Concepts

```
vehicle: a rdf:Class .
component: a rdf:Class .
person: a rdf:Class .
car: a rdf:Class ;
    rdfs:subClassOf vehicle: .
```

Properties

```
ex:hasDriver a rdf:Property, owl:FunctionalProperty ;
    rdfs:domain car: ;
    rdfs:range person: .
ex:hasPassenger a rdf:Property ;
    rdfs:domain car: ;
    rdfs:range person: .
```

Components

```
component:wheel1 a component: .
component:wheel2 a component: .
```

...

...

Light-weight vs heavy weight ontologies

Light-weight/upper ontologies

- Common agreement between many communities
- Have a small set of attributes
- Domain-independent: e.g. Schema.org, The Open Graph protocol
- In research domain: BFO, OBO (biology)

Heavy weight ontologies

- Usually domain specific

e.g the Chemical Functional Ontology (ChemFOnT) in chemistry [6]

515,332 terms or definitions

10.3 million chemical-functional relationships

Can also foster open science!

Scientific and scholarly ontologies

- Basic formal ontology (BFO)
- Discourse Element Ontology (DEO)
- Computer Science Ontology (CSO)
- Machine Learning Schema (MLS)
- DataCite Metadata Schema
- Vocabulary of Interlinked Datasets
- Semantic Publishing and Referencing Ontology (SPAR)
- ...

Scientific and scholarly ontologies – e.g DEO

Classes: Motivation, Problem statement, Methods, Model, Results, Acknowledgments, Prologue, ...

Object properties: hasPart, has relation, is part of

Scientific and scholarly ontologies – e.g DEO

Classes: Motivation, Problem statement, Methods, Model, Results, Acknowledgments, Prologue, ...

Object properties: hasPart, has relation, is part of

Tool for FAIR research results!

Comparison of research results - example


[View](#)
[Tools](#)
[About](#)
[NFDI4DataScience](#)


[+ Add new](#)
[Sign in](#)

Diverse approaches for the reproducibility of scientific work

January 2023

The comparison compare different approaches and tools for improving the reproducibility of the scientific work

DOI: <https://doi.org/10.48366/r285347>



Properties	An Invitation to Teaching Reproducible Research: Lessons from a Symposium <i>Contribution 1</i>	Important Papers in the History of Document Preparation Systems: Basic Sources <i>Contribution 1 - 1992</i>	Facilitating reproducible research through direct connection of data analysis with manuscript preparation: StatTag for connecting statistical software to Microsoft Word <i>Contribution 1</i>	Open Science Framework (OSF) <i>Contribution 1 - 2017</i>	Data reuse and the open data citation advantage <i>Contribution 1 - 2013</i>	Using Docker Containers to Improve Reproducibility in Software Engineering Research <i>Contribution 1 - 2016</i>	Binder 2.0 - Reproducible, interactive, sharable environments for science at scale <i>Contribution 1 - 2018</i>
approach	Sharing data and code	Documenting methodology	Documenting methodology	data management	version control	using containers	Interactive and Reproducible Environment (IRE)
				Pre-registration of studies			
				Sharing data and code			
				version control			
approach/version control/same as*				https://www.wikidata.org/entity/Q189439	https://www.wikidata.org/entity/Q189439		
reference url	https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2099489	https://www.wiumlie.no/2006/p/hd/archive/cajun.cs.nott.ac.uk/c ompsci/epo/papers/volume5/iss ue1/ep057rf.pdf	https://pubmed.ncbi.nlm.nih.gov/33215069/	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5370619/	https://peerj.com/articles/175/	https://dl.acm.org/doi/abs/10.1145/2889160.2891057?casa_token=3QCzgTnnWx4AAA:AA:VNMgCBkK8Kcl629n2cbC3HFBCybkI22UD3H-jPC_seyXG7ASUQIP4GMNxOosJwXFGj2oksxBzEgpb8E	https://pdfs.semanticscholar.org/c043/bef741a9616d1144e0205ac21ceae881485d.pdf
research problem	Computational Reproducibility	computer-based document preparation systems	Computational Reproducibility	research workflow	provenance of research findings	software quality in science	create interactive versions of repositories that exist on sites like GitHub

Linked Open Data

Linked Data (Wikipedia): In computing, linked data is structured data which is interlinked with other data so it becomes more useful through semantic queries.

Builds upon standard Web technologies: HTTP, RDF and URIs

Not only for human readers but for computers!

Should be accessible through open standards: e.g. SPARQL

Linked Open Data

Linked Data (Wikipedia): In computing, linked data is structured data which is interlinked with other data so it becomes more useful through semantic queries.

Builds upon standard Web technologies: HTTP, RDF and URIs

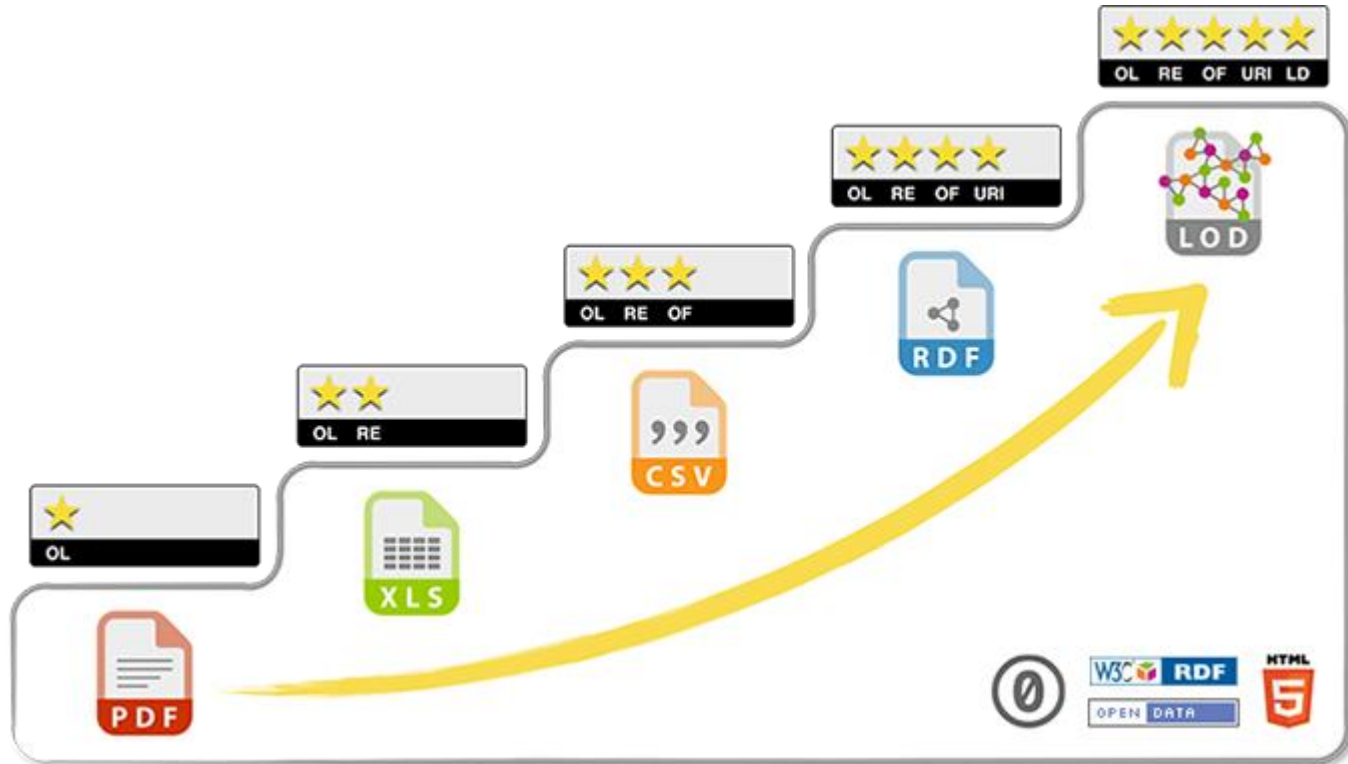
Not only for human readers but for computers!

Should be accessible through open standards: e.g. SPARQL

Linked Open Data (LOD) is Linked Data which is released under an open license, which does not impede its reuse for free [7].

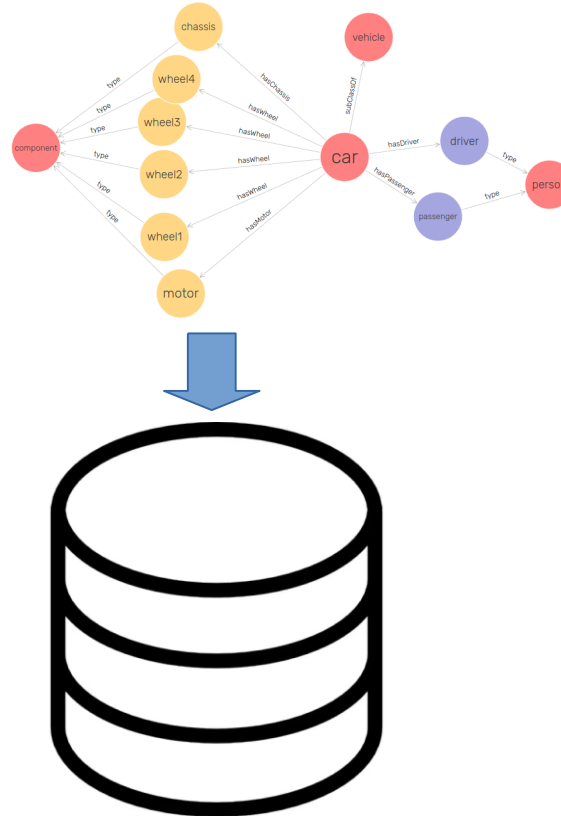
[7] https://en.wikipedia.org/wiki/Linked_data

Linked Open Data – 5 star model



RDF triple stores and graph databases

- AllgegroGraph
- Neo4J
- GraphDB
- Virtuoso
- Oracle RDF
- Jena
- ...



Should be accessible
over a SPARQL endpoint
via HTTP

SPARQL query language

- Query language for RDF
- Operations like JOINS, UNION, projections, filter, grouping, ... supported
- Supports update operations (insert and delete)

SPARQL query language - example

- What components does my car have?

```
select ?component where {
  ?component ?p <http://example.com/component/> .
}
```

Result

	Component
1	component:wheel1
2	component:wheel2
3	component:wheel3
4	component:wheel4
5	component:chassis
6	component:motor

Resources

- [1] [https://medium.com/@yassine.hammar1/inight-into-semantic-web-and-why-its-the-next-technological-revolution-24521cec4459#:~:text=The%20greatest%20advantage%20of%20the,Wide%20Web%20Consortium%20\(W3C\).](https://medium.com/@yassine.hammar1/inight-into-semantic-web-and-why-its-the-next-technological-revolution-24521cec4459#:~:text=The%20greatest%20advantage%20of%20the,Wide%20Web%20Consortium%20(W3C).)
- [2] <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006750>
- [3] <https://www.w3.org/2001/12/semweb-fin/w3csw>
- [4] <https://sparontologies.github.io/deo/current/deo.html>
- [5] [https://en.wikipedia.org/wiki/Ontology_\(computer_science\)](https://en.wikipedia.org/wiki/Ontology_(computer_science))
- [6] <https://www.chemfont.ca/>
- [7] https://en.wikipedia.org/wiki/Linked_data#Linked_open_data
- [8] <https://5stardata.info/en/>



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna | Austria

StarVers - Timestamped- based versioning of RDF datasets

Filip Kovacevic
28.03.2022

- Versioning RDF datasets
- RDF-star and SPARQL-star
- Starvers
- Live Demo

Versioning RDF datasets

- External approaches
 - Tools like Git and SVN
 - Relational databases
- Different archiving mechanisms [1]
 - Full snapshots
 - Change-based versioning
 - **Timestamp-based versioning**
- Native RDF approaches to assign statement-level metadata [2],[3]
 - Reification
 - Named graphs
 - Singleton
 - **RDF-star**

- Materialize snapshots of your datasets

Snapshot released at 2022-01-01

Name	Occupation
Nguyen Van Anh	Tour guide
Tran Thi Thu	Tour guide
Le Van Thanh	Travel blogger

Different archiving mechanisms - snapshot

- Materialize snapshots of your datasets

Snapshot released at 2022-01-01

Name	Occupation
Nguyen Van Anh	Tour guide
Tran Thi Thu	Tour guide
Le Van Thanh	Travel blogger

Snapshot released at 2022-06-01

Name	Occupation
Nguyen Van Anh	Owner of travel agency
Tran Thi Thu	Tour guide
Le Van Thanh	Travel blogger

- Advantages
 - Retrieving data is fast
- Problems
 - Storage-intensive
 - Information gap between snapshots

Different archiving mechanisms - change-based

- Materialize only initial snapshot and then record only changes

Snapshot released at 2022-01-01 Change set released at 2022-03-01 Change set released at 2022-06-01

Name	Occupation
Nguyen Van Anh	Tour guide
Tran Thi Thu	Tour guide
Le Van Thanh	Travel blogger

Name	Occupation
Nguyen Van Anh	Hotel manager

Name	Occupation
Nguyen Van Anh	Owner of travel agency

- Advantages: Less storage-intensive compared to snapshots
- Problems
 - Doesn't scale well if there are too many change sets
 - Patches need to be consecutively applied starting from the initial snapshots
 - Alternative approach: Computing “in-between” snapshots
 - Potential information gap if changes are consolidated

Different archiving mechanisms - timestamp-based

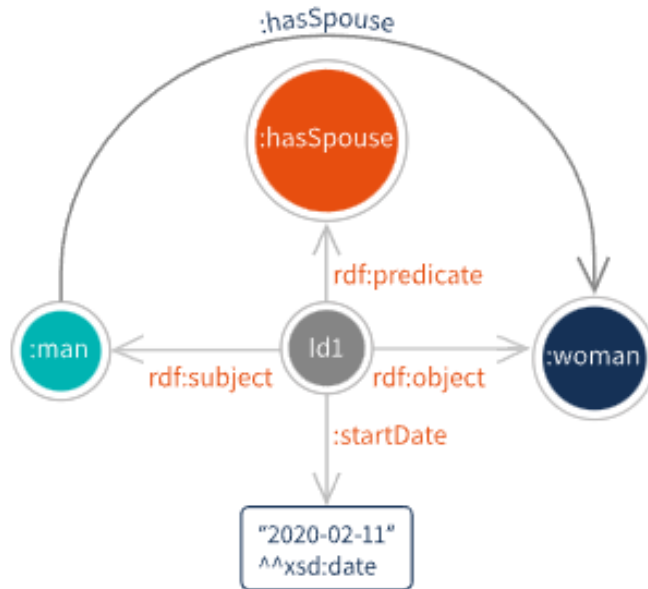
- Assign temporal metadata to every record
- Enables retrieving a subset as it was at a specific point in time

Name	Occupation	valid_from	valid_until
Nguyen Van Anh	Tour guide	2000-01-01T12:00:00	2010-01-01T12:00:00
Tran Thi Thu	Tour guide	2000-01-01T12:00:00	9999-12-31T00:00:00
Le Van Thanh	Travel blogger	2000-01-01T12:00:00	9999-12-31T00:00:00
Nguyen Van Anh	Hotel manager	2010-01-01T12:00:00	2020-01-01T12:00:00
Nguyen Van Anh	Owner of a travel agency	2020-01-01T12:00:00	9999-12-31T00:00:00

- Advantages
 - Scales well with high number of changes
 - Query performance can be comparable to the snapshot approach, if there is a good implementation of temporal indexes
- Problems
 - Storing indexes with additional temporal information might induce high space overhead, if no compression mechanisms are available [5]

Native RDF approaches to assign statement-level metadata - Standard reification

- Create new resource and link it to all entities of the triple
- Assign metadata to that new resource.



```

:man :hasSpouse :woman .
:id1 rdf:type rdf:Statement ;
    rdf:subject :man ;
    rdf:predicate :hasSpouse ;
    rdf:object :woman ;
    :startDate "2020-02-11"^^xsd:date .
  
```

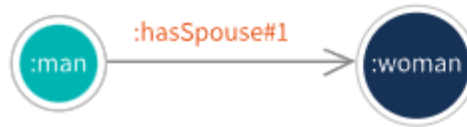
Native RDF approaches to assign statement-level metadata - Singleton

- Make unique identifiers out of every property
- Assign metadata to that property



```
:man :hasSpouse#1 :woman .
```

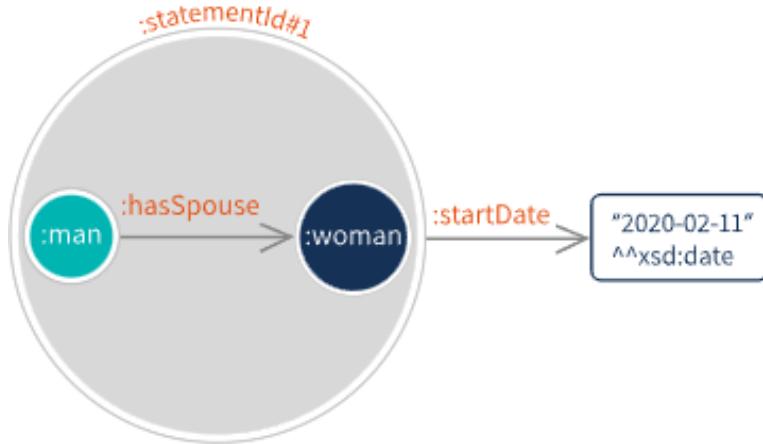
```
:hasSpouse#1 :startDate "2020-02-11"^^xsd:date .
```



Native RDF approaches to assign statement-level metadata - Named graphs

- Link every triple to a fourth entity, the named graph resource
- Assign metadata to that named graph

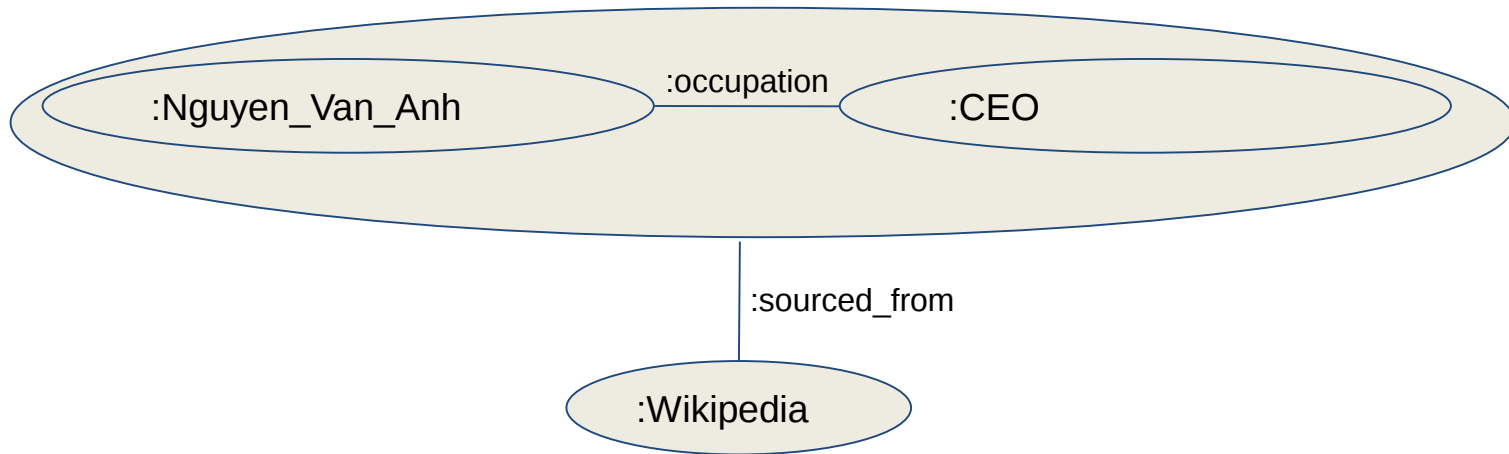
```
:man :hasSpouse :woman :singletonGraph#1 .
:singletonGraph#1 :startDate "2020-02-11"^^xsd:date :metadata .
```



RDF-star and SPARQL star

- “Quoted triples” approach
- Data model with the goal to make statements about triples

<< :Nguyen_Van_Anh :occupation :CEO >> :sourced_from :Wikipedia .



- Querying nested triples with SPARQL-star

```
Select ?s ?p ?o {
  <<?s ?p ?o >> :sourced_from :Wikipedia .
}
```

Result set

s	p	o
:Nguyen_Van_Anh	:occupation	:CEO

- Starvers is a [python module](#) for timestamp-based versioning of RDF data as RDF-star dataset. [6]
- Enables the evolution management of knowledge graphs and ontologies
- Arbitrary snapshots and subsets can be queried as they were at a specific point in time

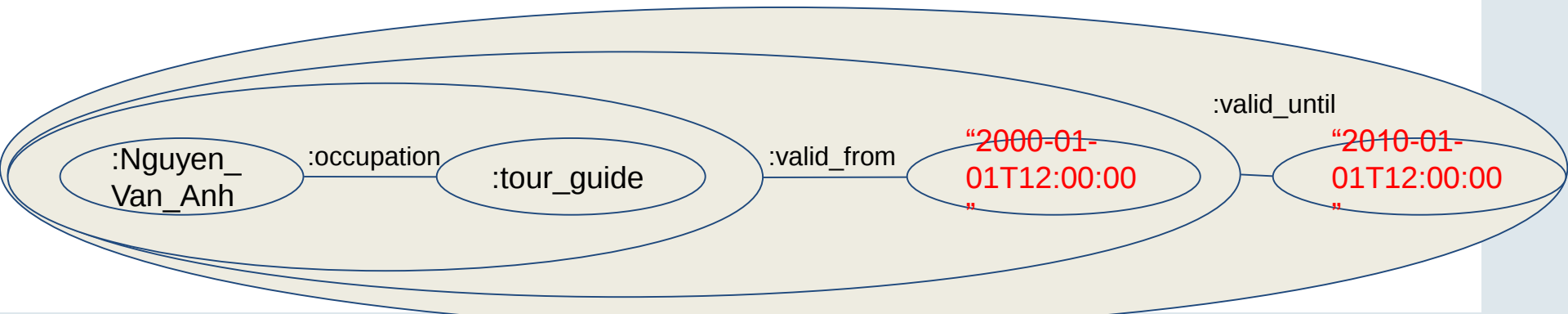
- Only needs two SPARQL endpoints to an RDF-star store (e.g. GraphDB) as configuration
- Features
 - Version all rows/Convert RDF to RDF star dataset
 - Insert
 - Update
 - “Outdate”
 - Query live or historical data

- Submitted to [Semantic Web Journal](#) [4]
- The revised version is going to be soon submitted soon

Timestamped-based versioning with RDF-star - The StarVers approach

Nested quoted triples, i.e. two nesting levels

```
<< << :Nguyen_Van_Anh :occupation :tour_guide >> :valid_from "2000-01-01T12:00:00" >>
valid_until "2010-01-01T12:00:00" >> .
<< << :Nguyen_Van_Anh :occupation :hotel_manager >> :valid_from "2010-01-01T12:00:00"
>> valid_until "2020-01-01T12:00:00" >> .
<< << :Nguyen_Van_Anh :occupation :CEO >> :valid_from "2020-01-01T12:00:00" >>
valid_until "9999-12-31T00:00:00" >> .
```

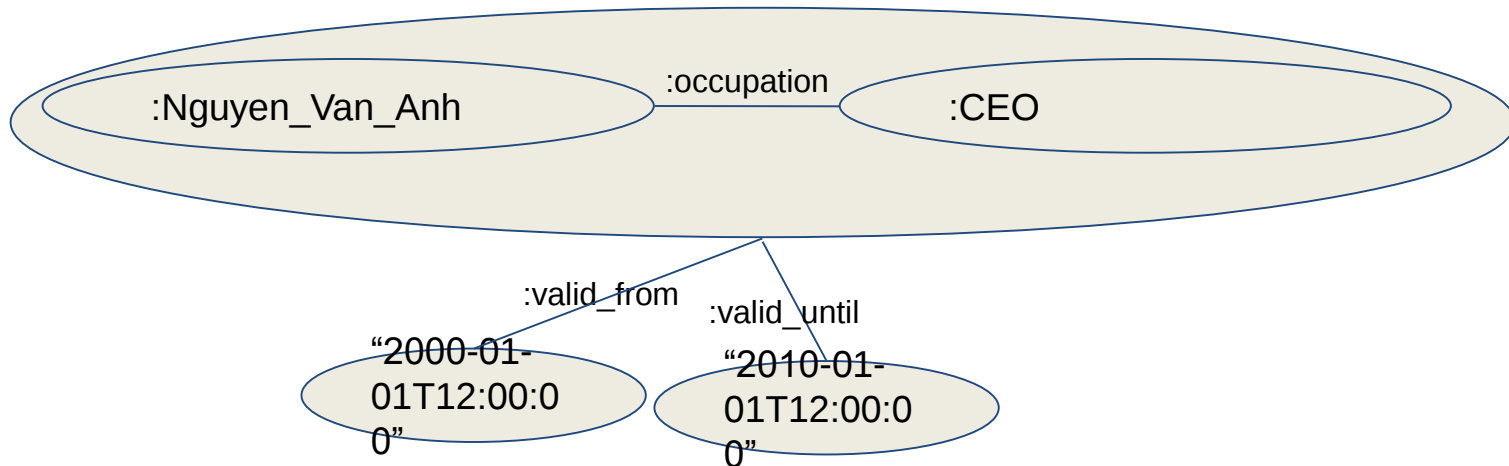


Excursion - Why not quoted triples approach?

- “Quoted triples” approach

<< :Nguyen_Van_Anh :occupation :CEO >> :valid_from “2000-01-01T12:00:00” .

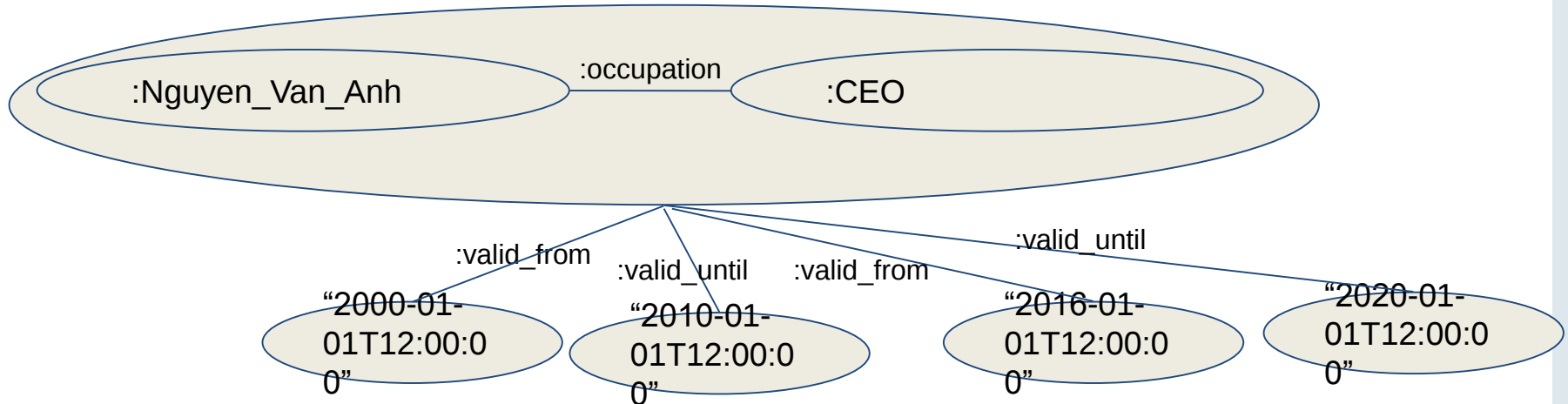
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_until “2010-01-01T12:00:00” .



Excursion - Why not quoted triples approach?

- “Quoted triples” approach

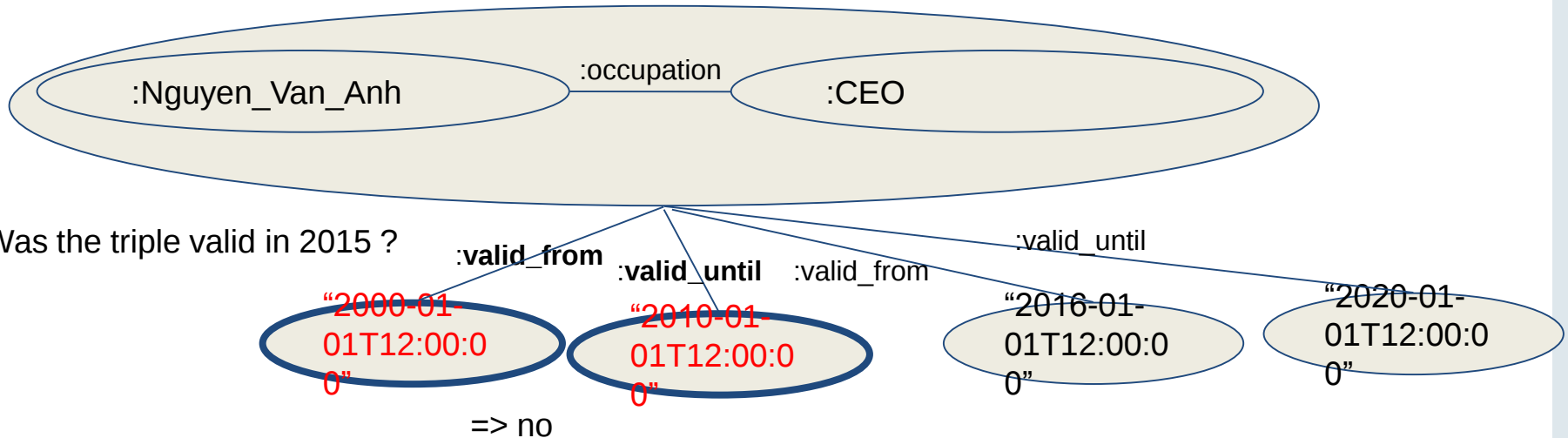
```
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_from "2000-01-01T12:00:00" .
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_until "2010-01-01T12:00:00" .
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_from "2016-01-01T12:00:00" .
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_until "2020-01-01T12:00:00" .
```



Excursion - Why not quoted triples approach?

- “Quoted triples” approach

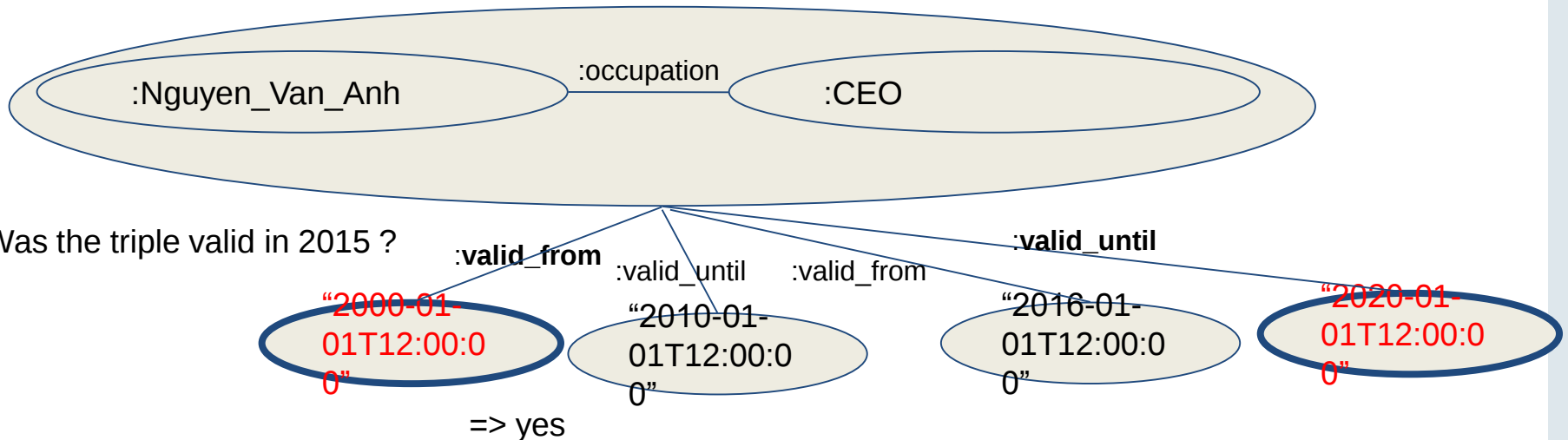
```
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_from "2000-01-01T12:00:00" .
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_until "2010-01-01T12:00:00" .
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_from "2016-01-01T12:00:00" .
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_until "2020-01-01T12:00:00" .
```



Excursion - Why not quoted triples approach?

- “Quoted triples” approach

```
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_from "2000-01-01T12:00:00" .
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_until "2010-01-01T12:00:00" .
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_from "2016-01-01T12:00:00" .
<< :Nguyen_Van_Anh :occupation :CEO >> :valid_until "2020-01-01T12:00:00" .
```



Live demo

- [1] Papakonstantinou, Vassilis, et al. "Versioning for Linked Data: Archiving Systems and Benchmarks." *BLINK@ ISWC* 1700 (2016).
- [2] Hernández, Daniel, Aidan Hogan, and Markus Krötzsch. "Reifying RDF: What works well with wikidata?." *SSWS@ ISWC* 1457 (2015): 32-47.
- [3] Hartig, Olaf. "Foundations of RDF* and SPARQL*:(An alternative approach to statement-level metadata in RDF)." *AMW 2017 11th Alberto Mendelzon International Workshop on Foundations of Data Management and the Web, Montevideo, Uruguay, June 7-9, 2017..* Vol. 1912. Juan Reutter, Divesh Srivastava, 2017.
- [4] RDF-star first paper version: <http://semantic-web-journal.org/content/starvers-versioning-and-timestamping-rdf-data-means-rdf-approach-based-annotated-triples>
- [5] Cerdeira-Pena, Ana, et al. "Self-indexing RDF archives." *2016 Data Compression Conference (DCC)*. IEEE, 2016.
- [6] <https://github.com/GreenfishK/starvers>