Introduction
From “Adam and Eve”

Henrik Askjer
University of Bergen
Norway

The Semantic Web

- 1989: Tim Berners-Lees proposal for the World Wide Web
  - A web of links between nodes representing not only documents but people, projects, concepts etc.
- A global database

Resource Description Framework (RDF)

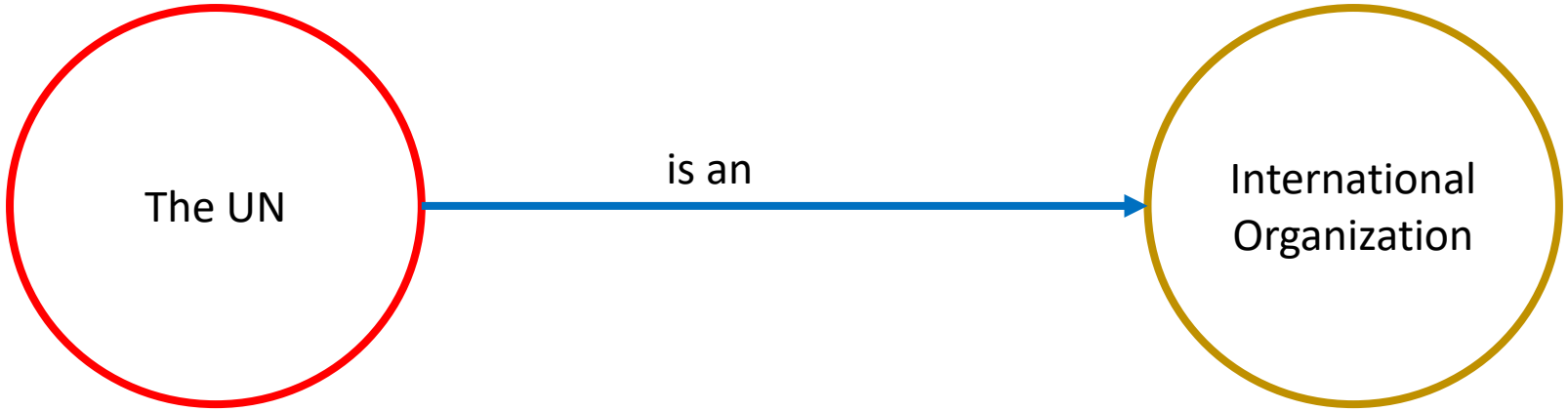
- Standard for metadata on the web developed by W3C in the late 1990s
- A framework for both **instance data** and **ontologies**. The latter facilitated by extensions to RDF such as Web Ontology Language (OWL)

The relational model in traditional databases: link rows using identifiers that are unique within each table.

The RDF-model: a graph composed of triple statements of identifiers that are unique across all databases

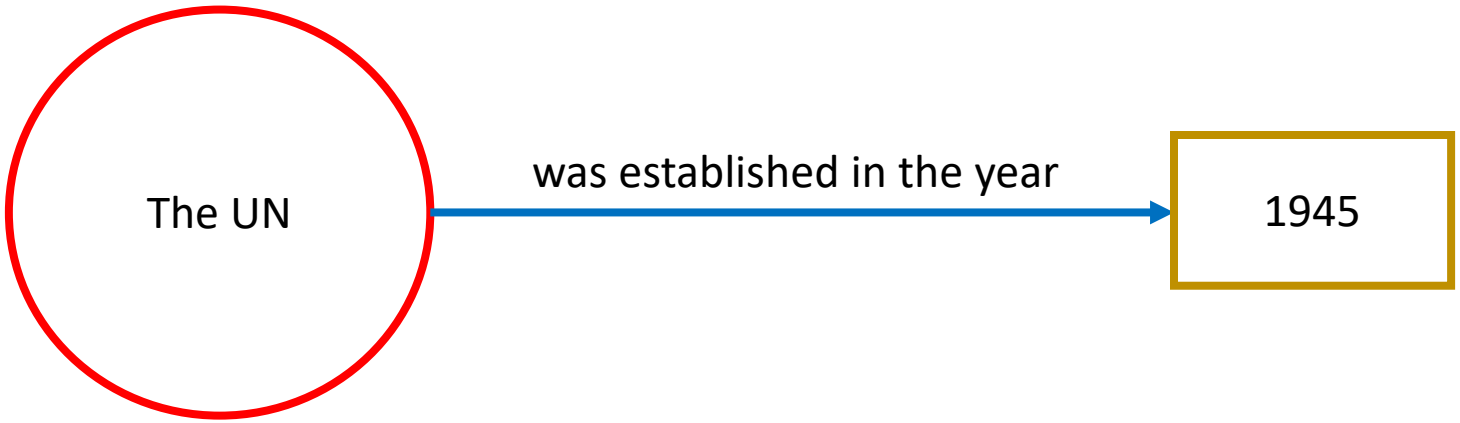
Triples

<http://wikidata.org/entity/Q1065> <http://example.org/instanceOf> <http://wikidata.org/wiki/Q484652> .



Triples

<http://wikidata.org/entity/Q1065> <http://example.org/establishedYear> “1945” .

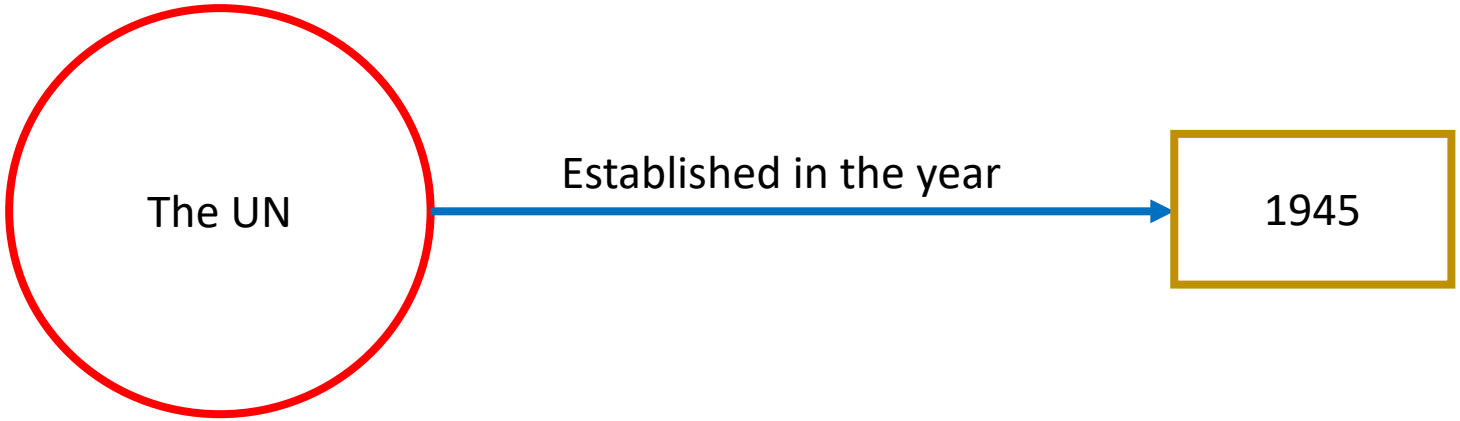


Triples

<http://wikidata.org/entity/Q1065> <http://example.org/establishedYear> “1945” .

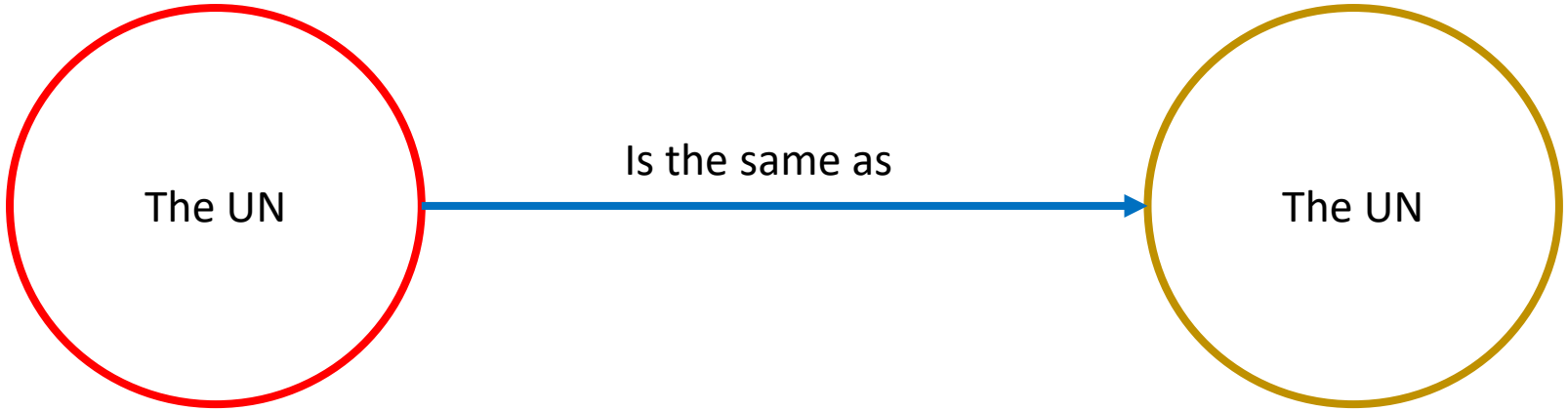
<http://wikidata.org/entity/Q1065> <http://www.w3.org/2000/01/rdf-schema#label> “The UN” .

<http://example.org/establishedYear> <http://www.w3.org/2000/01/rdf-schema#label> “Established in the year” .



Triples

<http://wikidata.org/entity/Q1065> <http://www.w3.org/2002/07/owl#sameAs> <http://viaf.org/viaf/140413306> .



Linked data

Coined by Tim Berners-Lee in a design note from 2006

1. Use Uniform Resource Identifiers (URI/IRIs) as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information using open standards such as RDF
4. Include links to other URIs. so that they can discover more things.

SPARQL

- RDF query language, similar to SQL
- Created by W3C in 2008
- SPARQL-endpoints allow developers/users to query the database directly without requesting a data dump
- Easier to write complex queries

Triplestores

- Graph databases purpose built for linked data
- Challenges
  - Scalability
  - Lack of expertise
- Advantages
  - Flexibility: Add heterogenous data without having to modify the schema/structure of the database
  - Complex joins in SPARQL are faster than joins in SQL
  - Reasoning: infer new facts from existing data

Linked open data

- Some challenges
  - Privacy
  - Copyright/Accreditation – web 2.0
  - Mainly adopted in academia and the public sector
    - Lack of mainstream software
- Some advantages
  - URIs/IRIs make it easier to combine data from different sources
  - Self documenting data
  - Promote a more open web

Where to start?

- Standardize authority data, and to some extent choice of ontology
- Convert and publish your data in a RDF-format
- Provide your own resources or authority data with IRIs that point to a landing page or an RDF file on the web

Summary

- **The semantic web**: an ambitious vision of a global database
- **RDF**: a standard that formalized representing data as triple statements of URIs
- **Linked data**: principles for how to publish data on the web, or data that fulfill these criteria. A pragmatic approach to the semantic web?
- **SPARQL and Triplestores**: technologies that realize some of the potential of linked data

Thank you for listening

United Nations Group of Experts
On Geographical Names (UNGEGN)

Promoting the collection, standardization and dissemination of geographical names