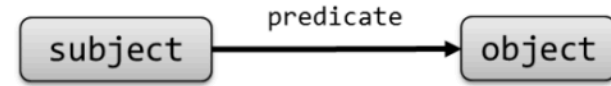




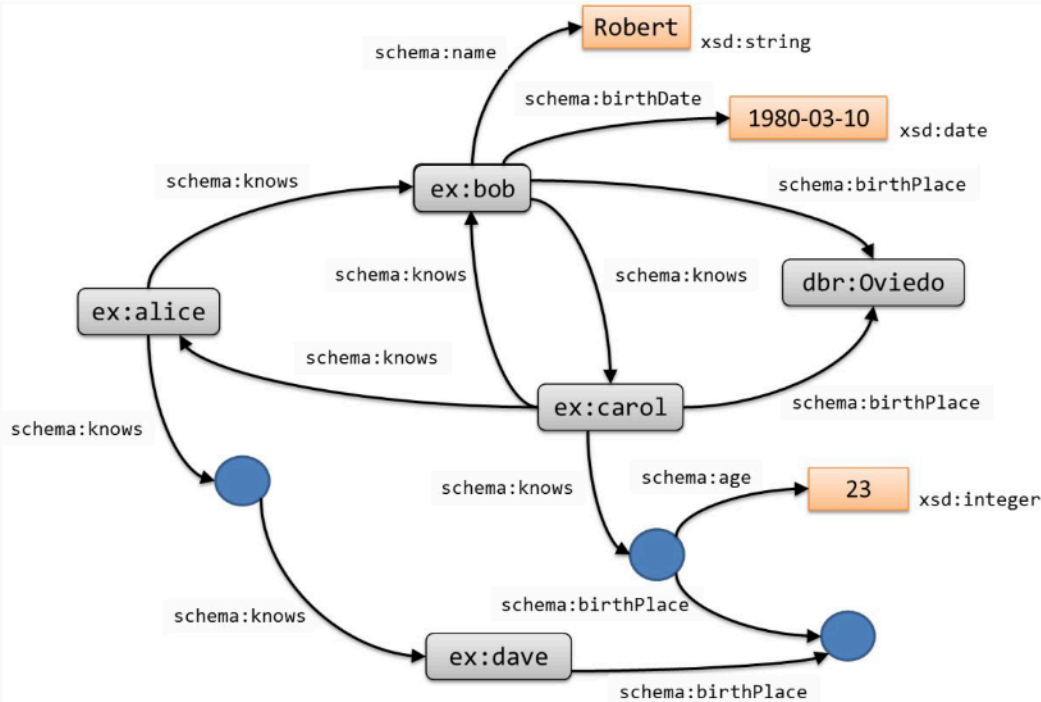
# Metadata models and ontologies for knowledge representation

*Jouni Tuominen*

# Graph-based data model (RDF)



**<subject> <predicate> <object>**



**ex:alice schema:knows ex:bob**

**ex:bob schema:birthPlace dbr:Oviedo**

# Internationalized Resource Identifier (IRI)

- Globally unique identifier
- Enables references to shared metadata models, ontologies, data
- Can act as locator to data (URL): enabling publishing and dereferencing Linked Data on the web

- “Follow your nose” principle:

ID: <http://www.wikidata.org/entity/Q1757> →

Human: <https://www.wikidata.org/wiki/Q1757>

Machine:

<https://www.wikidata.org/wiki/Special:EntityData/Q1757.json>

# How to design a (meta)data model and ontologies for your data

- **How to structure your data?**
  - What data (fields) do you have?
  - What kind of (research) questions do you want to pose to your data?
- **How to be interoperable with other datasets?**
  - What established data models and/or ontologies do exist (on your domain)?
- **For using your data, which applications do you plan to use?**
- **“Everything should be made as simple as possible, but not simpler”**

# Cultural heritage metadata models

- **Document/artifact/object-centric metadata models**
  - E.g. Dublin Core + extensions
- **Event-centric metadata models**
  - E.g. CIDOC CRM
- **Other examples of metadata models: IFLA Library Reference Model (LRM), BIBFRAME, ...**
- **You can combine “standard” models, use parts that you need, extend as necessary**
  - When extending: map your own properties to existing ones (rdfs:subClassOf, rdfs:subPropertyOf, ...)

# Dublin Core: generic data model for information resources

# Dublin Core

- **Originates from information and library sciences, 1995–**
- **Core elements (properties) for describing information resources**

# The original Dublin Core Metadata Element Set (15 properties)

**Contributor** – “An entity responsible for making contributions to the resource.”

**Coverage** – “The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.”

**Creator** – “An entity primarily responsible for making the resource.”

**Date** – “A point or period of time associated with an event in the lifecycle of the resource.”

**Description** – “An account of the resource.”

**Format** – “The file format, physical medium, or dimensions of the resource.”

**Identifier** – “An unambiguous reference to the resource within a given context.”

**Language** – “A language of the resource.”

**Publisher** – “An entity responsible for making the resource available.”

**Relation** – “A related resource.”

**Rights** – “Information about rights held in and over the resource.”

**Source** – “A related resource from which the described resource is derived.”

**Subject** – “The topic of the resource.”

**Title** – “A name given to the resource.”

**Type** – “The nature or genre of the resource.”



# DCMI Metadata Terms

Properties in the [/terms/](#) namespace: [abstract](#), [accessRights](#), [accrualMethod](#), [accrualPeriodicity](#), [accrualPolicy](#), [alternative](#), [audience](#), [available](#), [bibliographicCitation](#), [conformsTo](#), [contributor](#), [coverage](#), [created](#), [creator](#), [date](#), [dateAccepted](#), [dateCopyrighted](#), [dateSubmitted](#), [description](#), [educationLevel](#), [extent](#), [format](#), [hasFormat](#), [hasPart](#), [hasVersion](#), [identifier](#), [instructionalMethod](#), [isFormatOf](#), [isPartOf](#), [isReferencedBy](#), [isReplacedBy](#), [isRequiredBy](#), [issued](#), [isVersionOf](#), [language](#), [license](#), [mediator](#), [medium](#), [modified](#), [provenance](#), [publisher](#), [references](#), [relation](#), [replaces](#), [requires](#), [rights](#), [rightsHolder](#), [source](#), [spatial](#), [subject](#), [tableOfContents](#), [temporal](#), [title](#), [type](#), [valid](#)

Term Name: creator <a href="#">More details</a>	
<b>URI</b>	<a href="http://purl.org/dc/terms/creator">http://purl.org/dc/terms/creator</a>
<b>Label</b>	Creator
<b>Definition</b>	An entity responsible for making the resource.
<b>Comment</b>	Recommended practice is to identify the creator with a URI. If this is not possible or feasible, a literal value that identifies the creator may be provided.
<b>Type of Term</b>	Property
<b>Range Includes</b>	<ul style="list-style-type: none"><li><a href="http://purl.org/dc/terms/Agent">http://purl.org/dc/terms/Agent</a></li></ul>
<b>Equivalent Property</b>	<ul style="list-style-type: none"><li><a href="http://xmlns.com/foaf/0.1/maker">http://xmlns.com/foaf/0.1/maker</a></li></ul>
<b>Subproperty of</b>	<ul style="list-style-type: none"><li><a href="http://purl.org/dc/elements/1.1/creator">Creator</a> (<a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a>)</li><li><a href="http://purl.org/dc/terms/contributor">Contributor</a> (<a href="http://purl.org/dc/terms/contributor">http://purl.org/dc/terms/contributor</a>)</li></ul>

<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

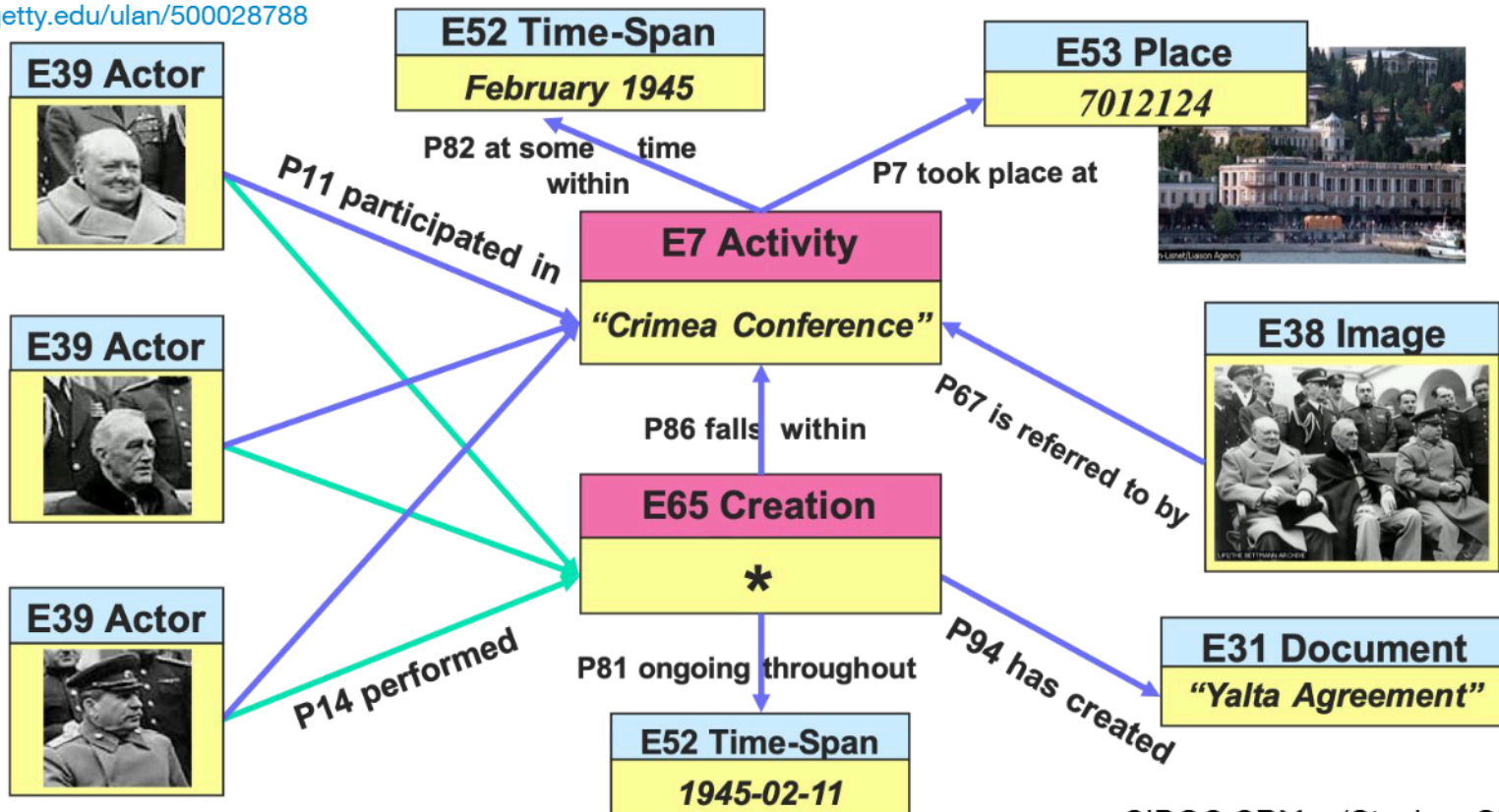
# CIDOC CRM: event-based data model for cultural heritage

# CIDOC CRM

- **Originates from the museum domain, 1990s**
- **Conceptual Reference Model: provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation**
- **Useful for information integration**

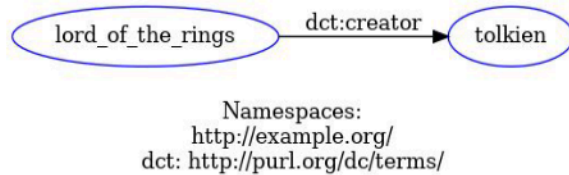
# Metadata models

<http://vocab.getty.edu/ulan/500028788>

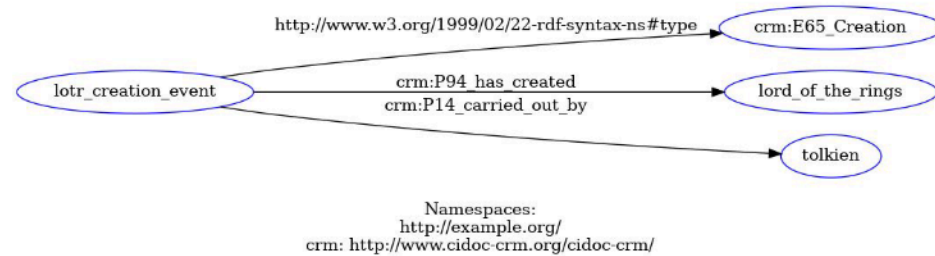


# Dublin Core vs. CIDOC CRM example: “Tolkien is the creator of The Lord of the Rings.”

## Dublin Core

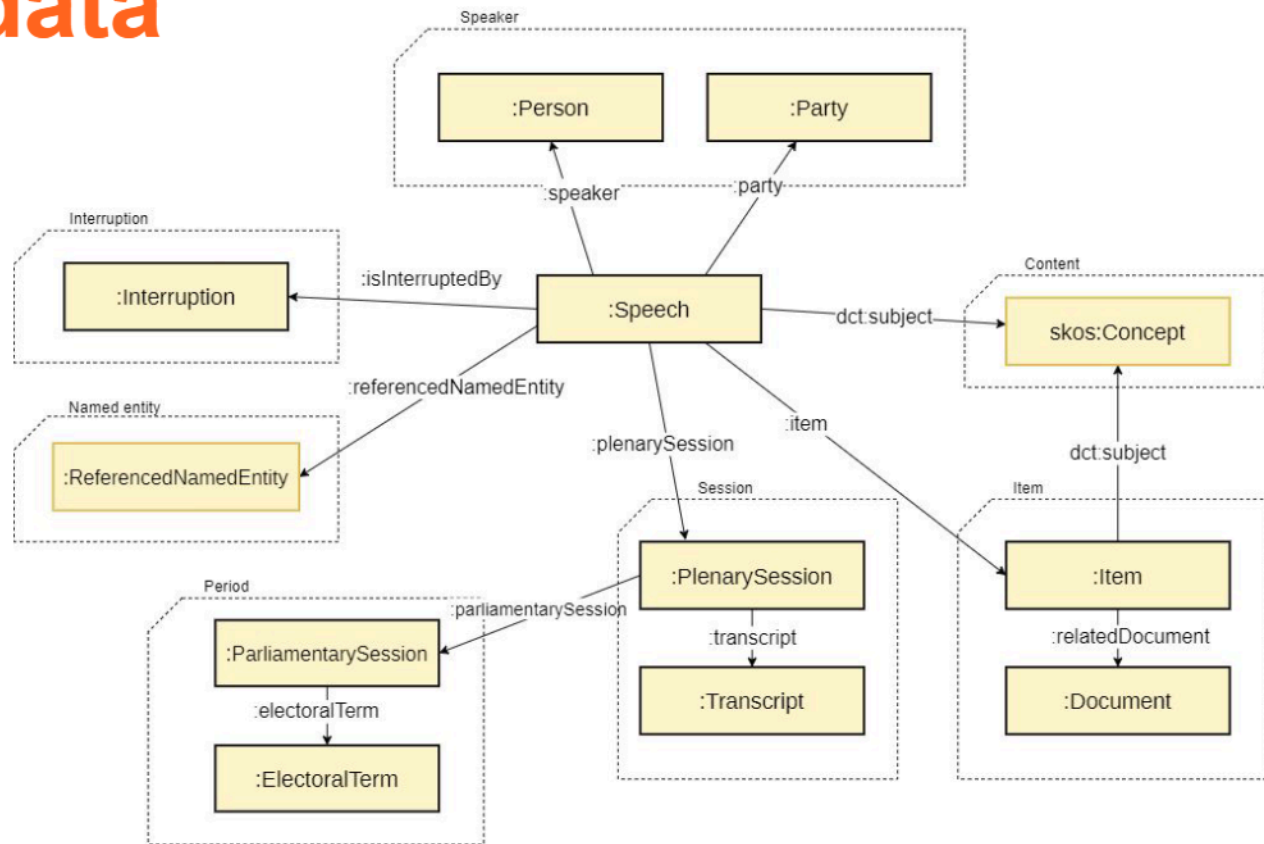


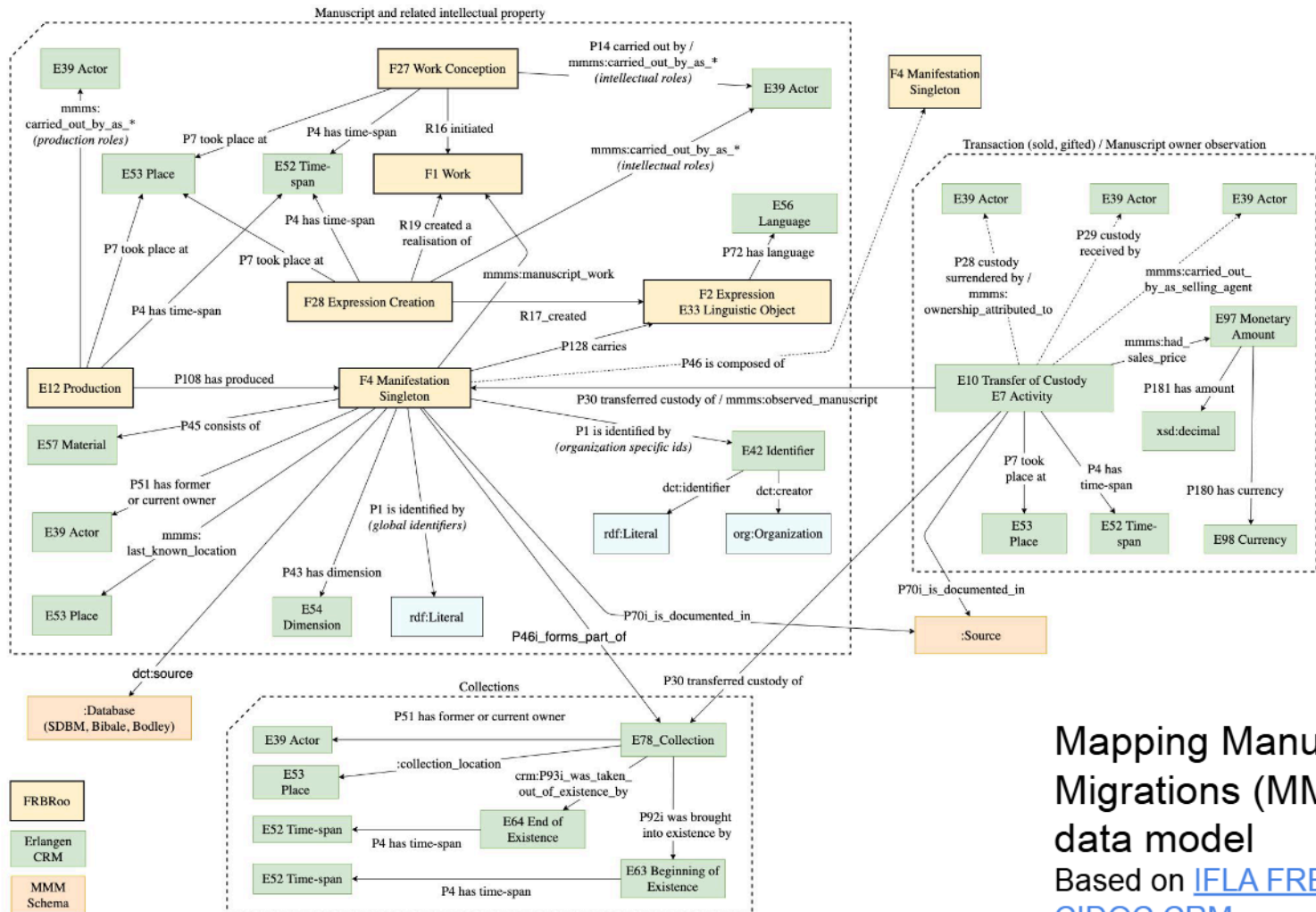
## CIDOC CRM



# Sampo data model examples

# ParliamentSampo Speech data model





Mapping Manuscript Migrations (MMM) data model  
Based on [IFLA FRBR](#) and [CIDOC CRM](#)

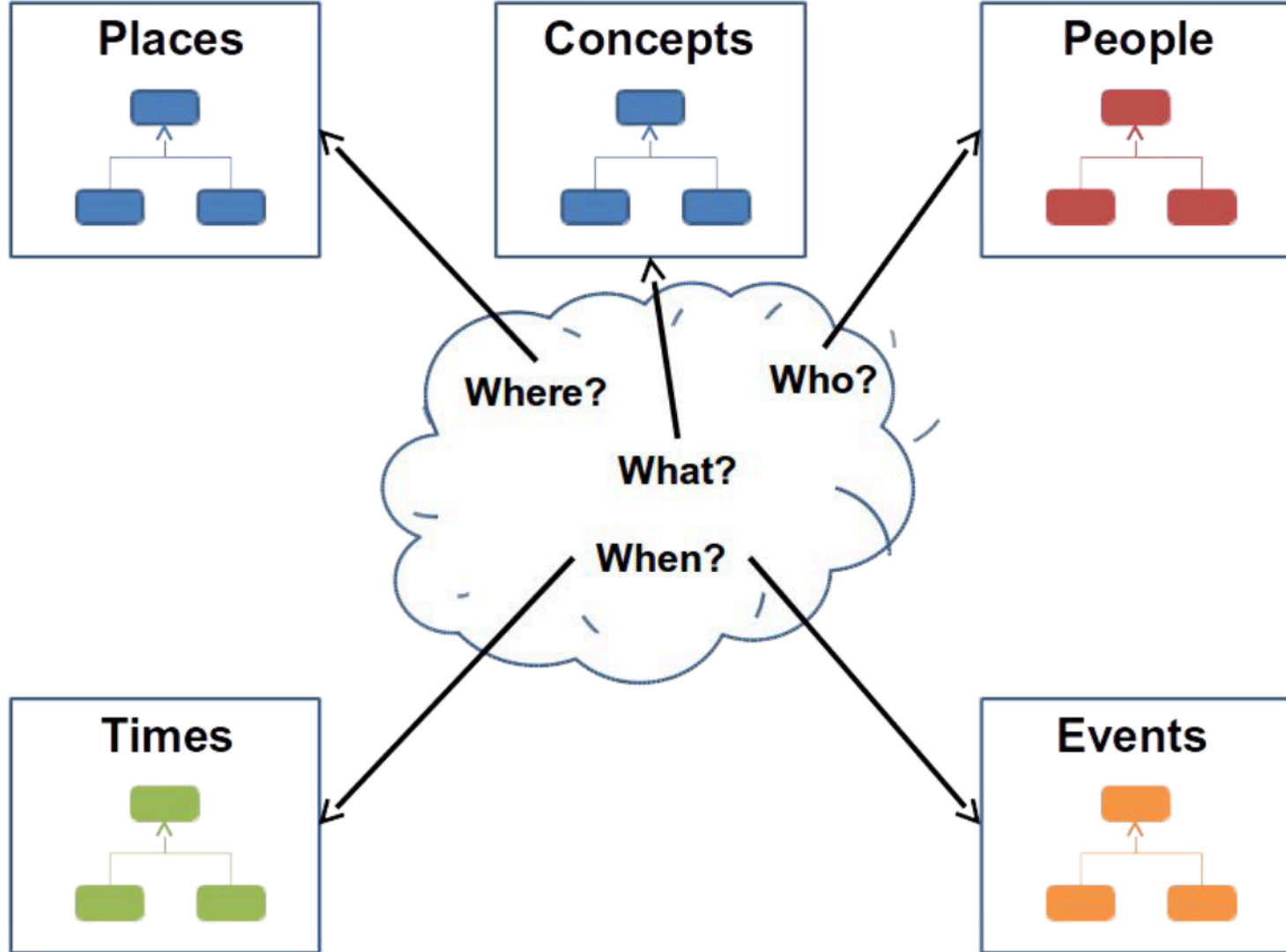


# Ontologies

# Ontology (in computer science)

- An ontology is a *formal, explicit* specification of a *shared conceptualization*.<sup>1</sup>
- For knowledge organization, information retrieval, reasoning: *knowledge organization systems (KOS)*
  - Also (meta)data models can be thought of as ontologies
- **Ontologies can be used as metadata values, e.g.:**
  - Ontology of places (e.g. GeoNames)
  - Ontology of actors (e.g. Getty ULAN)
  - Ontology of keywords/subject matter (e.g. General Finnish ontology)

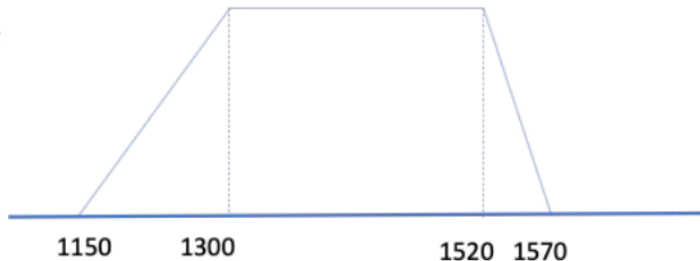
[1] Studer, R., Benjamins, V. R., and Fensel, D. Knowledge engineering: Principles and methods. *Data & Knowledge Engineering* 25, 1–2 (1998), 161–197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)



# Domain ontologies

Times and periods

Middle Age (Keskiäika) in Finland



Keyword concepts



Actors

URI: [http://dbpedia.org/resource/Pyotr\\_Ilyich\\_Tchaikovsky](http://dbpedia.org/resource/Pyotr_Ilyich_Tchaikovsky)



[Pjotr Tšaikovski](#) (fi)  
[Петр Ильич Чайковский](#) (ru)  
[Pyotr Ilyich Tchaikovsky](#) (en)  
[Pjotr Tšajkovskij](#) (sv)  
[Pjotr Tšajkovskij](#) (no)  
[Pjotr Iljitsch Tšchaikowski](#) (de)  
[Piotr Ilitch Tchaikowski](#) (fr)  
[Piotr Ilch Chaikowski](#) (es)  
[Petr Il'ič Čajkovskij](#) (it)  
[Pjotr Iljitsj Tšajkovskij](#) (nl)  
[Piotr Ilitch Tchaikowski](#) (pt)  
[Piotr Czajkowski](#) (pl)  
[Piotr Ilici Ceaikovski](#) (ro)  
[Pjotr Iljics Csaikovszkj](#) (hu)

Events

SAHA3 | historia - search

<http://agricola.utu.fi/df/phe83>

historiallinen tapahtuma: YYA-sopimus raukesi

ladi!

nimi	(fi) YYA-sopimus raukesi
tapahtumatyyppi	kahdenkeskiset kansainväliset sopimukset
kuuluu teemaan	poliittinen historia
kuka	Aho, Esko, Burbulis, Gennadi Eduardovitš (1945-) Suomi (1917-) Venäjä (1990-)
toiminta	allekirjottaminen, irtisanominen, poliittinen päätöksenteko
tulos	Suomen tasavallan ja Venäjän federaation välinen sopimus suhteiden perusteista 03/1992
tapahtuma-aika	1992-01-20
tapahtumapaikka	Helsinki(1966-2008)
asiasana	diplomatia, kansainväliset sopimukset, kansainväliset suhteet, ulkopoliittika, YYA-sopimus
liittyy	YYA-sopimus solmittiin Neuvostoliiton kanssa, YYA-sopimusta jälkeltin kahdella kymmenellä vuodella
liittyvä web-sivu	(fi) <a href="http://www.finlex.fi/fi/sopimukset/sopsteksti/1992/19920063">http://www.finlex.fi/fi/sopimukset/sopsteksti/1992/19920063</a>
syy	Neuvostoliitto lakkaufetaan
tyyppi	historiallinen tapahtuma

Places

# Data models for ontologies: SKOS, OWL

- **SKOS**: for thesauri, “lightweight” ontologies
  - Concept hierarchy, associative relations
  - “A Little Semantics Goes a Long Way”<sup>1</sup>
- **OWL**: for semantically richer ontologies
  - Class/property hierarchy, instances, property domains and ranges, property cardinalities, class disjointness, ...

[1] James Hendler. <https://www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html>

# Finto.fi ontology service (RDF, SPARQL, REST API)

Vocabularies About Feedback Help | suomeksi på svenska sámegielli

from all English Search

## finto

### Welcome to Finto.fi!

Finto.fi is a centralized service for interoperable thesauri, ontologies and classification schemes for different subject areas. You can use Finto.fi to browse vocabularies or integrate the vocabularies into your own system using the open APIs.

Finto AI provides a user interface and API for automated subject indexing.

#### Available vocabularies and ontologies

##### GENERAL

- Allärs - General thesaurus in Swedish
- FGF - Finnish genre and form vocabulary
- HCLCS - Helsinki City Library Classification System
- KANTO - Kansalliset toimijatiedot
- KOKO Ontology
- Metatietosanasto
- PLC - Finnish Public Libraries Classification System
- Pondus categories
- UCUM - The Unified Code for Units of Measure
- UDC Summary
- YSA - General Finnish thesaurus
- YSO - General Finnish ontology
- YSOn käsite-ehdotukset
- YSO-time

##### SOCIETY

- IPTC NewsCodes
- JUHO - Julkishallinnon ontologia
- JUPO - Finnish Ontology for Public Administration Services
- KEKO - Ontology for Education for Sustainable Development
- Korkeakoulujen tutkimustiedonkeruussa käytettävä tieteenalaluokitus
- Lapponica
- LIIKO - Liikenteen ontologia
- OIKO - Oikeushallinnon ontologia
- PUHO - Puolustushallinnon ontologia
- SOTO - Sotatieteen ontologia
- Tietotermit
- TSR ontology

### Finto 10 years!

.....finto<sup>10</sup>

Finto service turns 10 this year! The Anniversary Seminar is held on Friday, 1 November at Hotel Arthur's Ballroom, Helsinki. Streaming is available. The programme will be in Finnish. Welcome!

[Register](#)

# Wikidata

**Wikidata is a free and open knowledge base that can be read and edited by both humans and machines.**

**Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wiktionary, Wikisource, and others.**

**Wikidata is a linking hub for external datasets; Wikidata items can contain external identifiers.**

Wikidata SPARQL query service:

[https://www.wikidata.org/wiki/Wikidata:SPARQL\\_query\\_service](https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service)

The image shows a screenshot of a Wikidata item page for Douglas Adams (Q42). The page is annotated with various labels and boxes to highlight specific features:

- label**: Points to the name "Douglas Adams (Q42)".
- description**: Points to the text "English writer and humorist" and "Douglas Noël Adams | Douglas Noel Adams".
- aliases**: Points to the text "Douglas Noël Adams | Douglas Noel Adams".
- property**: Points to the "educated at" property.
- value**: Points to the value "St John's College".
- qualifiers**: Points to the table of qualifiers for the "educated at" statement, including "end time", "academic major", "academic degree", and "start time".
- rank**: Points to the "rank" column in the statements table.
- statement group**: Points to the entire "Statements" section.
- opened references**: Points to the "reference URL" and "title" fields in the "Encyclopaedia Britannica Online" reference.
- collapsed reference**: Points to the "Brentwood School" reference, which is collapsed.

The "Statements" section contains the following data:

rank	property	value	qualifiers	references
	educated at	St John's College	end time: 1974 academic major: English literature academic degree: Bachelor of Arts start time: 1971	2 references
		Brentwood School	end time: 1970 start time: 1959	0 references

# Thank You!