Aalto University
School of Science

UNIVERSITY OF HELSINKI

SeCo

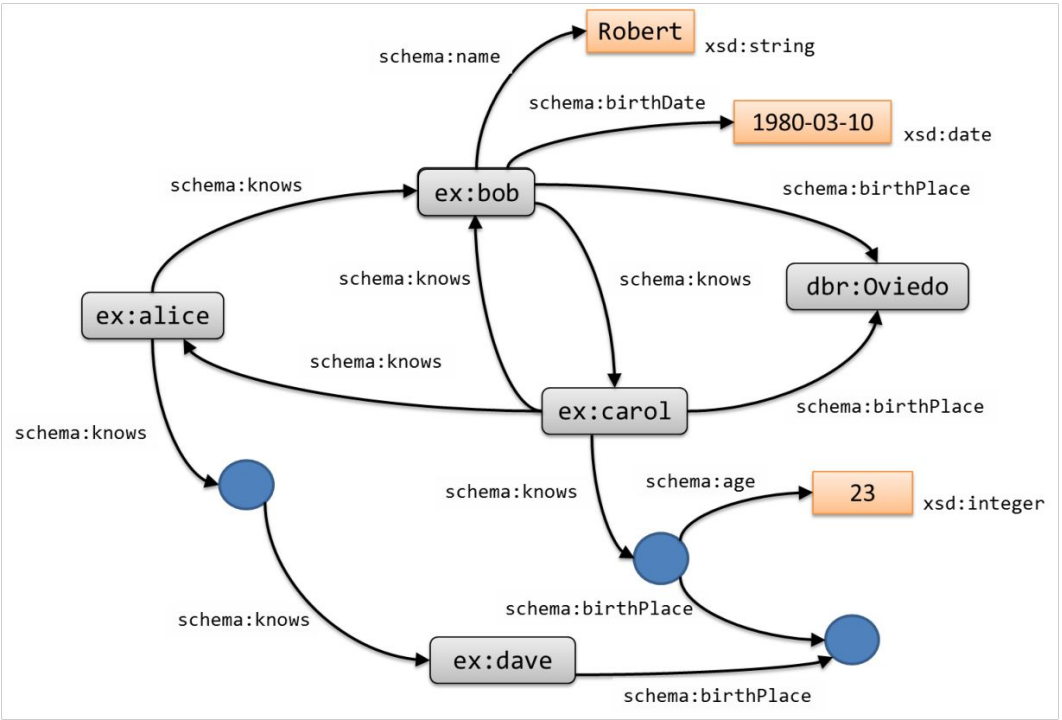
HELDIG
Helsinki Centre for Digital Humanities

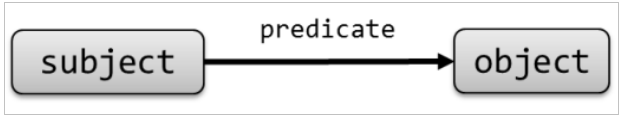
HELSINKI INSTITUTE FOR
SOCIAL SCIENCES
AND HUMANITIES

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

Gayo, Prud'hommeaux, Boneva, Kontokostas (2018).
Validating RDF Data. Morgan & Claypool.
https://book.validatingrdf.com/bookHtml008.html

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/Q13972 →
    Human: https://www.wikidata.org/wiki/Q13972
    Machine:
    https://www.wikidata.org/wiki/Special:EntityData/Q13972.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

Aalto University
School of Science
UNIVERSITY OF HELSINKI
SeCo
HELDIG
Helsinki Centre for Digital Humanities
HELSINKI INSTITUTE FOR
SOCIAL SCIENCES
AND HUMANITIES

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	More details
URI	http://purl.org/dc/terms/creator
Label	Creator
Definition	An entity responsible for making the resource.
Comment	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.
Type of Term	Property
Range Includes	• http://purl.org/dc/terms/Agent
Equivalent Property	• http://xmlns.com/foaf/0.1/maker
Subproperty of	• Creator (http://purl.org/dc/elements/1.1/creator) • Contributor (http://purl.org/dc/terms/contributor)

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

CIDOC CRM:
event-based data model for cultural heritage

Aalto University
School of Science

UNIVERSITY OF HELSINKI

SeCo

HELDIG
Helsinki Centre for Digital Humanities

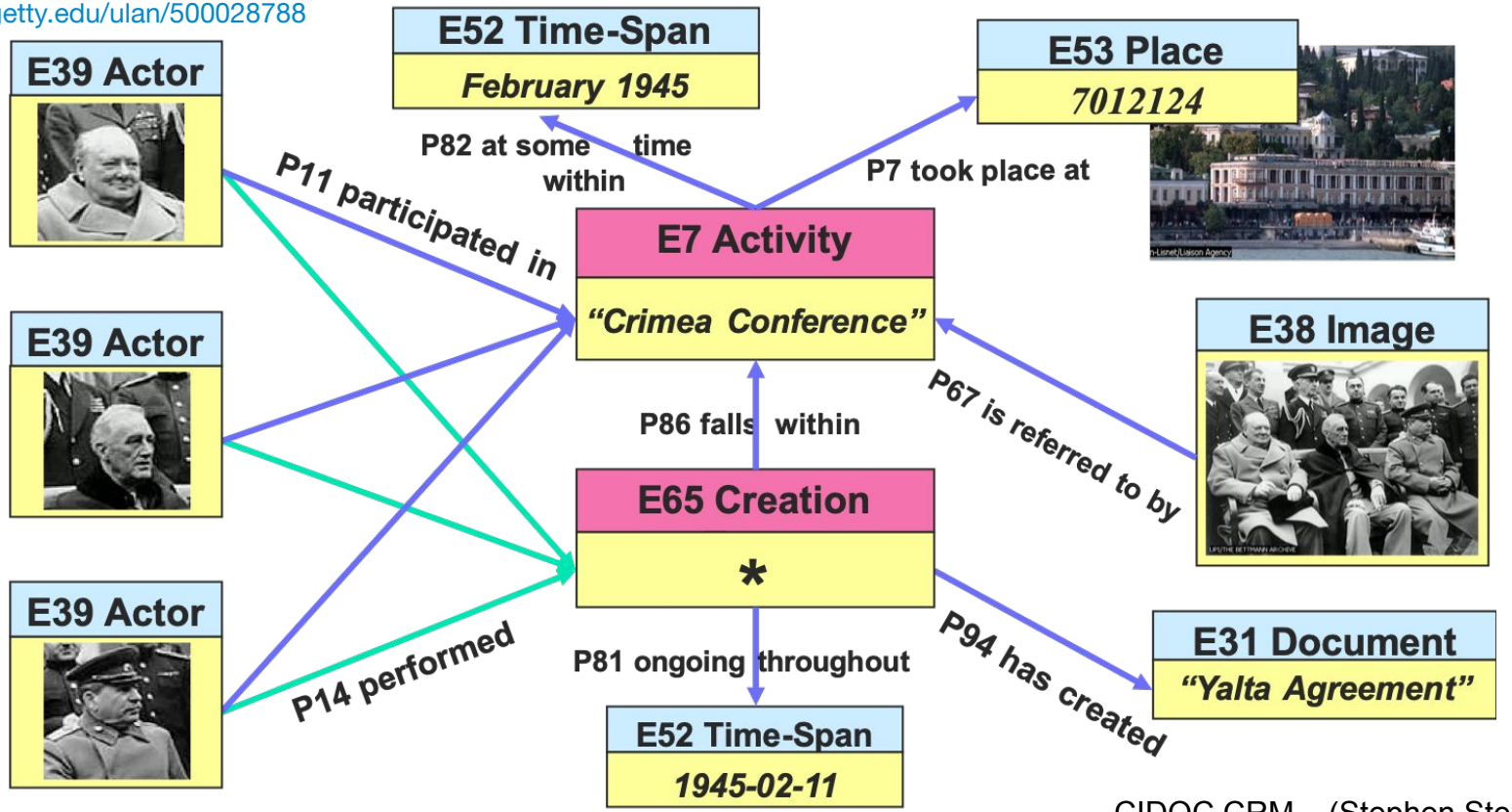
HELSINKI INSTITUTE FOR
SOCIAL SCIENCES
AND HUMANITIES

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

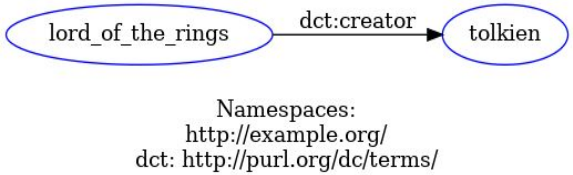
http://vocab.getty.edu/ulan/500028788



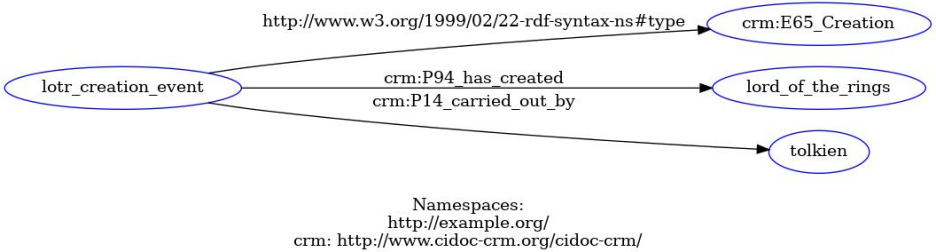
CIDOC CRM (Stephen Stead, 2008)

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

Dublin Core

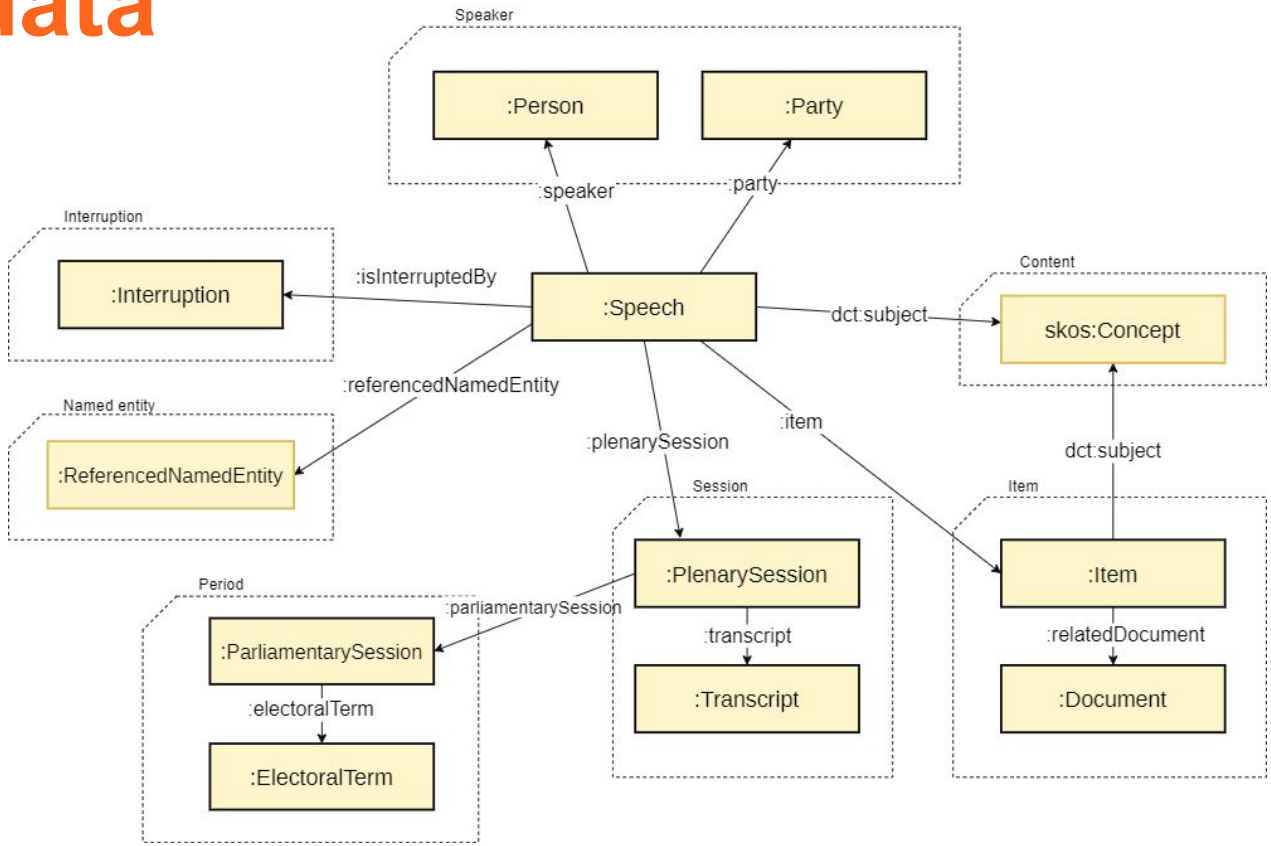


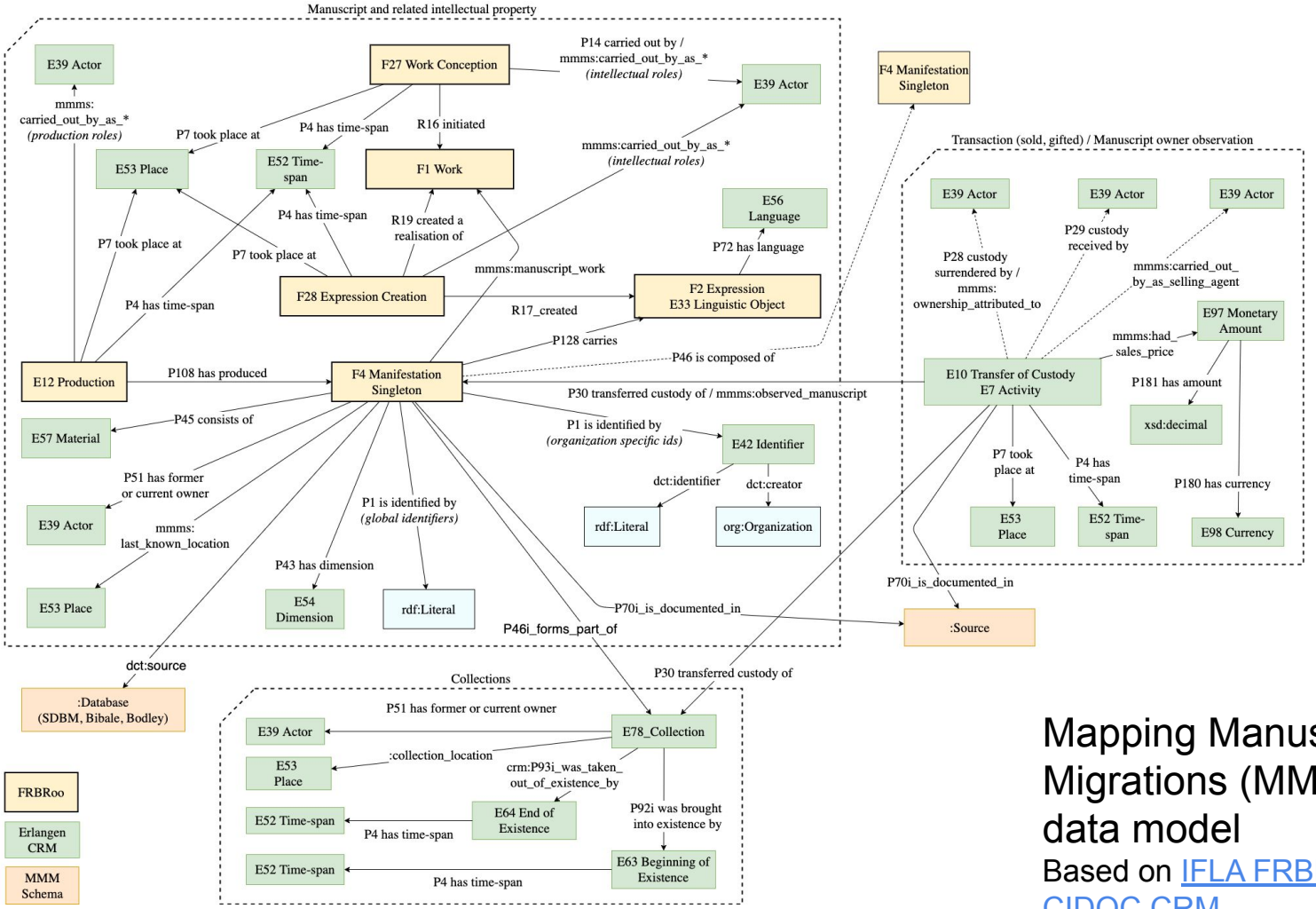
CIDOC CRM



Sampo data model examples

ParliamentSampo Speech data model





Mapping Manuscript Migrations (MMM) data model

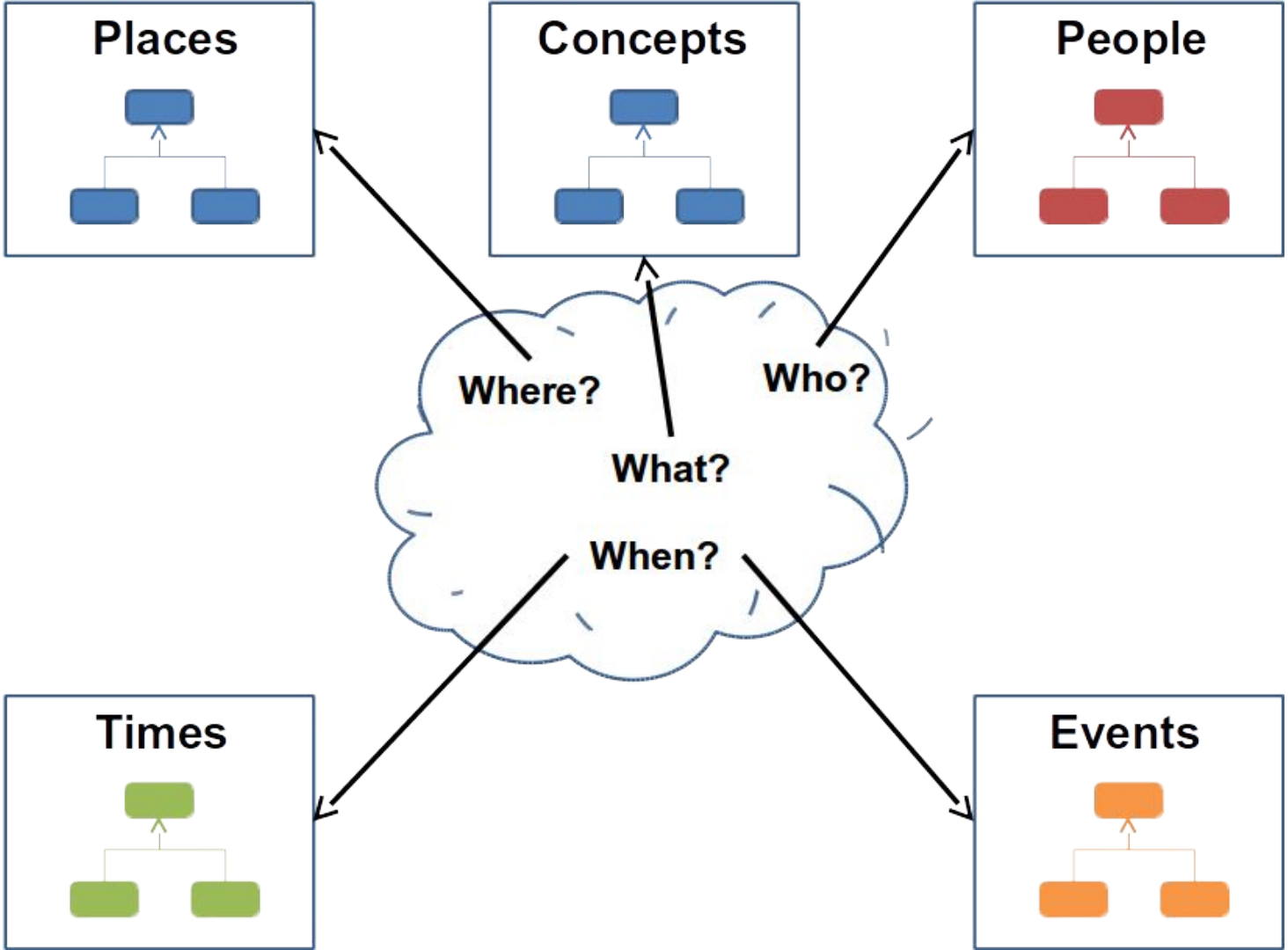
Based on IFLA FRBR and CIDOC CRM

Ontologies

Ontology (in computer science)

- **An ontology is a *formal*, *explicit* specification of a *shared conceptualization*.**[1]
- **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


Places
Concepts
People
Where?
What?
Who?
When?
Times
Events

Domain ontologies

Actors

URI: http://dbpedia.org/resource/Pyotr_Ilyich_Tchaikovsky

Pjotr Tšaikovski (fi)
Пётр Ильич Чайко́вский (ru)
Pyotr Ilyich Tchaikovsky (en)
Pjotr Tjajkovskij (sv)
Pjotr Tsjajkovskij (no)
Piotr Iljitsch Tschaikowski (de)
Piotr Ilitch Tchaïkovski (fr)
Piotr Ilich Chaikovski (es)
Pëtr Il'ič Čajkovskij (it)
Pjotr Iljitsj Tsjaikovski (nl)
Piotr Ilitch Tchaikovsky (pt)
Piotr Czajkowski (pl)
Piotr Ilici Ceaikovski (ro)
Pjotr Iljics Csajkovszkij (hu)

Events

SAHA3 | historia - search

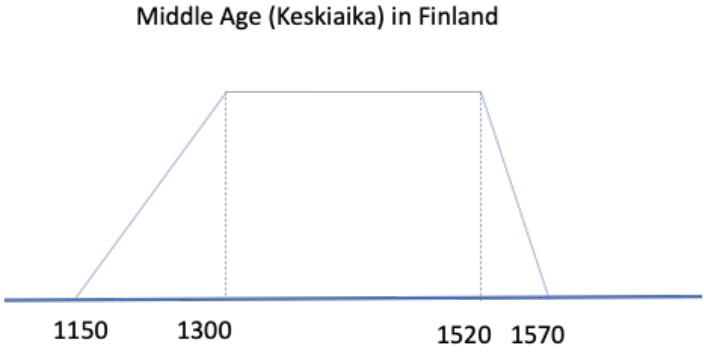
http://agricola.utu.fi/rdf/phe863

historiallinen tapahtuma: YYA-sopimus raukesi

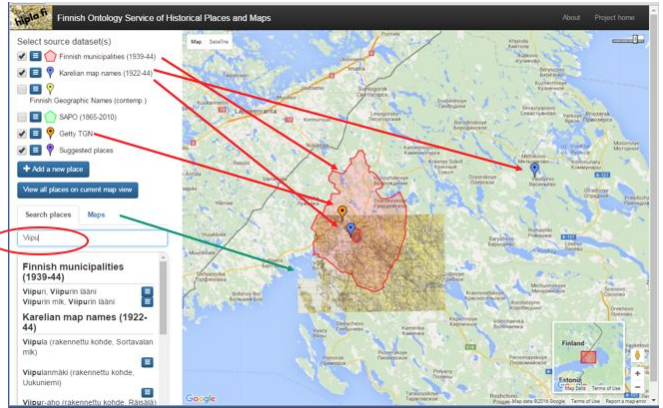
[edit]

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	allekirjoittaminen, irtisanominen, poliittinen päätöksenteko
tulos	Suomen tasavallan ja Venäjän federaation välinen sopimus suhteiden perusteista 63/1992
tapahtuma-aika	1992-01-20
tapahtumapaikka	Helsinki(1966-2008)
asiasana	diplomatia, kansainväliset sopimukset, kansainväliset suhteet, ulkopolitiikka, YYA-sopimus
liittyy	YYA-sopimus solmittiin Neuvostoliiton kanssa, YYA-sopimusta jatkettiin kahdellakymmenellä vuodella
liittyvä web-sivu	(fi) http://www.finlex.fi/fi/sopimukset/sopsteksti/1992/19920063
syy	Neuvostoliitto lakkautetaan
tyyppi	historiallinen tapahtuma

Times and periods



Places



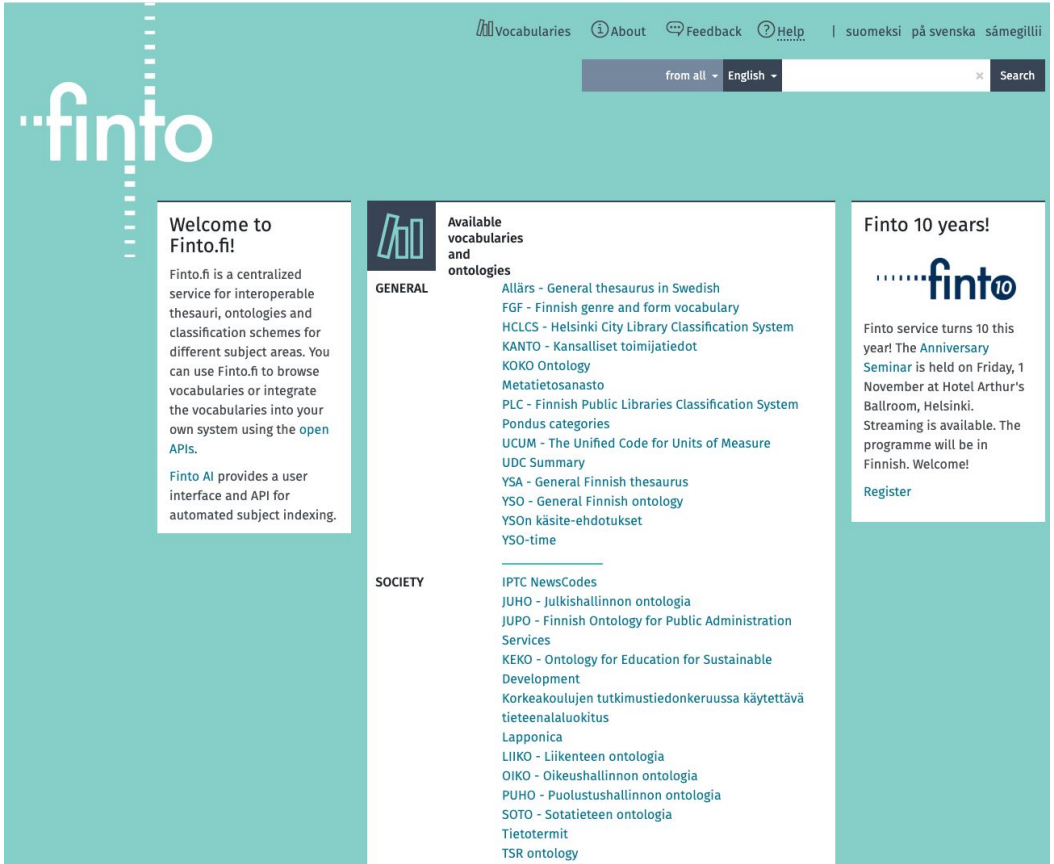
Keyword concepts

Data models for ontologies: SKOS, OWL

- **SKOS: for thesauri, "lightweight" ontologies**
  - Concept hierarchy, associative relations
  - "A Little Semantics Goes a Long Way"[1]
- **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)



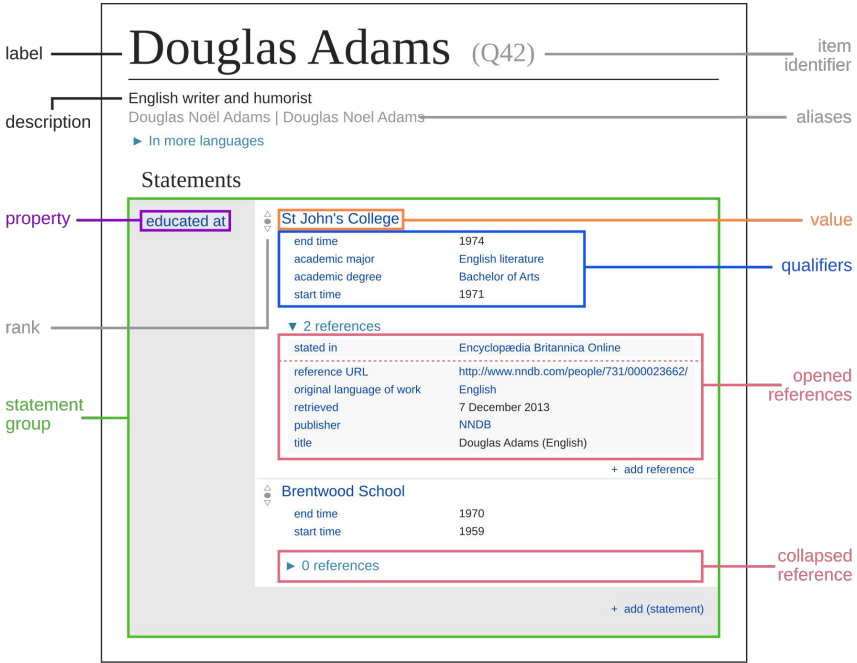
Wikidata

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



Thank You!