Knowledge Organisation Systems

- Form and Utility -

Martin Doerr

**Center for Cultural Informatics,
Institute of Computer Science
Foundation for Research and Technology - Hellas**

Heraklion, June 18, 2015

We distinguish:

- **Syntactic Interoperability:**
  - Information systems **can exchange** data objects and all their elements **without loss** of data. (EXCEL, HTML, XML, RDF…)
- *Semantic interoperability:*
  - Information systems **can communicate** and **combine** data objects and all their parts **consistent with the meaning** intended by the data creators or maintainers.
  - Systems **seem to understand**, help people talk to each other:

    "**Dream**: ….input in Chinese & English, but: Query and answers all in Russian".

Semantic interoperability can be divided into:

1) *Interoperable data structures/schemata*
- understand types of relationships: context
- either use of a standard schema
- or map & transform data between schemata
- ➔ **"formal ontologies"** describe concepts as common reference for schema equivalence.

2) *Identity of items referred to by data*
- local people, places, objects, events **never** appearing elsewhere, or with a publicly clear local ownership (collection items!).
- people, places, objects, events appearing **elsewhere** and **no local** ownership.
- **concepts**, categories, typologies characterizing items or subjects.
- ➔ "Knowledge Organisation Systems" describe identity of common references

Distinguish particulars from universals as a perceived truth.

o   Particulars do **not** have specializations.

o   Universals have **instances**, which can be either particulars or universals.

o   **particulars**: me, "hello", 2, WW II, the Mona Lisa, the text on the Rosetta Stone, 2-10-2006, 34N 26E, City of London

o   **universals**: patient, word, number, war, painting, text, car model, species

o   "strange" universals: colors, materials, mythological beasts

o   "strange" particulars: literary characters

o   Dualisms:

  ▪   *Texts as "equivalence classes" of documents containing the "same text".*

  ▪   *concepts as objects of discourse, e.g. "this is a 'chaffinch'" versus "Linné defined 'Fringilla coelebs Linnaeus, 1758' in 1758".*

The term KOS comes from library/information science:

"indexing languages", i.e.,

authoritative lists of items and concepts frequently referred in information systems in order to avoid using different names or identifiers for the same thing,

describing properties and definitions for identification ("matching") and names and identifiers for reference

often extending into useful relations to inform people and allow systems to make automated inferences for search and retrieval

Such inferences are

- identity (get all cats by "cat")
- generalization (get "cats" by "felines")
- related terms (get Heraklion by "Candia", get "bridge construction" by "bridges", get Heraklion by "Crete")

KOS

KOS can be divided into:

1) **Terminology of Universals**
   o describe things by their nature and behavior (form, function, structure…)
   o generalize over universals for searching
   o provide typical/general relations of universals for searching
   o divide a domain for administration and searching ("classification")

2) **Identification of Particulars: persons, things, places, events**
   o describe items by unique combinations of properties
   o understand we talk about the same item/instance (regardless classification!!):
   o provide important relations between particulars for searching

   a database schema contains ~ 20-500 concepts
   a terminology contains ~100-10 million concepts
   modern information systems may contain more than 10 billions of particulars.

KOS

A controlled vocabulary is a limited list of terms to be used in a database field.

- only an authority may add terms.
- highly ambiguous for particulars (typically place names)

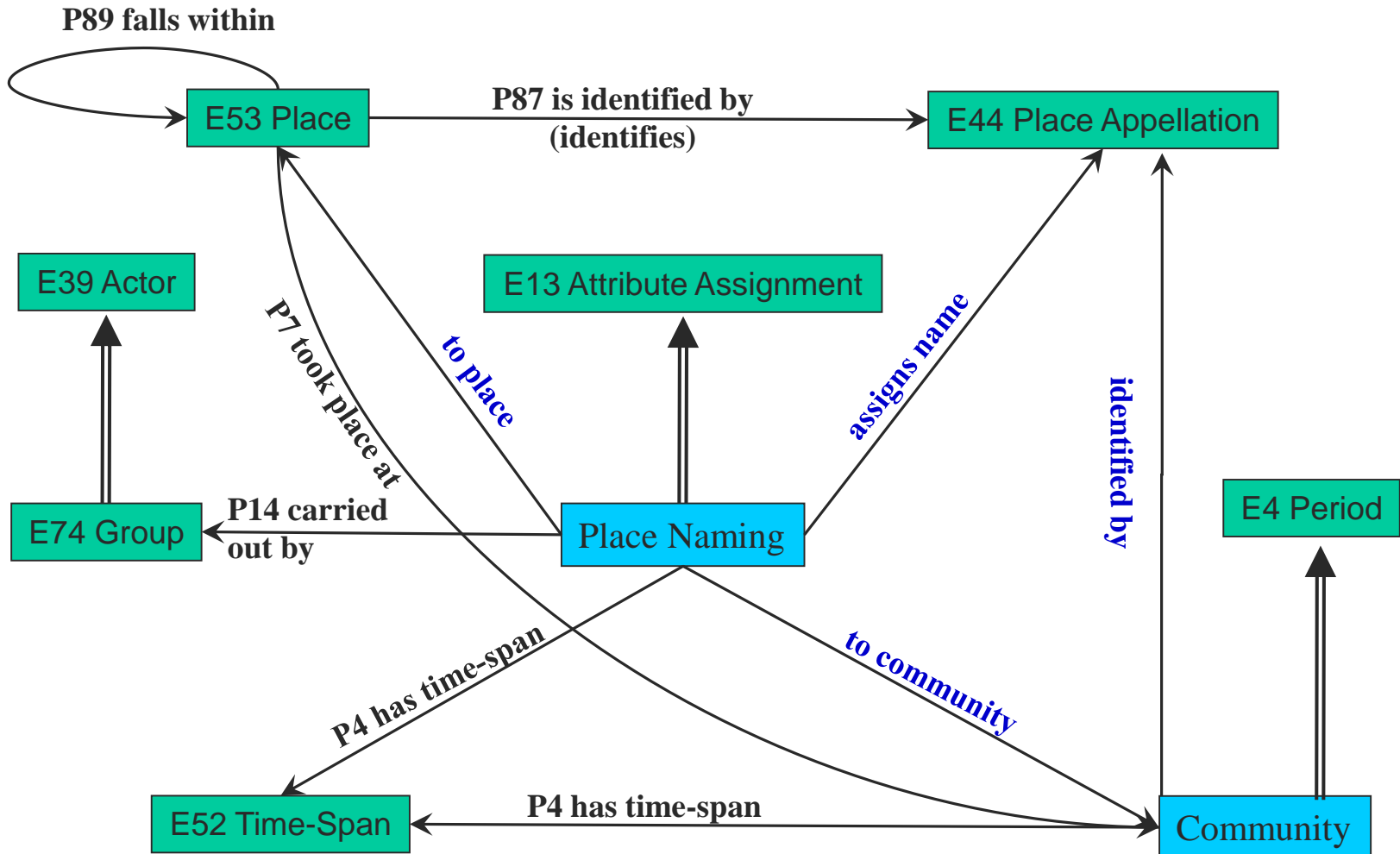*Authority files with identifying properties and recommended names*
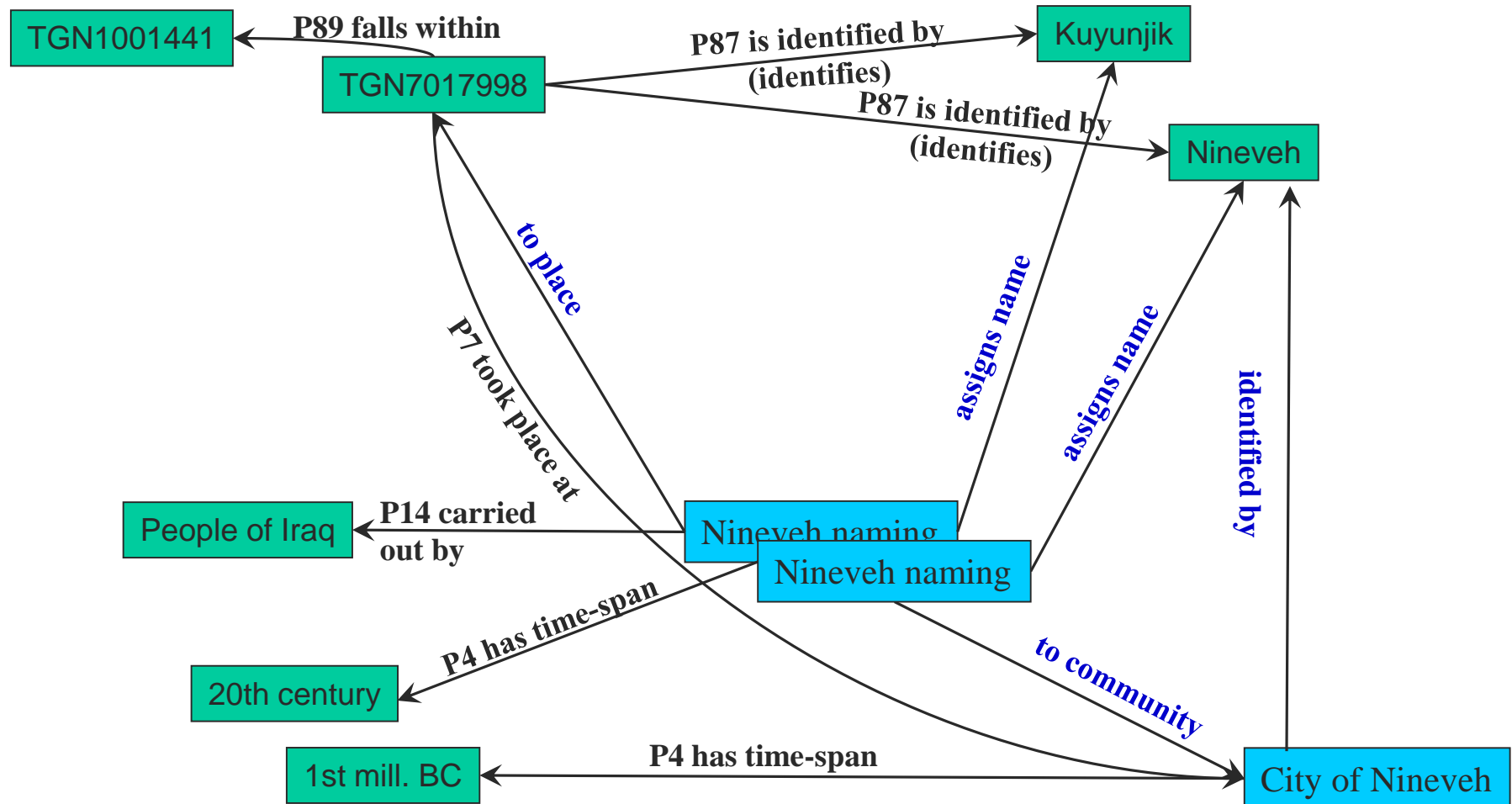
- only an authority may add terms.

Distinguishing:

- lists of persons (authors !) with life-dates, names, titles, roles, family and business relations
- "gazetteers": Lists and hierarchies of places together with recommended names (controlled) and geographic area.
- "thesauri of events or periods" : virtually non-existing yet!

P89 falls within

E53 Place

P87 is identified by (identifies)

E44 Place Appellation

E39 Actor

E13 Attribute Assignment

P7 took place at

to place

assigns name

identified by

E74 Group

P14 carried out by

Place Naming

E4 Period

P4 has time-span

to community

E52 Time-Span

P4 has time-span

Community

Nineveh (deserted settlement)

Note: Succeeded Nimrud as the Assyrian capital. It was made into a spectacular center by Sennacherib and developed into cultural center by Ashurbanipal in the 7th century BCE. A large library of cuneiform tablets was here. The city fell to Nabopolassar of Babylon in 612 BCE.

Names:
Nineveh (**preferred**,C,V)
Ninevah (C,V)
Ninive (H,O,French-P,U,N)
Ninus (H,O) .... ancient
Ninos (H,O)
Nineve (H,O)
Ninua (H,O)
Tell Nabi Yunus (C,V)
Kuyunjik (C,V)
Quyunjig (C,V)

Hierarchical Position:
World (facet)
.... Asia (continent) (P)
.... Iraq (nation) (P)
.... Nīnawā (governorate) (P)
.... Nineveh (deserted settlement) (P)

Additional Parents:
World (facet)
.... Asia (continent) (P)
.... Assyria (former nation/state/empire) (P,H)
.... Nineveh (deserted settlement) (P,H)

Place Types:
deserted settlement (**preferred**, C) .... was great city in 1st mill. BC
archaeological site (C)
capital (H) .... of Assyria, late 8th cen. BC
trade center (H)
inhabited place (H)

KOS

Hierarchical Position:
- World (facet)
- .... Europe (continent) (P)
- .... United Kingdom (nation) (P)
- .... England (country) (P)
- .... Greater London (metropolitan area) (P)
- .... London (inhabited place) (P)

Additional Parents:
- World (facet)
- .... Roman Empire (former nation/state/empire) (P,H)
- .... Britannia (province) (P)
- .... London (inhabited place) (P,H)

Place Types:

inhabited place (**preferred**, C)	....	established as a Roman town ca. 43 CE, after the Emperor Claudius invaded England
city (C)	....	since the late 6th century
national capital (C)	....	of the United Kingdom
capital (C)	....	of the country of England
county seat (C)		
manufacturing center (C)		
cultural center (C)		
episcopal see (C)	....	since the late 6th century
provincial capital (H)	....	of the Roman provinces Britannia Superior and Maxima Caesariensis
Roman (H)		
Romano-British (H)		

Related geographic places:
- capital of .... **Britannia Superior** .... (province)
  - .... (World, Roman Republic and Empire, Britannia) [6000962]
- capital of .... **Maxima Caesariensis** .... (province)
  - .... (World, Roman Republic and Empire, Britannia, Britannia Superior) [7030355]
  - .... from the early 4th century CE
- capital of .... **United Kingdom** .... (nation)
  - .... (World, Europe) [7008591]

ICS-FORTH June 18, 2015

KOS

London (geonames)

London 🌐 City of London,Gorad Londan,ILondon,LON,Lakana,Landen,Ljondan,Llundain,Londain,Londan,Londar,Londe,L...	**United Kingdom**, England Greater London	capital of a political entity population 7,556,900	N 51° 30' 3
City of London 🌐 Bandaraya London,Cathair Londan,Ceety o Lunnon,Cidade de Londres,Cite de Londres,Citta di Londra,Cit...	**United Kingdom**, England Greater London > City of London	seat of a third-order administrative division population 7,556,900	N 51° 30' 4
Gatwick Airport 🌐 Aerfort Londain-Gatwick,Aerodrom Getvik,Aeroport Gatvik,Aeroport de Londres Gatwick,Aeroport de Lond...	**United Kingdom**, England West Sussex > Crawley District	airport elevation 61m	N 51° 9' 21
Heathrow 🌐 Aehraport Khitrou,Aerfort Londain-Heathrow,Aerodrom Khitrou,Aeroport de Londres Heathrow,Aeroport de...	**United Kingdom**, England Greater London > Hillingdon	airport	N 51° 28' 1
London City Airport 🌐 Aerborzh Ker Londrez,Aerborzh Kêr Londrez,Aeroport de Londres City,Aeroport de la Ciutat de Londres,...	**United Kingdom**, England Greater London > Newham	airport elevation 5m	N 51° 30' 1
Sutton 🌐 London Borough of Sutton,Satton,Sutton,Sutton keruelet,Sutton kerület,Саттон	**United Kingdom**, England Greater London > Sutton	seat of a third-order administrative division population 187,600	N 51° 21' 0
Bexley 🌐 Bexley,Bexley keruelet,Bexley kerület,London Borough of Bexley,Londonska opshtina Beksli,Лондонска о...	**United Kingdom**, England Greater London > Bexley	populated locality population 228,000	N 51° 26' 2
Barnet 🌐 Barnet,Barnet keruelet,Barnet kerület,Barneto,Buirg Londan Barnet,Chipping Barnet,High Barnet,La Bar...	**United Kingdom**, England Greater London > Barnet	populated place population 30,000	N 51° 39' 0
Islington 🌐 District londonien d'Islington,Islington,ijeullingteon,izurinton,Ислингтон,ایزلینگتن,...	**United Kingdom**, England Greater London > Islington	seat of a third-order administrative division population 319,143	N 51° 32' 1

ID: 500005977　　　　　　　　　　　　　　**Record Type: Person**

Master of the Paradise Garden (German painter, active ca. 1410-1430)

Names:
　　Master of the Paradise Garden (**preferred**,V,display)
　　Master of the Frankfurt Paradise Garden (V)
　　Master Of The Paradise Garden (V)

Nationalities:
　　German (**preferred**)

Roles:
　　artist (**preferred**)
　　painter

Gender: male

List/Hierarchical Position:
　　.... Persons, Artists
　　........ Master of the Paradise Garden (I)

Biographies:
　　(German painter, active ca. 1410-1430) ..... [BHA Preferred]
　　(German artist, active 1410) ..... [WL-Courtauld]
　　(German artist, op.c.1410) ..... [WCP]
　　(German painter, act. ca.1410) ..... [GRLPSC]

*The **Virtual International Authority File** (**VIAF**) is an international <u>authority file</u>. It is a joint project of several national libraries and operated by the <u>Online Computer Library Center</u> (OCLC).[1][2] The project was initiated by the <u>German National Library</u> and the US <u>Library of Congress</u>.*

***The VIAF** combines multiple name authority files into a single OCLC-hosted name authority service. It is MARC based.*

VIAF** developed the idea of consolidated **identity** as a **cluster of identifiers
(<u>https://viaf.org/viaf/95161780/#Traven,_B</u>.
<u>https://viaf.org/processed/ISNI|0000000121441436</u>)

Kinds of KOS of Universals

A *dictionary* is a listing of *words* and phrases giving information such as

- spelling, morphology and part of speech,
- senses, definitions, usage, equivalents in other languages (bi- or multilingual dictionary).
- etymology

A *controlled vocabulary* is a limited list of terms to be used in a database field. Only an authority may add terms.

A **classification system** is a structure that organizes concepts into a (mono) hierarchy in order to **partition** some material following a sequence of *decision criteria.*

A ***thesaurus*** *is a controlled vocabulary of categorical terms related to concepts, and with scope notes and semantic relationships between concepts.*

  ○ semantic relationships are: IsA, related terms

subject catalogues *may use thesaurus relationships but interpret IsA as a generalization of "talking about".*

*A **monolingual** **thesaurus** has terms form one expert group or community*

*A **multilingual** **thesaurus** relates terms and concepts from two or more expert groups or communities (see next slide)*

ID: 300211087   **Record Type:** concept

🔱 **dolls** (figurines, statues, ... Visual and Verbal Communication (Hierarchy Name))

Note: Figurines representing humans or animals, including but not restricted to those intended as toys for children, usually girls, or as collectibles for adults. Dolls are commonly in the form of a baby or woman, often with changeable clothes; they may have moveable arms and legs. Doll may be made of cloth (rag dolls), wood, clay, porcelain, wax, paper, plastic, celluloid, corn husks, or other materials. Dolls may also be figurines used for ceremonial, religious, or decorative purposes. Archaeological evidence suggests that dolls were the first playthings; they have been found in Babylonian and Egyptian tombs from ca. 3000 BCE. In ancient Greece and Rome, maturing girls consecrated their childhood dolls to the goddesses. Cloth dolls in the form of animals are generally called "stuffed toys."

Terms:
    dolls (**preferred**,C,U,LC,English-P,D,U,PN)
    doll (C,U,English,AD,U,SN)
    poppen (C,U,Dutch-P,D,U,U)
    pop (C,U,Dutch,AD,U,U)
    poupées (C,U,French-P,D,U,PN)
    poupée (C,U,French,AD,U,SN)
    muñecas (C,U,Spanish-P,D,U,PN)
    muñeca (C,U,Spanish,AD,U,SN)
    muñeco (C,U,Spanish,UF,U,SN)

Facet/Hierarchy Code: V.VC

Hierarchical Position:
    🔱   Objects Facet
    🔱   .... Visual and Verbal Communication (Hierarchy Name) (G)
    🔱   .... Visual Works (Hierarchy Name) (G)
    🔱   .... visual works (works) (G)
    🔱   .... <visual works by material or technique> (G)
    🔱   .... sculpture (visual work) (G)
    🔱   .... <sculpture by subject type> (G)
    🔱   .... statues (G)
    🔱   .... figurines (G)
    🔱   .... dolls (G)

Additional Parents:

- Objects Facet
- .... Furnishings and Equipment (Hierarchy Name) (G)
- .... Recreational Artifacts (Hierarchy Name) (G)
- .... recreational artifacts (equipment) (G)
- .... toys (recreational artifacts) (G)
- .... <toys by form> (G)
- .... doll-playing accessories (G)
- .... dolls (G)

Additional Notes:

Dutch ..... Beeldjes die mensen of dieren voorstellen, inclusief maar niet uitsluitend de figuurtjes die bedoeld zijn als speelgoed voor kinderen, vooral meisjes, of als verzamelobjecten voor volwassenen. Poppen stellen vaak een baby of vrouw voor en hebben soms beweegbare armen en benen. Dikwijls is het mogelijk ze andere kleding aan te trekken. Een pop kan gemaakt zijn van stof (lappenpoppen), hout, klei, porselein, was, papier, plastic, celluloid, maïsvliezen of andere materialen. Poppen kunnen ook figuurtjes zijn die worden gebruikt voor ceremoniële, religieuze of decoratieve doeleinden. Er is archeologisch bewijs dat poppen de eerste speeltjes waren; ze zijn aangetroffen in Babylonische en Egyptische graftomben uit circa 3000 v. Chr. In het oude Griekenland en Rome wijdden meisjes als ze volwassen werden de poppen uit hun kindertijd aan de godinnen. Poppen van stof in de vorm van dieren worden meestal 'pluchen speelgoed' genoemd.

Spanish ..... Figuras humanas o humanoides, especialmente aquellas usadas para jugar, y algunas usadas para propósitos ceremoniales, religiosos o decorativos.

Related concepts:

created by .... **dollmakers**
.... (<people in crafts and trades by product>, <people in crafts and trades>, ... People) [300025399]
meaning/usage overlaps with ....   **stuffed toys**
.... (soft toys, <toys by form>, ... Furnishings and Equipment (Hierarchy Name)) [300211252]
produced by .... **dollmaking**
.... (<object-making processes and techniques>, <processes and techniques by specific type>, ... Processes and Techniques (Hierarchy Name)) [300053665]

IsA Generalization is based on strict inheritance of properties:

- All narrower concepts must have all properties or potential of properties as the more general ones (plus their own).

Robust criteria for IsA regard:

- An a priori fixed scope of use
- compatibility of substance ("paper" or "letter"?)
- ways or reasons for coming into existence ("tree" or "rosacea" ?)
- compatibility of behavior or function ("flying" or "disc shape" ?)
- ability of recognizing/knowing ("professor" or "intelligent" ?)

Applying such criteria increases chance of

- individuals coming to the same generalizations/ decisions
- larger groups agreeing on the same general concepts
- individuals learning the "indexing language"

SKOS Simple Knowledge Organization System *(Home Page)*

SKOS is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading lists and taxonomies within the framework of the Semantic Web
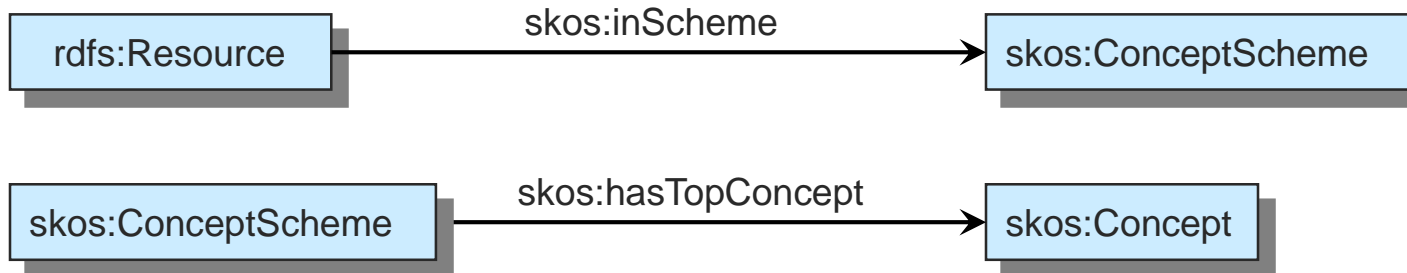
Classes

- **skos:Concept**
- **skos:ConceptScheme**
- **skos:collection**
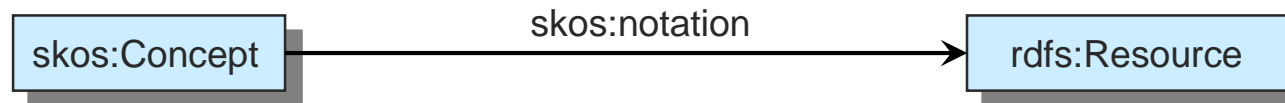- **skos:OrderedCollection (sub-class of skos:collection)**

Interthesaurus relations

Concept scheme properties

| rdfs:Resource | — skos:inScheme → | skos:ConceptScheme |

| skos:ConceptScheme | — skos:hasTopConcept → | skos:Concept |

Notation property

| skos:Concept | — skos:notation → | rdfs:Resource |

Interthesaurus relations

Hierarchical Relations

skos:Concept — skos:broader → skos:Concept

skos:Concept — skos:narrower → skos:Concept

skos:Concept — skos:broaderTransitive → skos:Concept

skos:Concept — skos:narrowerTransitive → skos:Concept
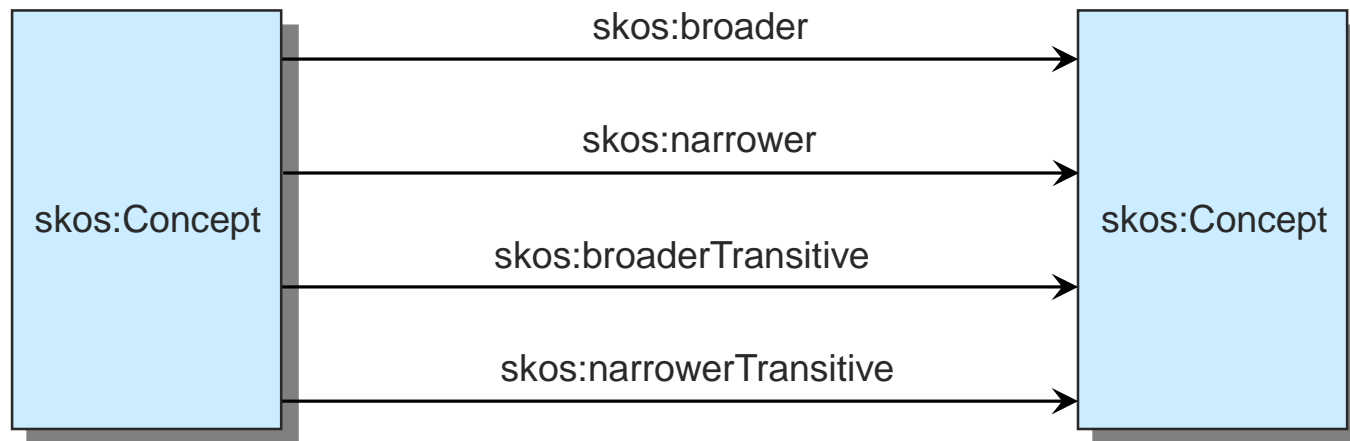
We take "skos:broaderTransitive" for IsA

KOS

Interthesaurus relations

Associative Relations

| skos:Concept | → skos:related → | skos:Concept |

Equivalence Relations

| skos:Concept | → skos:altLabel → | rdfs:Literal |

| skos:Concept | → skos:hiddenLabel → | rdfs:Literal |

Grouping relations

| skos:Collection | → skos:member → |

skos:Concept

skos:Collection
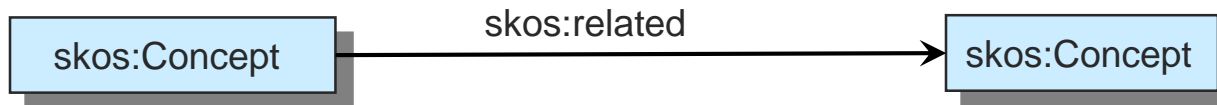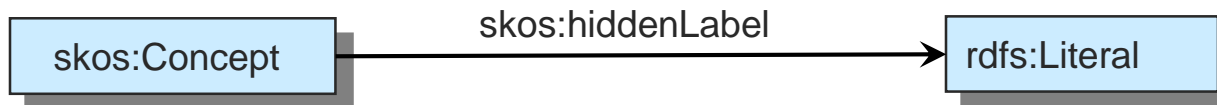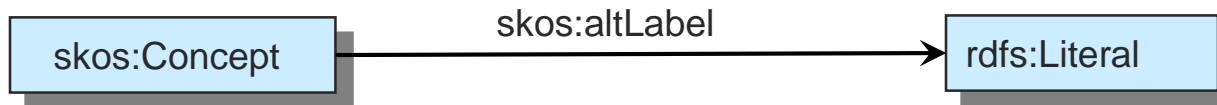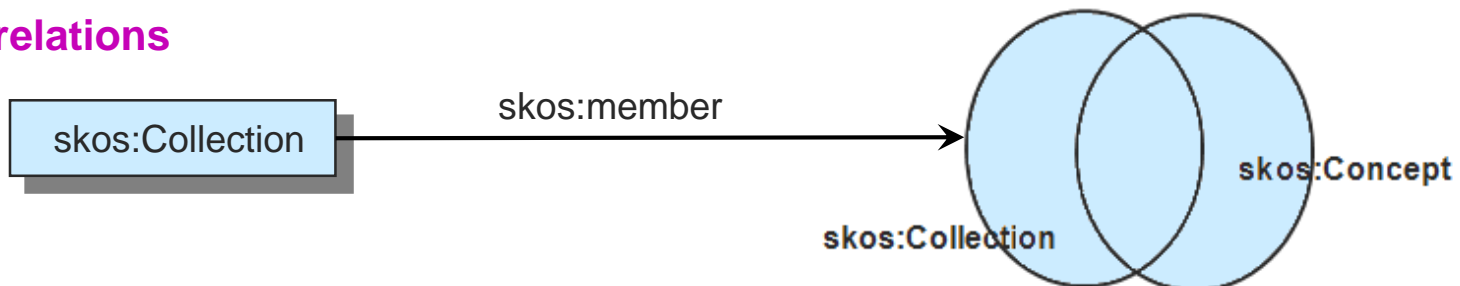
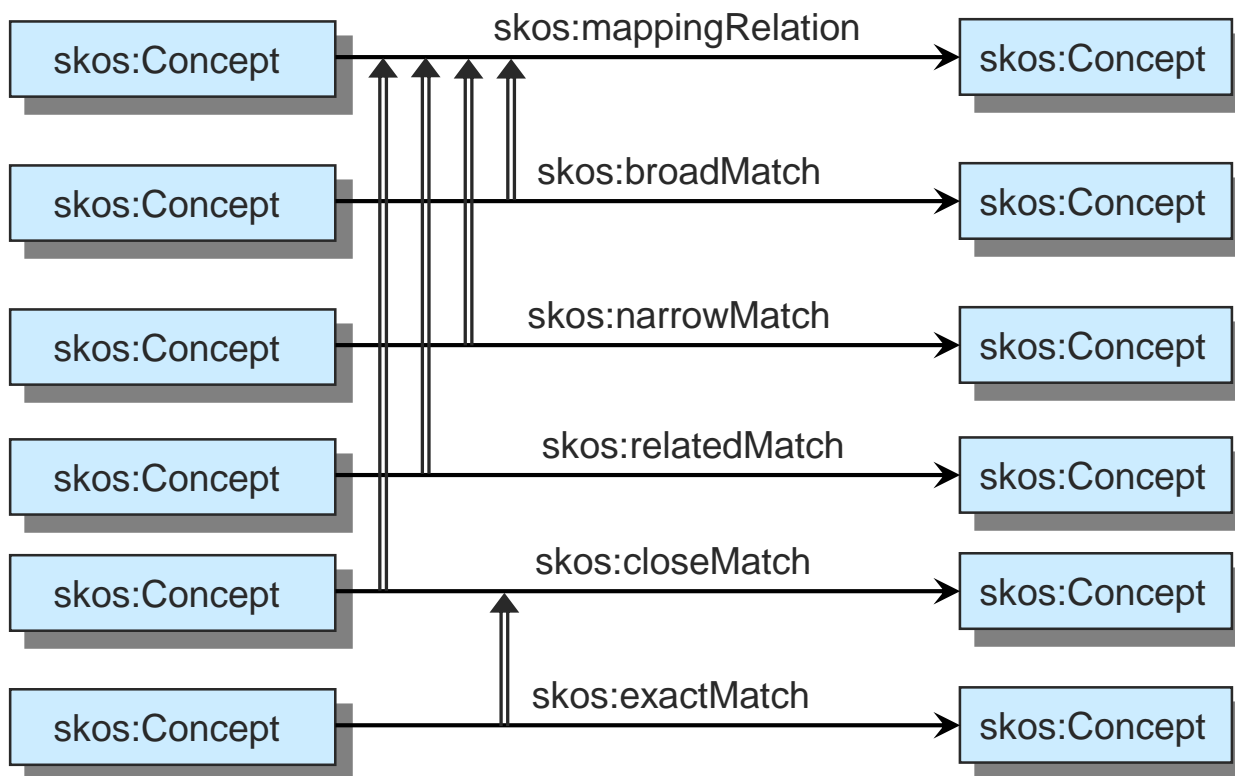## Interthesaurus relations

Documentation properties
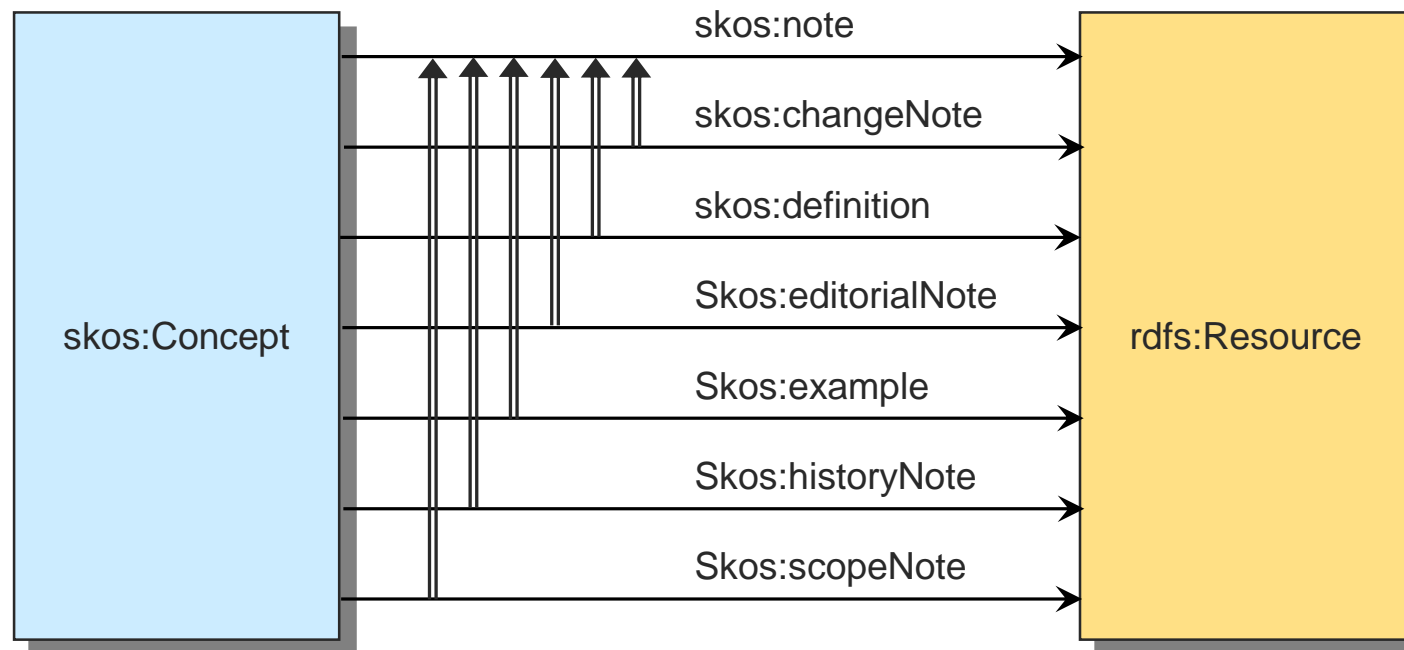
Intrathesaurus relations

Mapping properties

Making a person a SKOS:Concept would mean

- there is something "broader" of me ("gvp:parentString"!!)
- something "narrower" of me
- something is a kind of "me"
- ⇒ *impossible to consolidate schema integration on that base*

  http://www.getty.edu/vow/ULANFullDisplay?find=Paradise+Garden&role=&nation=&prev_page=1&subjectid=500005977

Literary characters are not persons!

- In FRBRoo: F38 Character . R57 is based on (is basis for) : E39 Actor
- Characters vary between authors. I do not vary depending on other people.
- Characters are concepts, with generalizations/specializations
- As instances we may regard their (propositional) role in stories.

Making a "Place" a SKOS:Concept would mean

- Asia is "broader" of Istanbul…
- Completely impossible t describe spatiotemporal overlap as "hierarchy".

BT partitive (BTP/NTP) is defined in ISO2788. Example:

- central nervous systems
  BTP nervous systems
- nervous systems
  NTP central nervous systems
- In analogy: "Heraklions" use to be part of "Greeces" ???
  (http://www.getty.edu/vow/TGNFullDisplay?find=Heraklion&place=&nation=&prev_page=1&english=Y&subjectid=7002690)

Talking about a place is not a place!

- http://id.loc.gov/authorities/subjects/sh85056605.html

"**Great Britain**

*Here are entered **works on** the United Kingdom of Great Britain and Northern Ireland, which comprises England, Northern Ireland, Scotland, and Wales, as well **as works on** the island of Great Britain. Works on the Republic of Ireland and on the island of the British Isles called Ireland are entered under [Ireland.] Works on the non-jurisdictional island group comprising the islands of Great Britain, Ireland, and smaller adjacent islands are entered under [British Isles.]*

- Library of Congress does **not** confuse places with geographical subjects!

Never mix particulars and universals! Never use SKOS for particulars!

- Completely different fields and reasoning mechanisms
- Abusing "broader term" for spatial inclusion inhibits integration.

The subjectivity of criteria to define a place/period does not make a place/period a product of the mind. Telling stories about a person does not make the person a story.

There are no real standards for KOS of particulars

- For persons: VIAF and EAC-CPF are **good** prototypes
- For gazetteers, periods: We propose work of DAI.

*Semantic abuse of a standard format is **NOT** using a standard!*