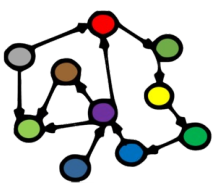
Welcome to INFO216:
Knowledge Graphs
Spring 2022

Andreas L Opdahl
<Andreas.Opdahl@uib.no>

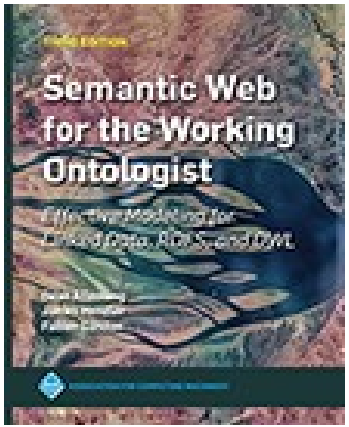
Session 9: Vocabularies (and ontologies)

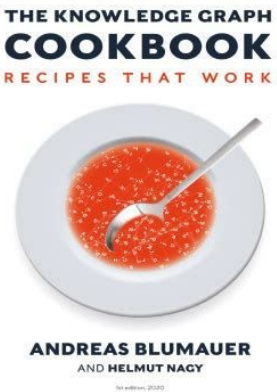
- Themes:
  - semantic vocabularies
  - *SKOS, schema.org, DC, FOAF, geo, VANN, VS, CC, PROV, Event, Time, Timeline (tl), BIO, SIOC, Bibo, Music (mo)...*
  - the open KG ontologies
    - *Wikidata, DBpedia, GeoNames…*
  - the *News Hunter ontologies*



Readings

- Sources:
  - Allemang, Hendler & Gandon(2020): Semantic Web for the Working Ontologist, 3rd edition (chapters 10-11, but chapters 9-10 in the 2nd edition)
  - Blumauer & Nagy (2020): Knowledge Graph Cookbook – Recipes that Work (about time to finish parts 1 and 3)
- Material at http://wiki.uib.no/info216:
  - supplementary links in the portal, including
  - Linked Open Vocabularies (LOV) http://lov.okfn.org/dataset/lov/

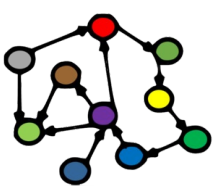




Why vocabularies (and ontologies)

Terms

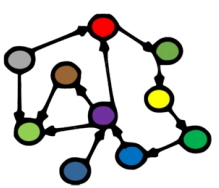
- *Semantic vocabularies / ontologies*
  - graphs/datasets (in RDFS, OWL...) that define:
  - standard URIs for *types of resources*
  - standard URIs for *properties*
  - standard URIs for some important *individuals*
  - standard datatypes (identified by URIs) for *literals*
- *Linked open semantic datasets / knowledge graphs*
  - graphs/datasets (in RDF, RDFS, OWL...) that define:
  - standard URIs for *individual resources*
  - facts (as triples) about those *individual resources*
  - *may* also define their own vocabularies



Terms

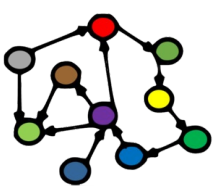
- *Semantic vocabularies / ontologies*
  - graphs/datasets (in RDFS, OWL...) that define:
  - standard URIs for *types of resources*
  - standard URIs for *properties*
  - standard URIs for some important *individuals*
  - standard datatypes (identified by URIs) for *literals*
- *Linked open semantic datasets / knowledge graphs*
  - graphs/datasets (in RDF, RDFS, OWL...) that define:
  - standard URIs for *individual resources*
  - facts (as triples) about those *individual resources*
  - *may* also define their own vocabularies

But: ontologies can contain individuals and KGs can contain classes and restrictions



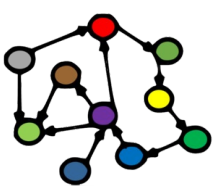
Why do we need vocabularies and ontologies?

- Shared, well-defined terms (dereferencable URIs) for types, properties and some individuals that can be used to represent a domain
- Formal rules and constraints
- Domains can be:
  - people, their friends and workplaces (FOAF, BIO)
  - electronic and other documents (DC, BIBO)
  - commerce (schema.org)
  - classification in libraries etc. (SKOS)
  - general encyclopedic information (DBpedia, Wikidata)
  - general time and place (OWL-Time, geo)
  - ...and *lots* of others



Why do we need vocabularies and ontologies?

- To make knowledge graphs more precisely defined
- To make semantic data sets easier to use
  - encourage reuse
  - avoid misunderstandings and errors
  - easier to understand, recombine, enrich...
- To support computer processing
  - more powerful
  - more general
  - automated reasoning over ontologies
    - inference rules a la RDFS
    - logic reasoning a la description logic (DL)
    - ...and there is also *query rewriting*

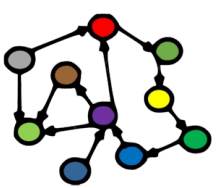


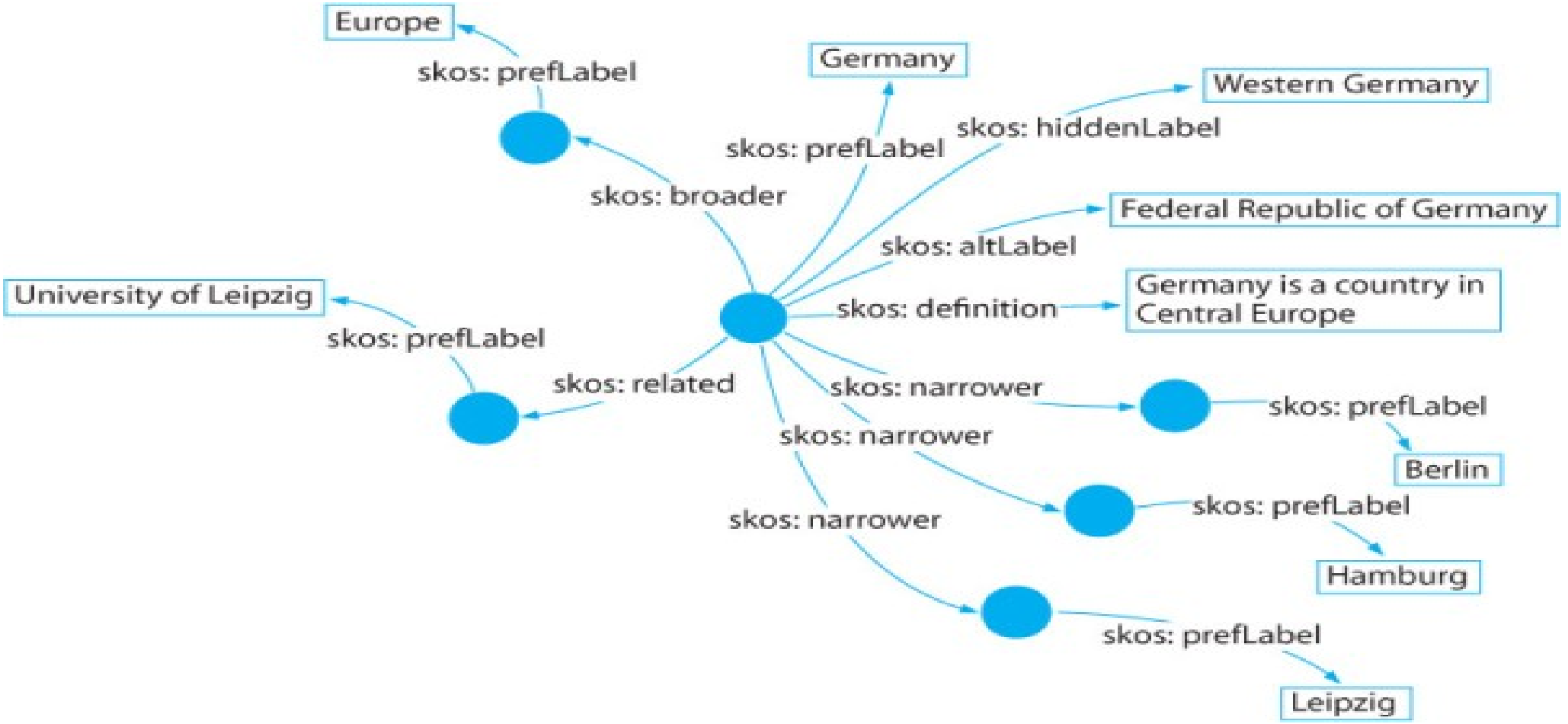
Semantic vocabularies

http://lov.okfn.org/dataset/lov/

Simple Knowledge Org. System (SKOS)

- Making classification schemes, subject heading lists, taxonomies and other fixed vocabularies (or knowledge organization systems, KOS) within the Web of Data
  - also: providing *mappings* between schemes
- @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
- Main concepts:
  - classes: Concept, ConceptScheme
  - properties: for describing concepts and for interrelating concepts within and between schemes
- *Uses:*
  - *widely used to represent, exchange and interrelate catalogues, e.g., by the Library of Congress*

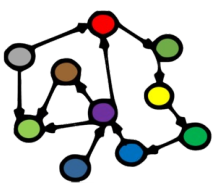




Source: www.w3c.it

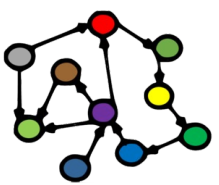
SKOS: describing concepts

- Class: skos:Concept
- Properties that describe concepts:
  - skos:prefLabel, skos:altLabel, skos:hiddenLabel
  - skos:note, and its subproperties:
    - skos:definition, skos:example, skos:changeNote, skos:editorialNote, skos:historyNote, skos:scopeNote
  - skos:notation (typed literals of *external classifications*)
- Properties that relate concepts:
  - skos:semanticRelation, and its subproperties:
    - skos:related, skos:broader, skos:broaderTransitive, skos:narrower, skos:narrowerTransitive
    - ...within the same *concept scheme*



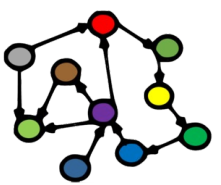
SKOS: property naming

- Undirectional property naming *(bad!)*:
  - skos:narrower means “hasNarrowerConcept”
  - skos:broader means “hasBroaderConcept”
- ex:animals rdf:type skos:Concept;
  skos:prefLabel "animals"@en;
  skos:narrower ex:mammals.
- ex:mammals rdf:type skos:Concept;
  skos:prefLabel "mammals"@en;
  skos:broader ex:animals.
  - SKOS uses rdfs:labels to make this clearer...



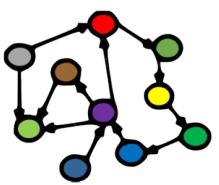
SKOS: concept schemes

- Class: skos:ConceptScheme
  - skos:inScheme, skos:topConceptOf, skos:hasTopConcept
- Properties that relate concepts in different schemes:
  - skos:mappingRelation, and its “rdfs:subProperties”:
    - skos:closeMatch, skos:exactMatch, skos:relatedMatch, skos:broadMatch, skos:narrowMatch

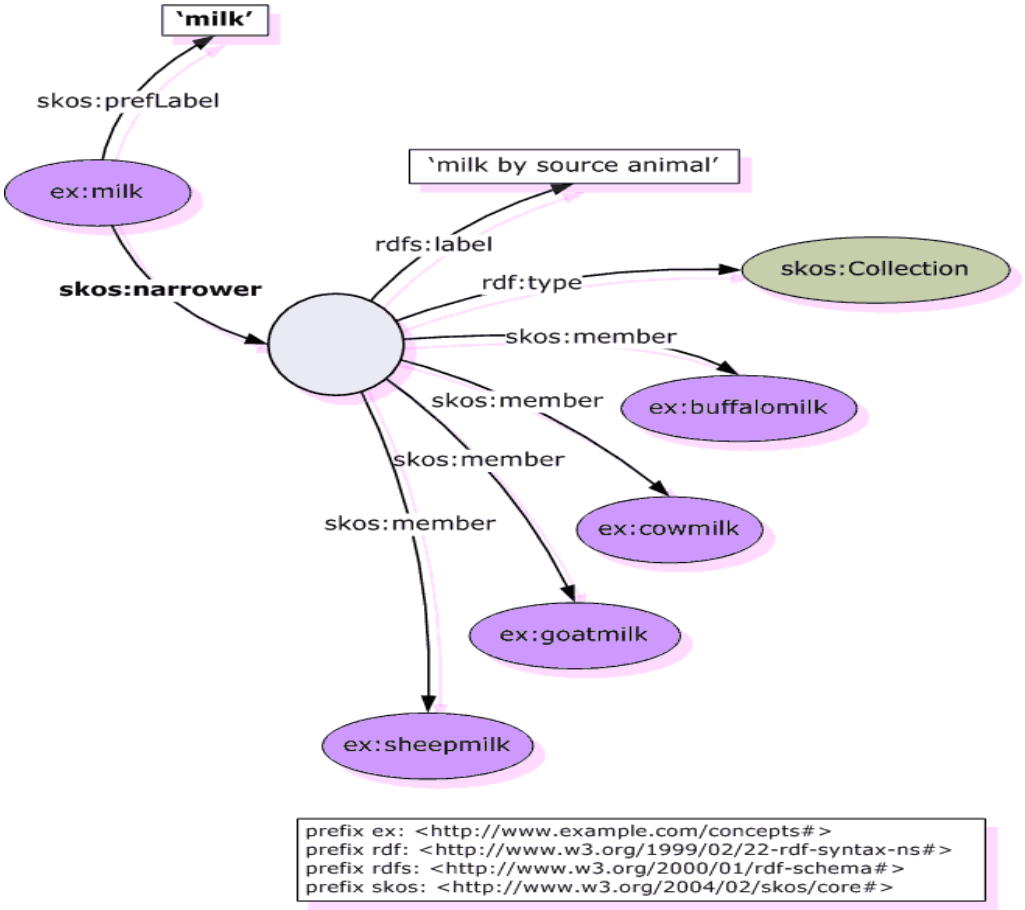


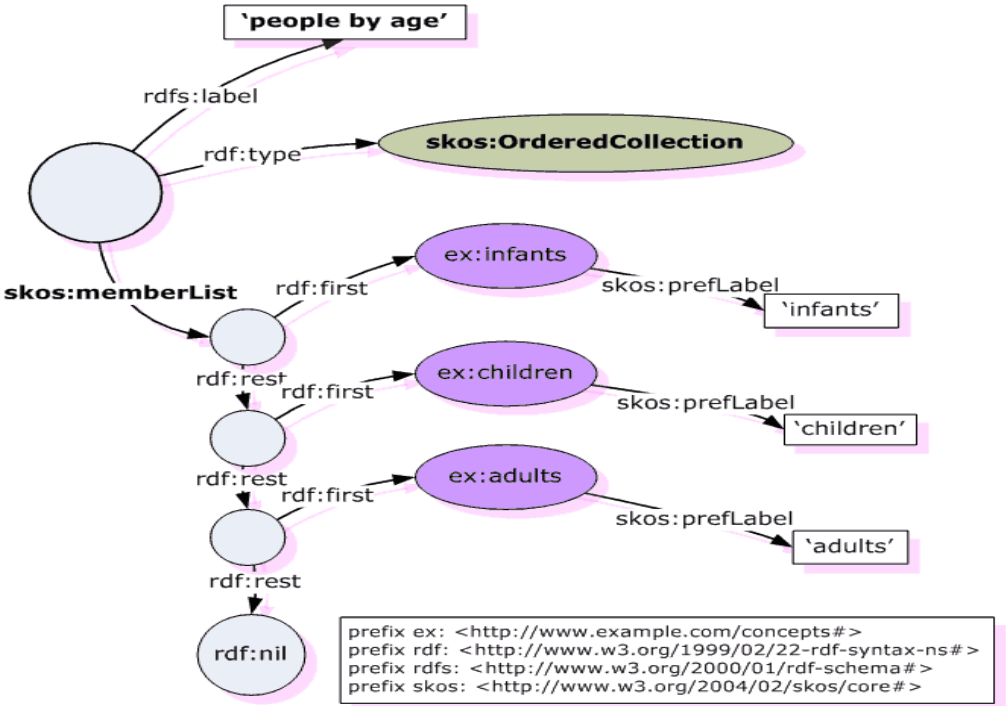
SKOS: concept schemes and collections

- Class: skos:ConceptScheme
  - skos:inScheme, skos:topConceptOf, skos:hasTopConcept
- Properties that relate concepts in different schemes:
  - skos:mappingRelation, and its “rdfs:subProperties”:
    - skos:closeMatch, skos:exactMatch, skos:relatedMatch, skos:broadMatch, skos:narrowMatch
- Classes: skos:Collection, skos:OrderedCollection
- Properties:
  - skos:member, skos:memberList



SKOS: collections

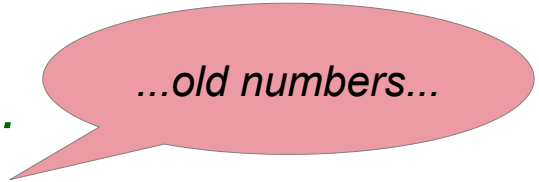


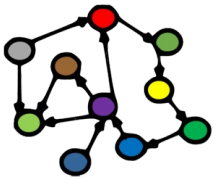


Source: www.w3c.it

schema.org

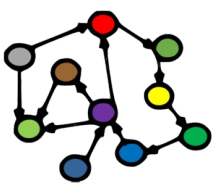
- Letting webmasters markup their pages in ways recognized by search providers such as Google, Microsoft, Yahoo and Yandex
  - ...and letting search providers improve the display of search results, enabling new tools and applications
  - schema.org has a commercial angle
- @prefix schema: <http://schema.org/> .
- Defines a hierarchy of classes
  - each with associated properties
  - markup with *Microdata, RDFa Lite, or JSON-LD...*
- *Uses: lots of web pages (> 10 000 000) that want to be searchable (but uptake is not fantastic either)*





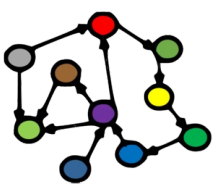
Microdata

- Not the same as schema.org
  - used to insert semantic data into HTML documents
  - comprises groups (items) of name-value pairs
  - (item, name, value) ~ (subject, predicate, object)
  - not visible as part of the page
  - typically used with schema.org types and properties
  - parsed and used extensively by search engines



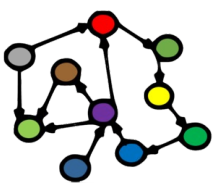
Microdata example

```
<dl itemscope>
 <dt>Title
 <dd itemprop="title">The Reality Dysfunction
 <dt>Author
 <dd itemprop="author">Peter F. Hamilton
 <dt>Publication date
 <dd><time itemprop="pubdate"
           datetime="1996-01-26">26 January 1996</time>
</dl>
```



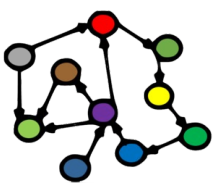
Microdata example

```
<dl itemscope
    itemtype="https://vocab.example.net/book"
    itemid="urn:isbn:0-330-34032-8">
 <dt>Title
 <dd itemprop="title">The Reality Dysfunction
 <dt>Author
 <dd itemprop="author">Peter F. Hamilton
 <dt>Publication date
 <dd><time itemprop="pubdate"
           datetime="1996-01-26">26 January 1996</time>
</dl>
```



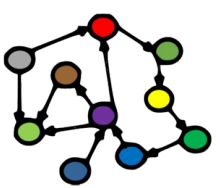
schema.org: Products

- “A product is anything that is made available for sale—for example, a pair of shoes, a concert ticket, or a car. Commodity services, like haircuts, can also be represented using this type.”
- Class: Thing → Product
- Properties:
  - general: name, description, image, sameAs, url...
  - specific: productId, brand, manufacturer, model, color, depth, width, height, weight, review, aggregateRating...
- More specific types:
  - IndividualProduct, ProductModel, SomeProducts



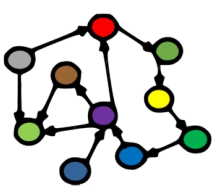
schema.org: Offers

- “An offer to transfer some rights to an item or to provide a service—for example, an offer to sell tickets to an event, to rent the DVD of a movie, to stream a TV show over the internet, to repair a motorcycle, or to loan a book.”
- Class: Thing → Intangible → Offer
- Properties:
  - general: ...as before...
  - specific: seller, itemOffered, businessModel (sell, lease, repair, dispose), price, priceCurrency, priceSpecification, acceptedPaymentMethod, availablility, warranty, validFrom, validTo...



schema.org: Data types

- Primitive data types (subclasses of DataType):
  - Boolean
  - Date (ISO 8601)
  - DateTime (also ISO 8601)
  - Number (Float, Integer)
  - Text (URI)
  - Time
- Structured values (subclasses of Thing → Intangible):
  - GeoCoordinates, GeoShape
  - OpeningHoursSpecification, PriceSpecification
  - QuantitativeValue, TypeAndQuantityNode
  - NutritionInformation...



Dublin Core (DC)

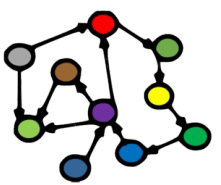
- Describing web resources (video, images, web pages...) and physical resources (books, CDs, artworks...)
- DC Metadata Element Set (DCMES, version 1.1):
  - @prefix dce: <http://purl.org/dc/elements/1.1/> .
  - 15 original properties
- DCMI (Metadata Initiative) Metadata Terms:
  - @prefix dcterms: <http://purl.org/dc/terms> .
  - more than 50 RDF properties
    - including the original 15
  - also a selection of types
- *Widely used!*

http://dublincore.org/

METADATA
Dublin Core® Metadata Initiative
INNOVATION

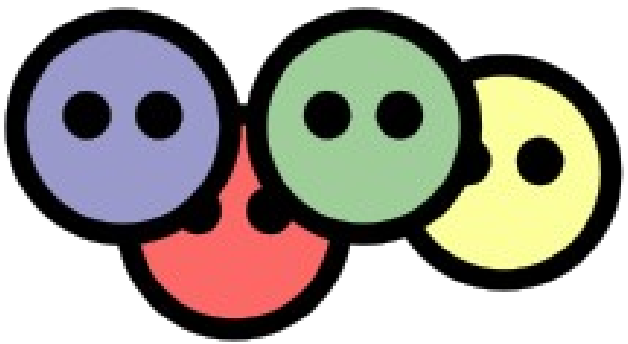
Important DCTerms

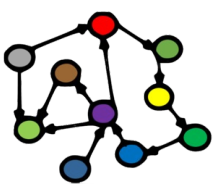
- dcterms:creator – URI for the resource (e.g., person, organisation, service...) that is primarily responsible for creating the dataset
- dcterms:title – the name of the dataset
- dcterms:description – a textual description of the dataset
- dcterms:publisher – URI for the resource that is responsible for making the dataset available
- dcterms:contributor – URI the resource that is responsible for making contributions to the dataset
- dcterms:source – URI of resource the dataset is derived from
- dcterms:date – xsd:date for a point or period of time associated with an event in the life-cycle of the resource
- dcterms:created – xsd:date of creation of the dataset
- dcterms:issued – xsd:data of publication of the dataset
- dcterms:modified – xsd:date when the dataset was changed



Friend of a Friend (FOAF)

- Creating a web of machine-readable pages describing people, the links between them and the things they create and do ... connecting social Web sites and people
- @prefix foaf: <http://xmlns.com/foaf/0.1/> .
- *RDFS*, since 2000
- Main concepts:
  - 13 classes: Agent, Person, Group, Organization, Project, Document, Image...
  - 62 properties, e.g.:
    - name, givenName, familyName, knows, based_near...
    - mbox, mbox_sha1sum, nick, msnChatId...
    - title, page, homepage, workplaceHomepage, weblog...
    - knows, maker ↔ made, publications



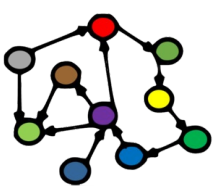


DBpedia ontology

- DBpedia Ontology:
  - 685 classes, 2795 properties
  - max depth: 5
- Available as download, SPARQL endpoint, Linked Data interface...

Instances per class

Class	Instances
Resource (overall)	4,233,000
Place	735,000
Person	1,450,000
Work	411,000
Species	251,000
Organisation	241,000

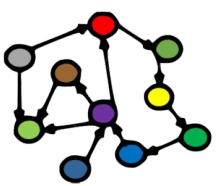


GeoNames ontology

- Vocabulary in OWL:
  - @prefix gn: <http://geonames.org/ontology#> .
  - gn:Feature class
  - 9 top-level feature codes:
    - **A** country, state, region, ...; **H** stream, lake, ...;
      L parks, area, ...; **P** city, village, ...; **R** road, railroad;
      S spot, building, farm; **T** mountain, hill, rock, ...;
      U undersea; **V** forest, heath, ...
    - 645 detailed feature codes (in a hierarchy)
  - gn:name, gn:alternateName, gn:locationMap, gn:countryCode, gn:featureClass, gn:featureCode, gn:nearbyFeatures, gn:parentADM1, gn:parentADM2, gn:parentCountry, gn:population, gn:wikipediaArticle
  - also uses properties from *geo, foaf, dcterms, cc, rdfs...*

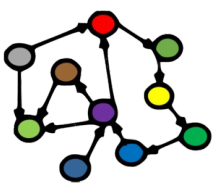
Geo (WGS84)

- A vocabulary for representing latitude, longitude and altitude according to WGS84 (World Geodetic Standard)
- @prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
- Classes: SpatialThing, Point
- Properties:
  - lat, long: latitude and longitude in decimal degrees
  - lat_long: comma-separated pair of lat and long
  - alt: altitude in meters (above *local reference ellipsoid*)
  - location: near something else (foaf:based_near)
- *Uses:*
  - *lat and long (also lat_long or point) are widely used!*



Annotating vocabulary descriptions (VANN)

- A vocabulary for annotating descriptions of vocabularies with examples and usage notes.
- @prefix vann: <http://purl.org/vocab/vann/> .
- In RDFS since 2005
- Main concepts:
  - no classes
  - six properties:
    - preferredNamespacePrefix, -Uri
    - example, usageNote
    - changes (delta from previous version)
    - termGroup

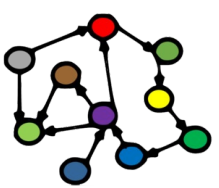


Vocabulary Status (VS)

- An RDF vocabulary for describing the status of vocabulary terms on the Web of Data
  @prefix vs: <http://www.w3.org/2003/06/sw-vocab-status/ns#> .
- Main concepts:
  - no classes
  - three properties:
    - term_status (e.g., *unstable*, *testing*, *stable* or *archaic*)
    - moreinfo
    - userdocs

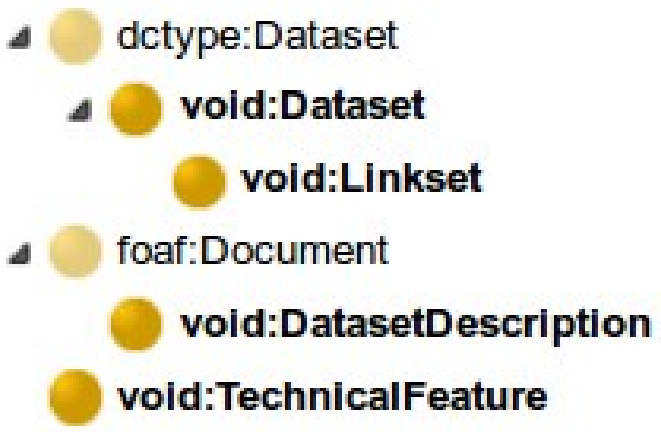
Also: Vocabulary of a Friend (VoaF)

- used by the Linked Open Vocabulary (LOV) site



Vocabulary of Interlinked Datasets (VoID)

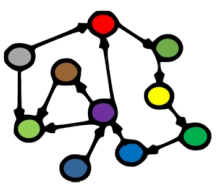
- Expressing metadata about RDF datasets.
  - *general metadata:* following DC and FOAF
  - *access metadata:* how RDF data can be accessed
  - *structural metadata:* the structure and schema of datasets, useful for querying and data integration
  - description of *links between datasets:* how multiple datasets are related and can be used together
- @prefix void: <http://rdfs.org/ns/void#> .
  - in RDFS since spring 2013
- 4 classes:
    - Dataset, Linkset, DatasetDescription...
- 27 properties



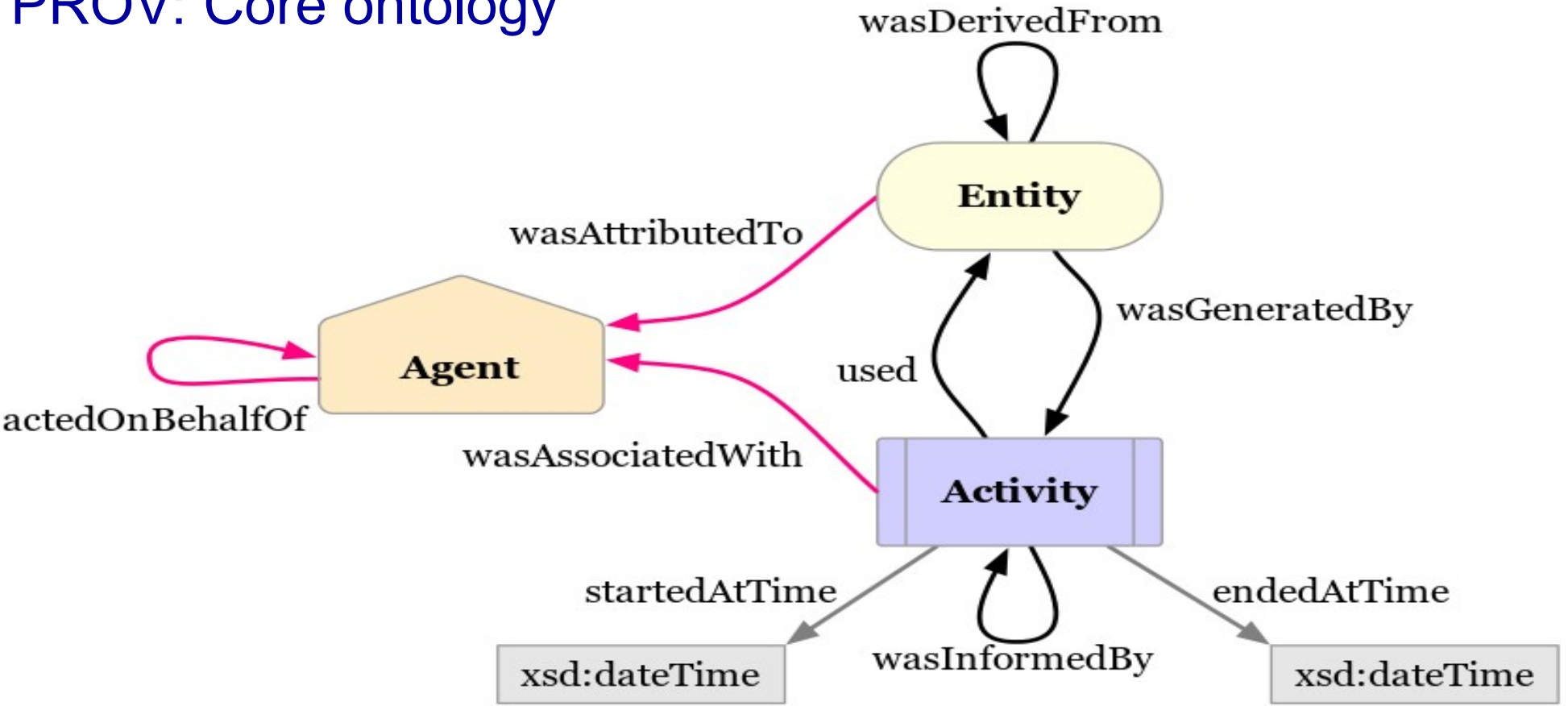
http://vocab.deri.ie/void

Provenance Interchange (PROV)

- A vocabulary for describing the origin of physical, conceptual, and digital entities, including RDF datasets ... it can be used to assess their quality, reliability and trustworthiness
- @prefix prov: <http://www.w3.org/ns/prov#> .
  - in RDFS since spring 2013
- Components:
  - data model (PROV-DM), *ontology (PROV-O)*
  - XML Schema (PROV-XML), notation
  - *mapping from DC*
- LOV: 50 classes, 83 properties, 1 instance
- PROV-O: 30 classes, 49 properties



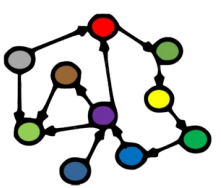
PROV: Core ontology

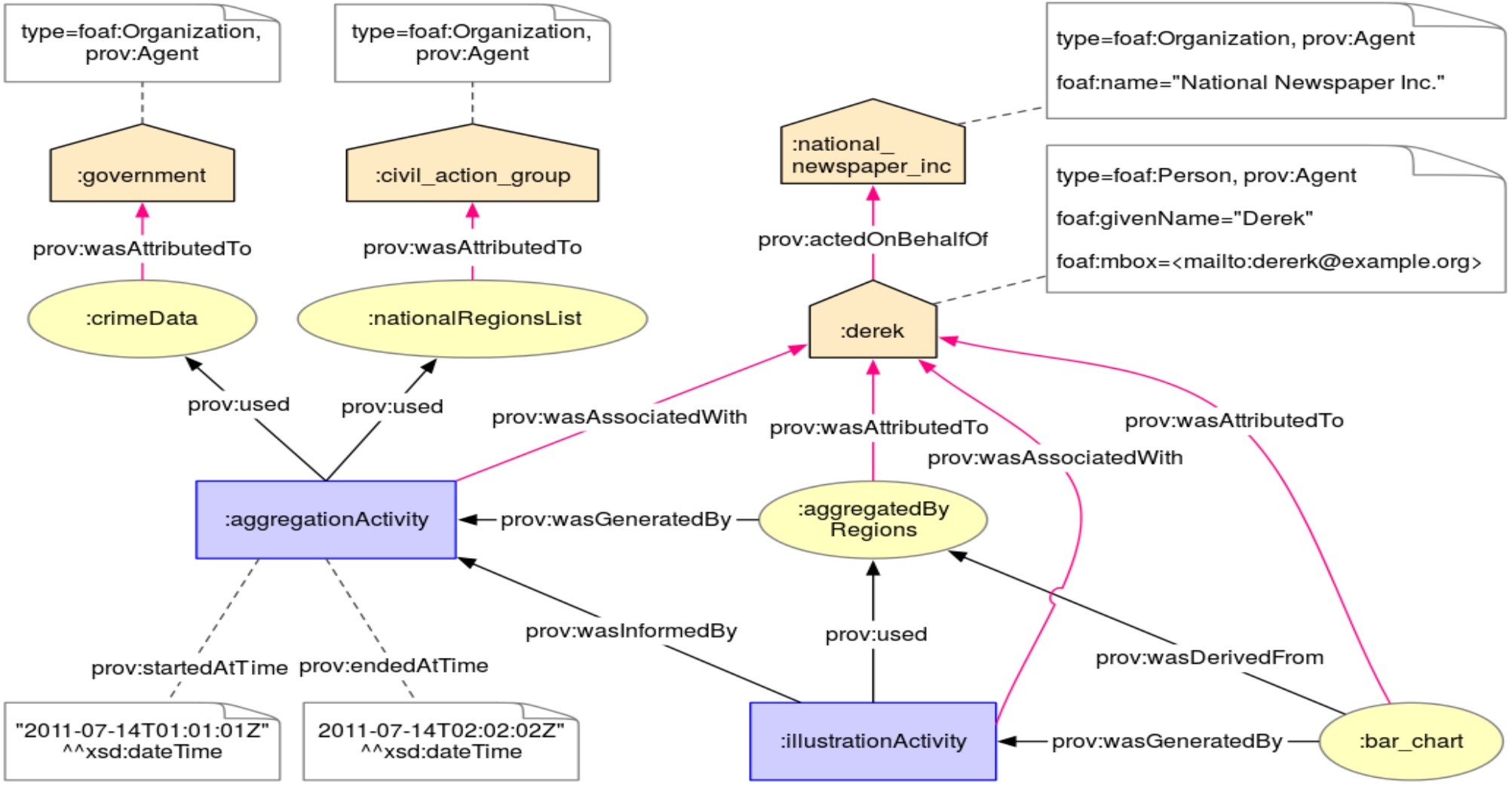


Source: https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/

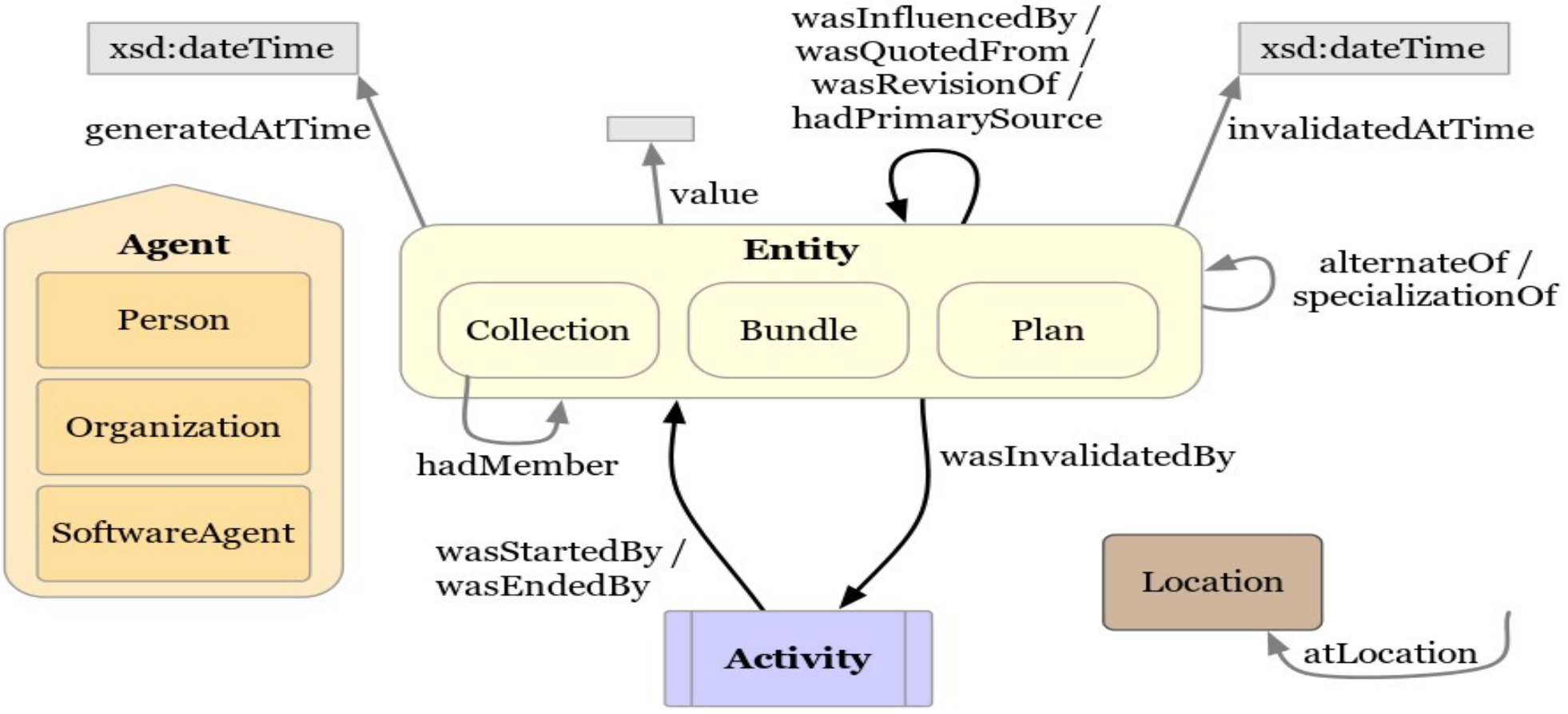
PROV: Entity, activity and agent

- Entity
  - a physical, digital, conceptual, or other kind of thing with some fixed aspects; entities may be real or imaginary
- Activity
  - occurs over a period of time and acts upon or with entities
  - may include consuming, processing, transforming, modifying, relocating, using, or generating entities
- Agent
  - bears some form of responsibility
  - for an activity taking place, for the existence of an entity, or for another agent's activity




type=foaf:Organization, prov:Agent
type=foaf:Organization, prov:Agent
type=foaf:Organization, prov:Agent
foaf:name="National Newspaper Inc."
:government
:civil_action_group
:national_ newspaper_inc
type=foaf:Person, prov:Agent
foaf:givenName="Derek"
foaf:mbox=<mailto:dererk@example.org>
prov:wasAttributedTo
prov:wasAttributedTo
prov:actedOnBehalfOf
:crimeData
:nationalRegionsList
:derek
prov:used
prov:used
prov:wasAssociatedWith
prov:wasAttributedTo
prov:wasAttributedTo
prov:wasAssociatedWith
:aggregationActivity
prov:wasGeneratedBy
:aggregatedBy Regions
prov:wasInformedBy
prov:used
prov:wasDerivedFrom
prov:startedAtTime
prov:endedAtTime
"2011-07-14T01:01:01Z" ^^xsd:dateTime
2011-07-14T02:02:02Z" ^^xsd:dateTime
:illustrationActivity
prov:wasGeneratedBy
:bar_chart

PROV: More specific terms

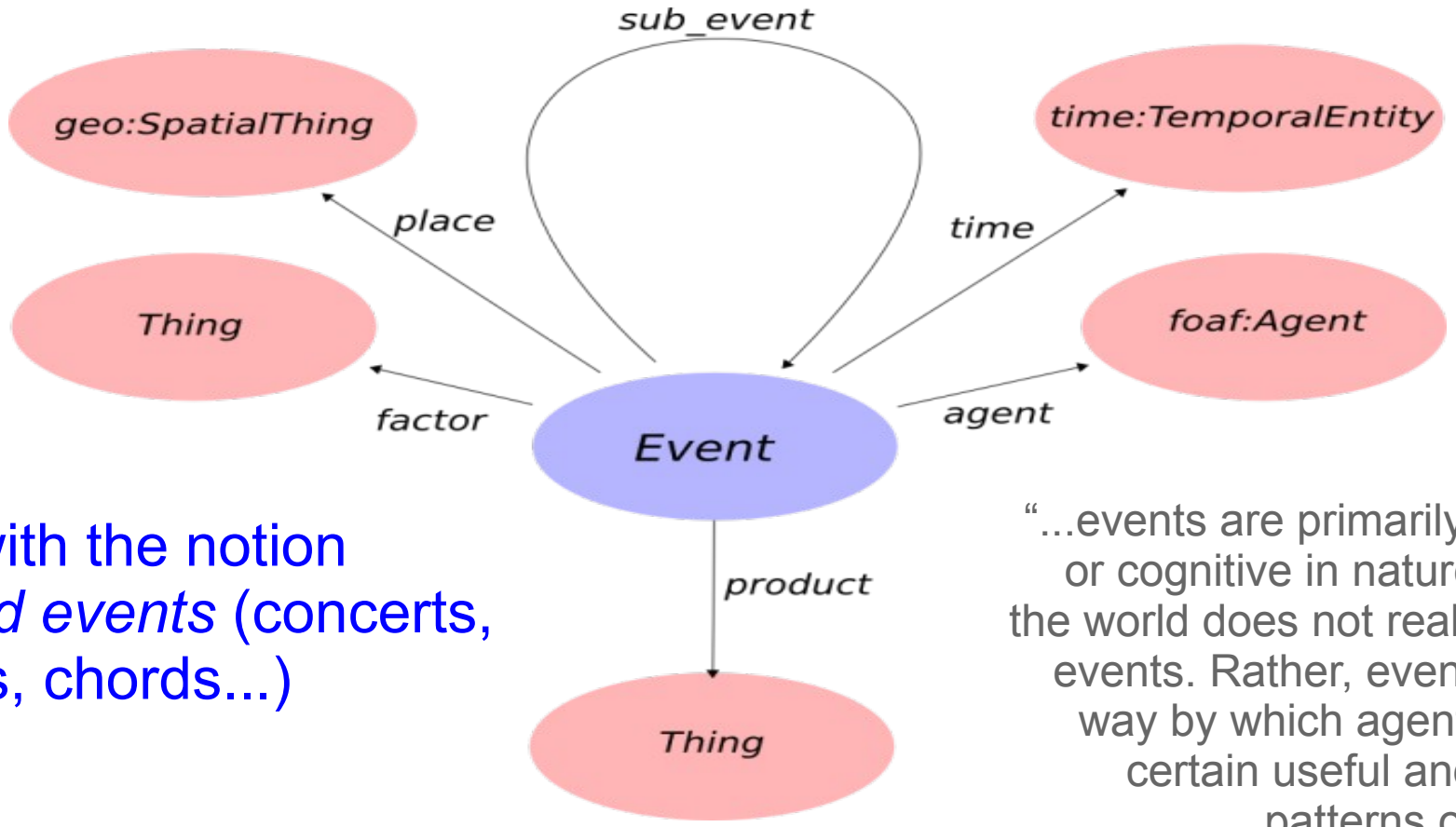


Creative commons (CC)

- The Creative Commons Rights Expression Language (CC REL) lets you describe copyright licenses in RDF
- @prefix cc: <http://creativecommons.org/ns> .
- Classes:
  - Work, License, Jurisdiction
  - Permission (Reproduction, Distribution, DerivativeWorks, Sharing)
  - Requirement (Notice, Attribution, ShareAlike, SourceCode, Copyleft, LesserCopyLeft)
  - Prohibition (CommercialUse, HighIncomeNationUse)
- Properties:
  - license, permits, requires, prohibits, legalCode...

Source: http://creativecommons.org/schema.rdf
Licences: https://creativecommons.org/about/cclicenses/

Event ontology (event)



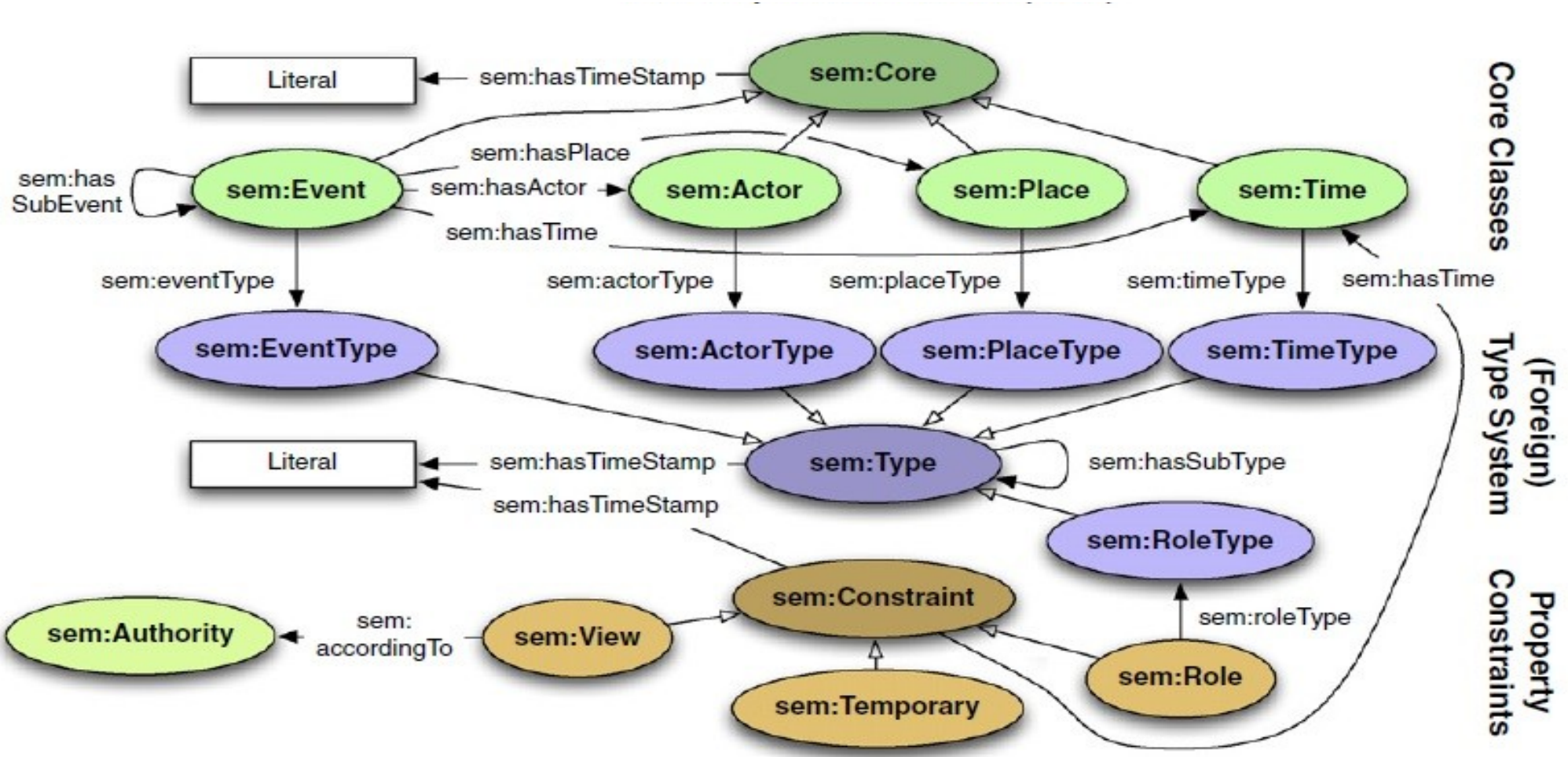
Deals with the notion of *reified events* (concerts, festivals, chords...)

"...events are primarily linguistic or cognitive in nature. That is, the world does not really contain events. Rather, events are the way by which agents classify certain useful and relevant patterns of change."

@prefix event: <http://purl.org/NET/c4dm/event.owl#> .

http://motools.sourceforge.net/event/event.html

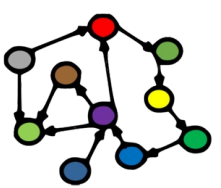
Simple Event Model (SEM)



https://semanticweb.cs.vu.nl/2009/11/sem/

Time ontology in OWL (time, OWL-time)

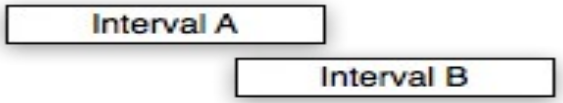
- Describing the temporal content of Web pages and the temporal properties of Web services.
- @prefix time: <http://www.w3.org/2006/time#> .
- 9 classes:
  - TemporalEntity: either Instant or Interval
  - Interval: ProperInterval and DateTimeInterval
  - the other classes are for complex data types
- More than 40 properties:
  - 18 describe relations between TemporalEntities:
    - before, after, inside, hasBeginning, hasEnd...
  - the rest describe the Instants and Intervals
    - mostly in terms of XSD types...
- Also provides a TimeZone ontology



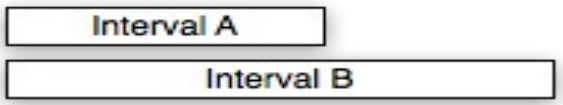
A is before B or
B is after A

A meets B or
B is met by A

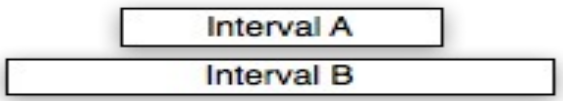
A overlaps with B or
B is overlapped by A



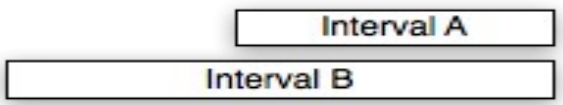
A starts B or
B is started-by A



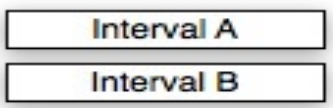
A during B or
B contains A



A finishes B or
B is finished-by A

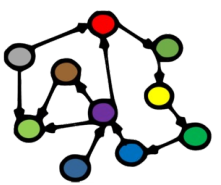


A and B are cotemporal



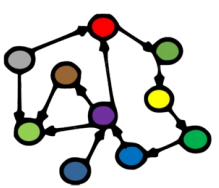
Timeline Ontology (tl)

- Annotating sections of a signal, video, score, work or any temporal object
  - often combined with the event and time ontologies
- @prefix tl: <http://purl.org/NET/c4dm/timeline.owl#> .
- Classes:
  - tl:TimeLine (of time:Intervals and time:Instants)
    - continuous or discrete, physical, relative or abstract
  - tl:TimeLineMaps between tl:TimeLines
    - sampling, shifting, windowing
- Properties:
  - timeline, domainTimeLine, rangeTimeLine
  - delay, sampleRate, windowsLength

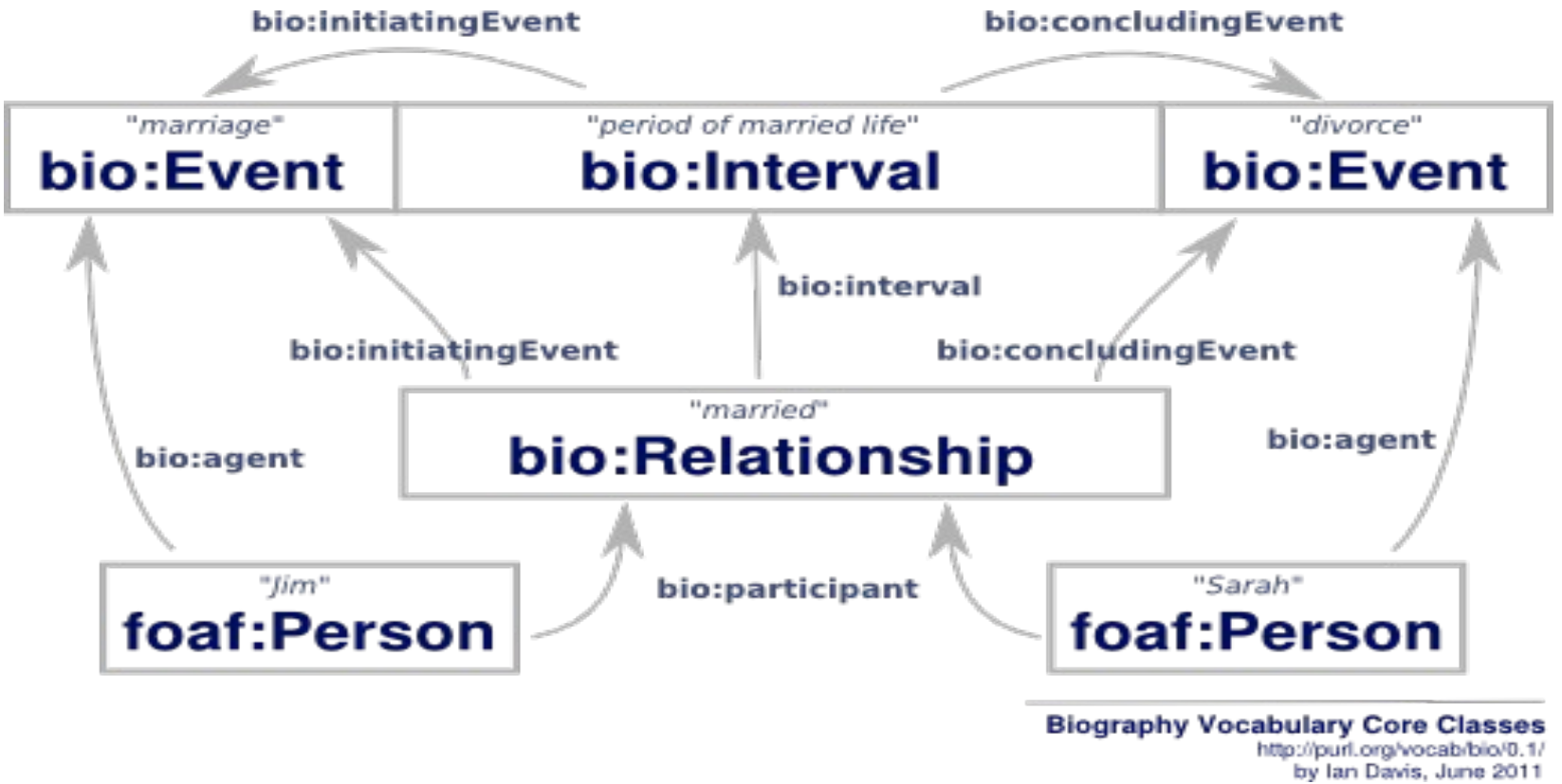


Biographical Information (BIO)

- Terms for finding out more about people and their backgrounds... some cross-over into genealogical information... describe a person's life as a series of interconnected key events
- @prefix bio: <http://purl.org/vocab/bio/0.1/> .
  - in RDFS since spring 2010
  - builds on, e.g., the Event and Time ontologies

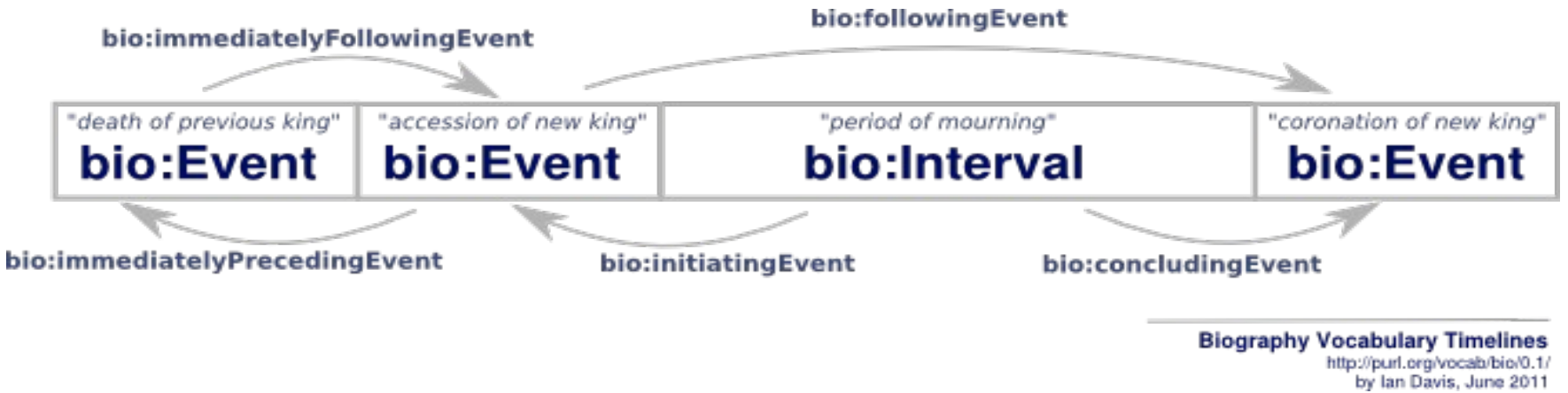


Biographical Information (BIO)



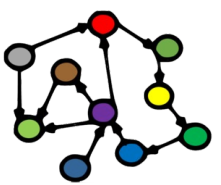
http://vocab.org/bio/0.1/.html

Biographical Information (BIO)



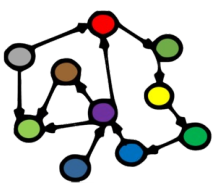
Bibliographic Ontology (bibo)

- Describing bibliographic things on the semantic Web in RDF, can be used as a citation ontology, as a document classification ontology, or as a way to describe documents in RDF
- @prefix bibo: <http://purl.org/ontology/bibo/1.3/> .
- 69 classes:
  - Article, Book, Chapter, Conference, Event, Film, Interview, Manual, Map, Newspaper, Patent, Slideshow, Website...
- More than 100 properties...
  - authorList, cites, producer, status
- 14 individuals:
  - document states, thesis degrees
- *Also uses dcterms, event and foaf*

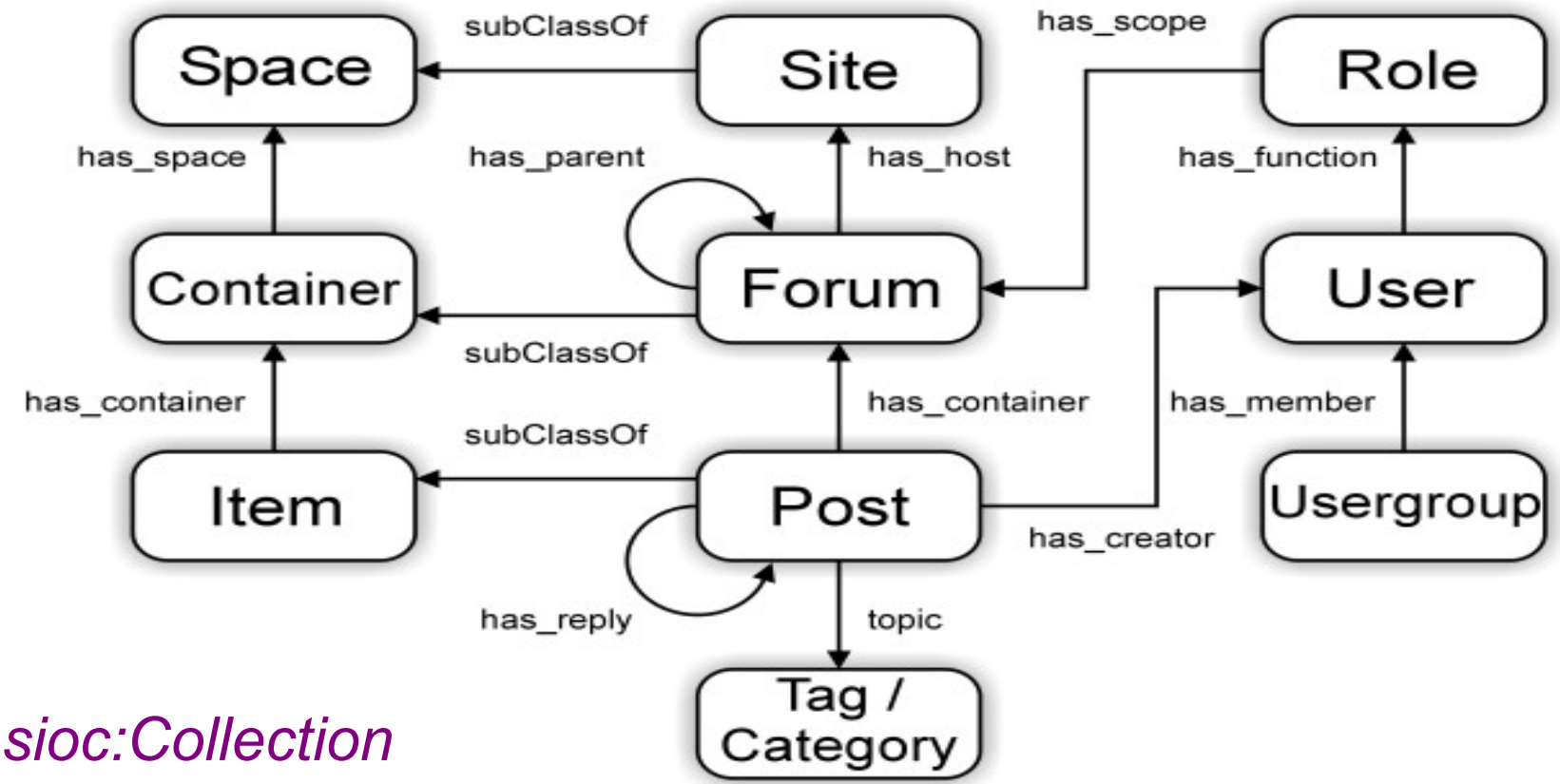


Semantic Interlinked Online Communities

- Semantic Interlinked Online Communities (SIOC):
  - Using Semantic Web technologies to describe the information that Online community sites (weblogs, message boards, wikis...) have about their structure and contents
  - also to find related information and new connections between content items and other community objects.
- @prefix sioc: <http://rdfs.org/sioc/ns#> .
- 12 classes and around 50 properties



SIOC: classes

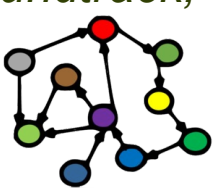


Also sioc:Collection and sioc:Thread.

http://vocab.org/bio/0.1/.html

Music Ontology (mo)

- Providing main concepts and properties for describing music (artists, albums, tracks...) on the Semantic Web.
- @prefix mo: <http://purl.org/ontology/mo/> .
- 54 classes, e.g.:
  - MusicalWork, MusicalExpression, MusicalManifestation, MusicalItem *(ligner: verk, fremførelse, utgivelse, eksemplar)*
  - CD, Composition, Festival, Genre, Instrument, Label, Lyrics, MusicArtist, Performance, Release, Torrent, Vinyl...
- 153 properties: release_type, release_status
- 13 individuals:
  - *album, audiobook, compilation, ep, interview, live, remix, single soundtrack, spokenword* (rdf:type mo:ReleaseType)
  - *bootleg, official, promotion* (rdf:type mo:ReleaseStatus)



Next week:
Description Logic (DL)