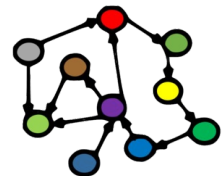


Welcome to INFO216:
Knowledge Graphs
Spring 2023

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

Session 9: Vocabularies (and ontologies)

- Themes:
 - semantic vocabularies
 - *SKOS, schema.org, DC, FOAF, geo, CC, PROV, Event, Time, SIOC...*
 - also the open KG ontologies
 - *Wikidata, DBpedia, GeoNames...*



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 (supplementary)
- Resources in the wiki <http://wiki.uib.no/info216>:
 - supplementary links in the portal, including
 - Linked Open Vocabularies (LOV)
<http://lov.okfn.org/dataset/lov/>

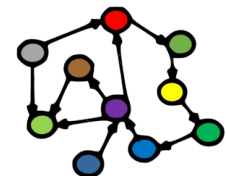


THE KNOWLEDGE GRAPH
COOKBOOK
RECIPES THAT WORK



ANDREAS BLUMAUER
AND HELMUT NAGY

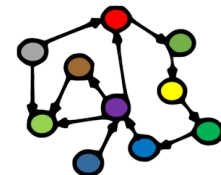
1st edition, 2020



Why vocabularies (and ontologies)

Terms

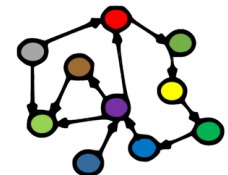
- *Semantic vocabularies / ontologies*
 - graphs/datasets (in RDFS, OWL...) that define:
 - standard URIs for *types of resources*
 - standard URIs for *properties*
 - standard URIs for some important *individuals*
 - standard datatypes (identified by URIs) for *literals*
- *Linked open semantic datasets / 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

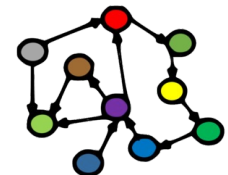
- *Semantic vocabularies / ontologies*
 - graphs/datasets (in RDFS, OWL...) that define:
 - standard URIs for *types of resources*
 - standard URIs for *properties*
 - standard URIs for some important *individuals*
 - standard datatypes (identified by URIs) for *literals*
- *Linked open semantic datasets / 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



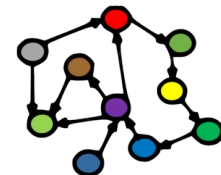
Why do we need vocabularies and ontologies?

- Shared, well-defined terms (dereferencable URIs) for types, properties and some individuals that can be used to represent a domain
- 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*

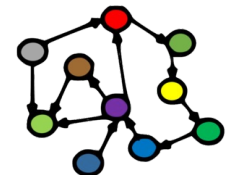
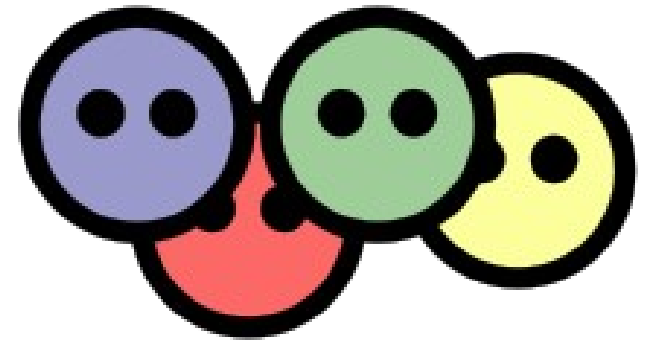


Semantic vocabularies

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

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/>> .
- 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



:Andreas_Lothe_Opdahl foaf:interest :MuellerReport .

foaf:interest rdfs:domain foaf:Agent ; rdfs:range Document . # FOAF rules

=> :AndreasLOpdahl a foaf:Agent .

:MuellerReport a foaf:Document .

:Robert_Mueller foaf:publications :MuellerReport .

foaf:publications rdfs:domain foaf:Person ; rdfs:range foaf:Document . # FOAF rules

=> :Robert_Mueller a foaf:Person, geo:SpatialThing .

:MuellerReport a foaf:Document .

:MuellerReport

foaf:primaryTopic dbpedia:Donald_Trump_2016_presidential_campaign .

foaf:primary_topic rdfs:domain foaf:Document ; # FOAF rules

a owl:FunctionalProperty ;

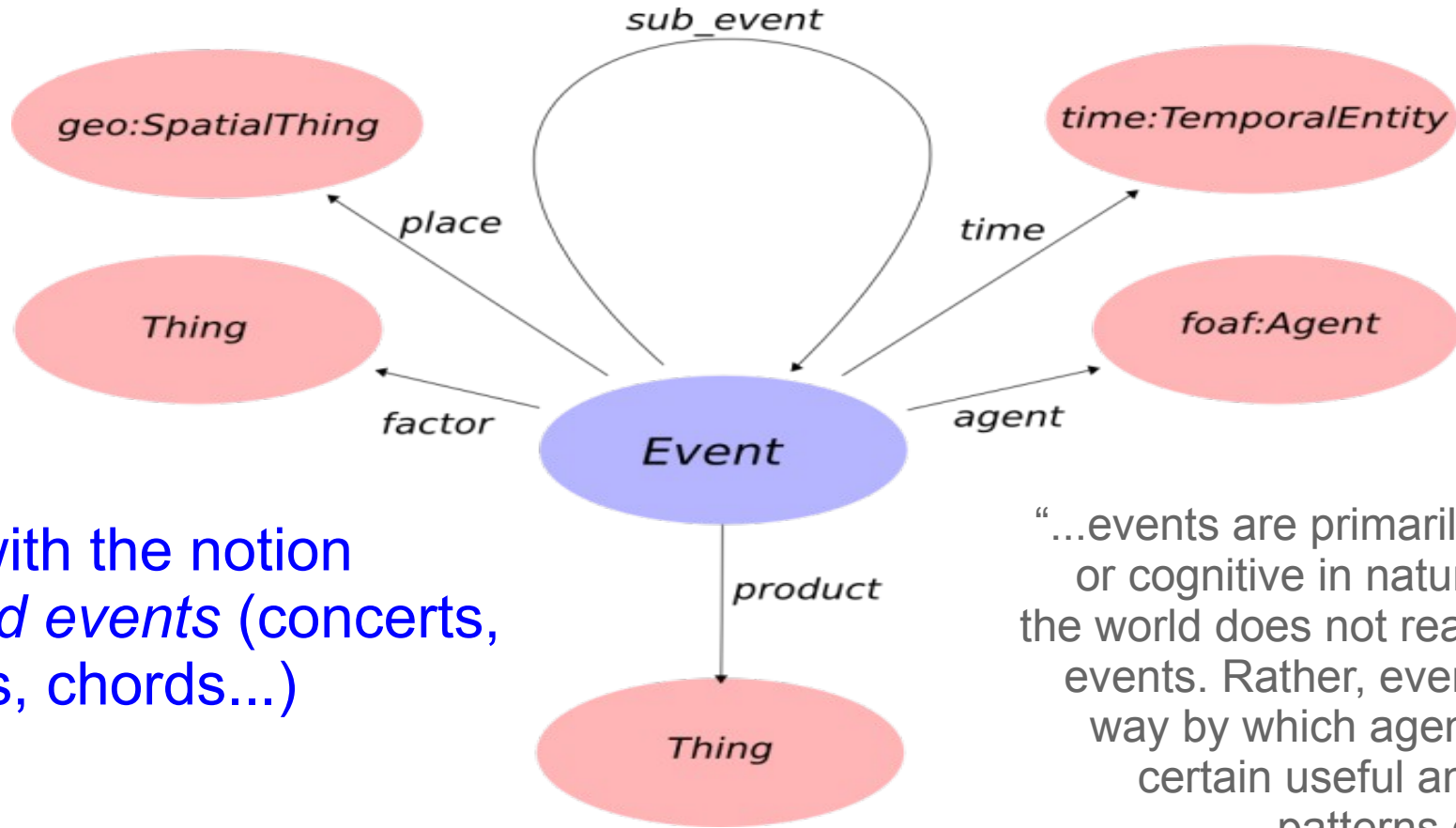
owl:inverseOf foaf:is_primary_topic_of .

=> dbpedia:Donald_Trump_2016_presidential_campaign a foaf:Document .

dbpedia:Donald_Trump_2016_presidential_campaign
foaf:member dbpedia:Roger_Stone, dbpedia:Rudy_Giuliani,
dbpedia:Paul_Manafort, dbpedia:George_Papadopoulos,
dbpedia:Steve_Bannon, dbpedia:Michael_D._Cohen_(lawyer),
dbpedia:Michael_Flynn, dbpedia:Rick_Gates_(political_consultant) .
foaf:member rdfs:domain foaf:Group ; rdfs:range foaf:Agent .

dbpedia:Roger_Stone foaf:knows dbpedia:Donald_Trump .

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.”

dbpedia:Donald_Trump_2016_presidential_campaign a event:Event .

:MuellerInvestigation a event:Event ;

event:actor :Robert_Mueller ;

event:factor dbpedia:Donald_Trump_2016_presidential_campaign ,
dbpedia:Roger_Stone, dbpedia:Rudy_Giuliani,
dbpedia:Paul_Manafort, dbpedia:George_Papadopoulos,
dbpedia:Steve_Bannon, dbpedia:Michael_D._Cohen_(lawyer),
dbpedia:Michael_Flynn, dbpedia:Rick_Gates_(political_consultant) .

event:product :MuellerReport ;

event:location :USA ;

event:time :MuellerInvestigationPeriod .

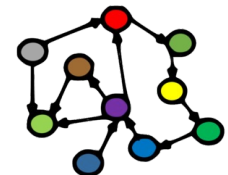
event:actor rdfs:domain event:Event ; rdfs:range foaf:Agent .

event:location rdfs:domain event:Event ; rdfs:range geo:SpatialThing .

event:time rdfs:domain event:Event ; rdfs:range time:TemporalEntity .

Time ontology in OWL (time, OWL-time)

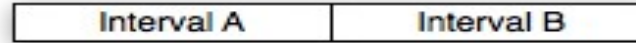
- Describing the temporal content of Web pages and the temporal properties of Web services.
- @prefix time: <<http://www.w3.org/2006/time#>> .
- 9 classes:
 - TemporalEntity: either Instant or Interval
 - Interval: ProperInterval and DateTimeInterval
 - the other classes are for complex data types
- 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



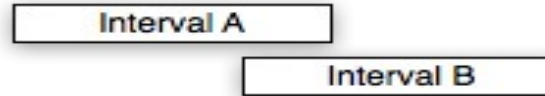
A is before B or
B is after A



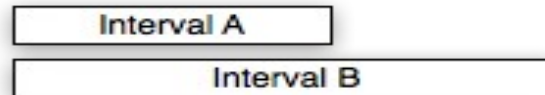
A meets B or
B is met by A



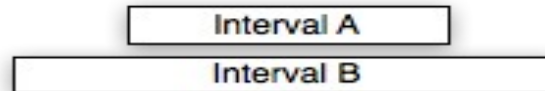
A overlaps with B or
B is overlapped by A



A starts B or
B is started-by A



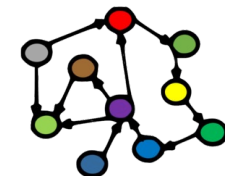
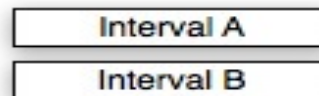
A during B or
B contains A



A finishes B or
B is finished-by A

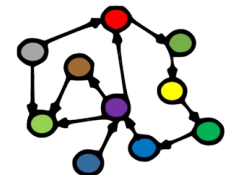


A and B are cotemporal



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!*

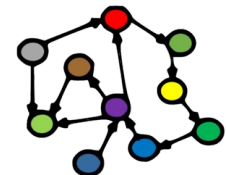


```
dbpedia:Donald_Trump_2016_presidential_campaign a geo:SpatialThing ;
    geo:location dbpedia:Manhattan, dbpedia:Trump_Tower .
geo:location rdfs:domain geo:SpatialThing . # GEO rules
geo:location rdfs:subPropertyOf foaf:based_near .
=> dbpedia:Trump_Tower a geo:SpatialThing .
```

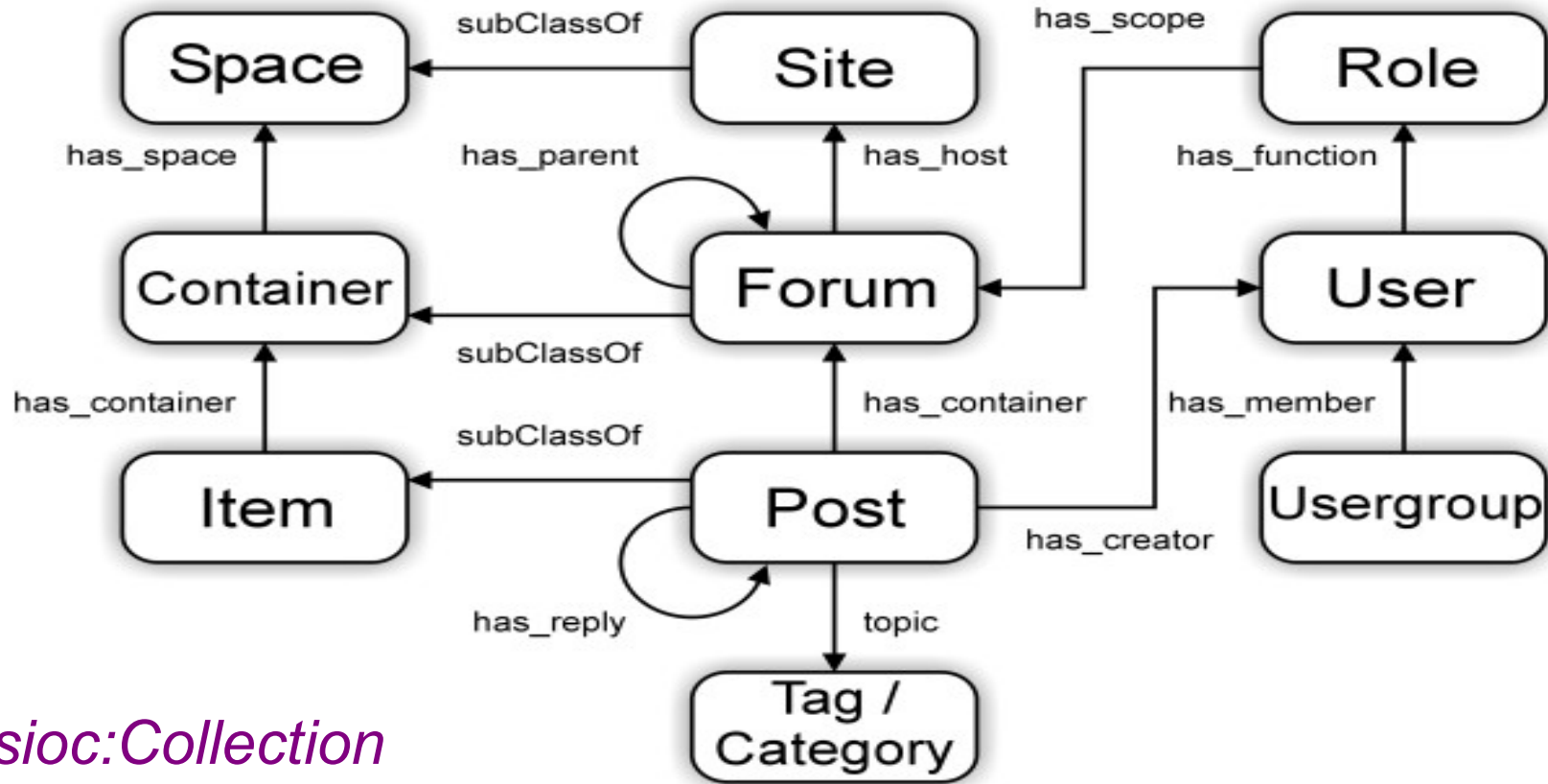
```
dbpedia:Trump_Tower geo:lat 40.7625 ; geo:long -73.9738 .
geo:lat rdfs:domain geo:SpatialThing . # GEO rules
geo:long rdfs:domain geo:SpatialThing .
=> dbpedia:Trump_Tower a geo:SpatialThing .
    dbpedia:Donald_Trump_2016_presidential_campaign
        foaf:based_near dbpedia:Trump_Tower .
```

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.*

twitter:AndreasLOpdahl a sioc:UserAccount ;
sioc:name “@AndreasLOpdahl” ;
sioc:account_of :AndreasLotheOpdahl ;
sioc:follows twitter:realDonaldTrump .


twitter:realDonaldTrump a sioc:UserAccount ;
sioc:name “@realDonaldTrump” ;
sioc:account_of :DonaldTrump .




A screenshot of a tweet from Donald J. Trump. The tweet text reads: "All that time and money spent, and it turns out that the Mueller Investigation was a total SCAM!". The tweet was posted at 10:27 PM on Sep 12, 2020. It has 102.2K likes and 22.2K replies. The interface includes a profile picture of Donald Trump, his name "Donald J. Trump" with a verified badge, his handle "@realDonaldTrump", and a "Follow" button. There are also icons for retweeting, replying, and a link icon. At the bottom, there is a button that says "Read 22.2K replies".

 **Donald J. Trump** 
@realDonaldTrump · [Follow](#) 

All that time and money spent, and it turns out that the Mueller Investigation was a total SCAM!

10:27 PM · Sep 12, 2020 

 102.2K  Reply  Copy link

[Read 22.2K replies](#)

twitter:tweet20200912

a sioc:Post ;

sioc:about :MuellerInvestigation ;

sioc:content “All that time and money spent, and it turns out that
the Mueller Investigation was a total SCAM!” ;

sioc:has_creator twitter:realDonaldTrump ;


sioc:delivered_at “20200912T222753+0100” .

twitter:MAGA_MAN

a sioc:UserAccount ;

sioc:name “@CNNIsFakeNews1” ;


sioc:likes twitter:tweet20200912 .




The image shows a screenshot of a tweet from Donald J. Trump. The tweet text is "All that time and money spent, and it turns out that the Mueller Investigation was a total SCAM!". The tweet is dated 10:27 PM on Sep 12, 2020. It has 102.2K likes and 22.2K replies. The user's name is Donald J. Trump, with the handle @realDonaldTrump and a verified account. The user's profile picture is visible. The tweet is displayed in a light blue border with a white background.

Donald J. Trump 
@realDonaldTrump · [Follow](#) 

All that time and money spent, and it turns out that the
Mueller Investigation was a total SCAM!

10:27 PM · Sep 12, 2020 

 102.2K  Reply  Copy link

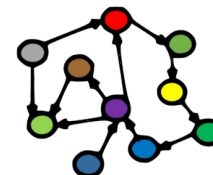
[Read 22.2K replies](#)

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><http://purl.org/dc/terms>> .
 - more than 50 RDF properties
 - including the original 15
 - also a selection of types
- *Widely used!*

Important DCTerms

- **dcterms:title** – the name of the dataset
- **dcterms:description** – a textual description of the dataset
- **dcterms:subject** – topic of the resource, represented using keywords, key phrases, or classification codes: using a controlled vocabulary is recommended
- **dcterms:creator** – URI for the resource (e.g., person, organisation, service...) that is primarily responsible for creating the dataset
- **dcterms:contributor** – URI the resource that is responsible for making contributions to the dataset
- **dcterms:publisher** – URI for the resource that is responsible for making the dataset available
- **dcterms:source** – URI of resource the dataset is derived from
- **dcterms:created**, **dcterms:issued**, **dcterms:modified** – xsd:date when the dataset was created, published, or changed
- **dcterms:date** – xsd:date for a time point or duration in the life-cycle of the resource



:MuellerReport dc:contributor :Robert_Mueller .

dc:contributor rdfs:domain foaf:Document ; rdfs:range foaf:Person . # *not* a DC rule

:MuellerReport

dc:creator :MuellerInvestigation ;

dc:publisher :USDepartmentOfJustice .

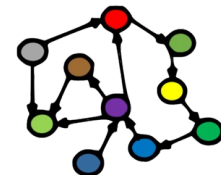
:MuellerReport

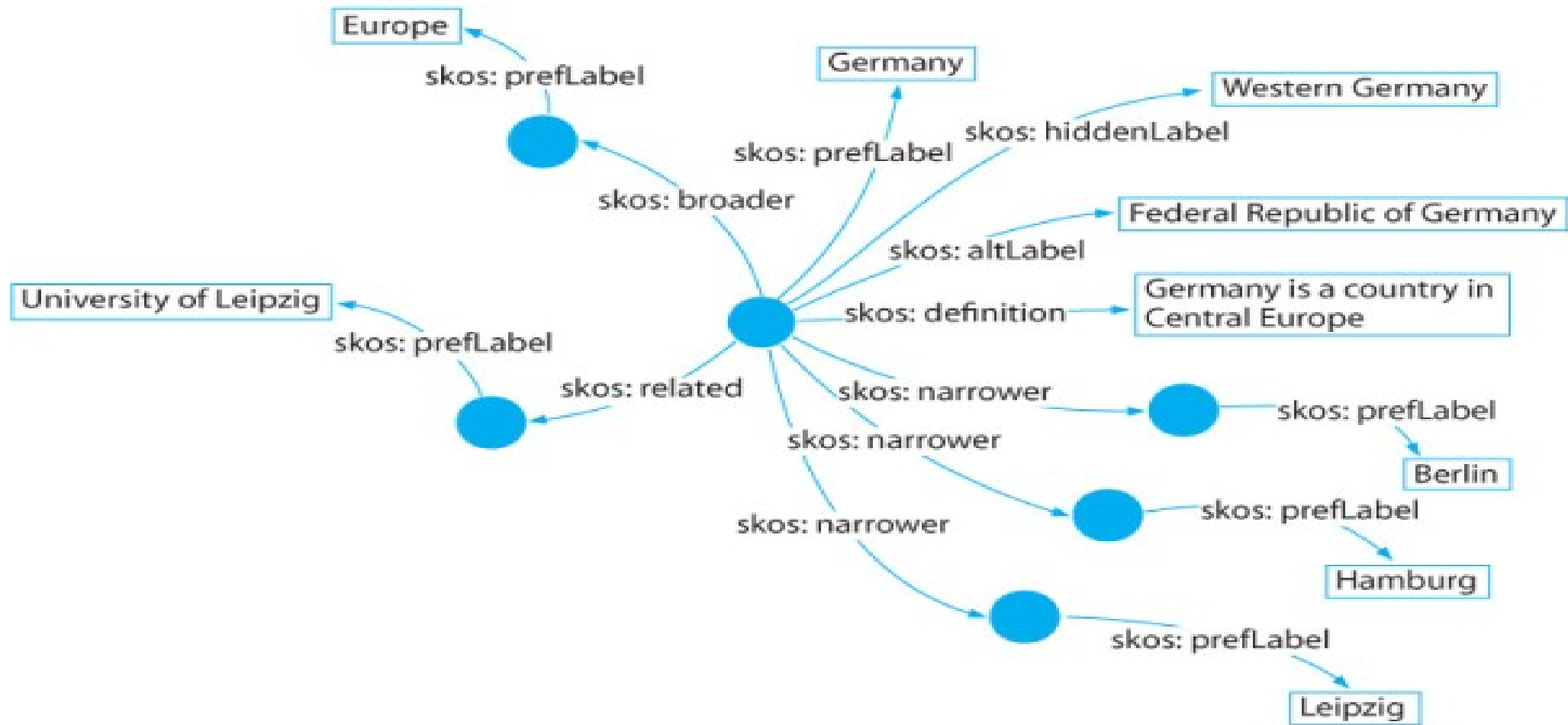
dc:subject :ElectionInterference, :Conspiracy, :ObstructionOfJustice .

dc:subject rdfs:domain foaf:Document ; rdfs:range skos:Concept . # *not* a DC rule

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*





cnet:Conspiracy # examples from ConceptNet

skos:prefLabel "conspiracy"@en, "conspiración"@es, "cabale"@fr ;

skos:altLabel "cabal"@en, "complot"@fr, "conspiration"@fr ;

skos:broader cnet:Agreement, cnet:Organization, cnet:Plot ;

skos:narrower cnet:ConspiracyOfSilence, cnet:GunpowderConspiracy ;

skos:related cnet:Conspirator .

skos:broader rdfs:domain skos:Concept ; rdfs:range skos:Concept . # SKOS rules

skos:narrower rdfs:domain skos:Concept ; rdfs:range skos:Concept .

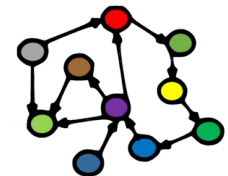
skos:related rdfs:domain skos:Concept ; rdfs:range skos:Concept .

=> cnet:Conspiracy a skos:Concept .

cnet:Conspirator a skos:Concept .

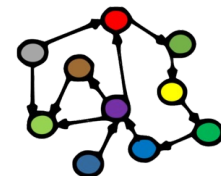
SKOS: describing concepts

- 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*



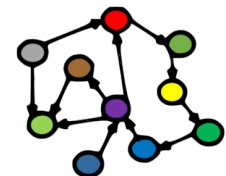
SKOS: property naming

- 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

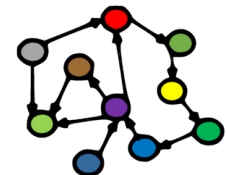
- Class: `skos:ConceptScheme`
 - `skos:inScheme`, `skos:topConceptOf`, `skos:hasTopConcept`
- Properties that relate concepts in different schemes:
 - `skos:mappingRelation`, and its “`rdfs:subProperties`”:
 - `skos:closeMatch`, `skos:exactMatch`, `skos:relatedMatch`,
`skos:broadMatch`, `skos:narrowMatch`



schema.org

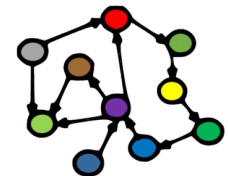
- Letting webmasters markup their pages in ways recognized by search providers such as Google, Microsoft, Yahoo and Yandex
 - ...and letting search providers improve the display of search results, enabling new tools and applications
 - schema.org has a commercial angle
- @prefix schema: <<http://schema.org/>> .
- 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)*

...old numbers...



Microdata

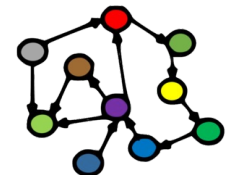
- Not the same as schema.org
 - used to insert semantic data into HTML documents
 - comprises groups (items) of name-value pairs
 - (item, name, value) ~ (subject, predicate, object)
 - not visible as part of the page
 - typically used with schema.org types and properties
 - parsed and used extensively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



Microdata example

...
Title The Reality Dysfunction
Author Peter F Hamilton
Publication date 26 January 1966
...

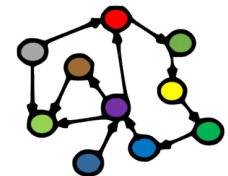
```
<dl itemscope>  
  <dt>Title </dt>  
  <dd itemprop="title">The Reality Dysfunction </dd>  
  <dt>Author </dt>  
  <dd itemprop="author">Peter F. Hamilton </dd>  
  <dt>Publication date </dt>  
  <dd><time itemprop="pubdate"  
    datetime="1996-01-26">26 January 1996</time></dd>  
</dl>
```



Microdata example

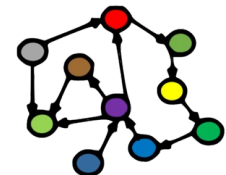
```
<dl itemscope
  itemtype=
    "https://vocab.example.net/book"
  itemid="urn:isbn:0-330-34032-8">
  <dt>Title
  <dd itemprop="title">The Reality Dysfunction
  <dt>Author
  <dd itemprop="author">Peter F. Hamilton
  <dt>Publication date
  <dd><time itemprop="pubdate"
    datetime="1996-01-26">26 January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

```
...
Title           The Reality Dysfunction
Author         Peter F Hamilton
Publication date 26 January 1966
...
```



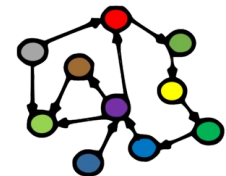
schema.org: Products

- “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**



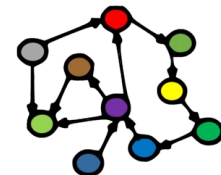
schema.org: Offers

- “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, availability, 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...

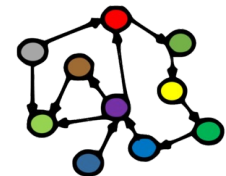


DBpedia ontology

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

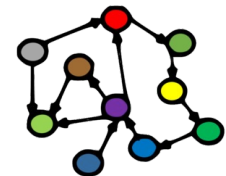
Instances per class

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



Wikidata ontology

- WikiProject Ontology:
 - reaching deep into the nature of being, becoming, existence, and reality
 - ...applying those insights during Wikidata's maintenance tasks
 - support a broad semantic interoperability between notable ontologies like DOLCE, BFO, SUMO, Lemon, RDA, etc.
 - *DUL: Descriptions and Situations (DnS) ontology + DOLCE Ultra Lite*
 - build consensus around the main branches of our core concept tree
- Central concepts:
 - entities: class (Q16889133) \approx rdfs:Class and entity (Q35120) \approx rdfs:Resource
 - properties: instance of (P31) \approx rdf:type and subclass of (P279) \approx rdf:type



https://www.wikidata.org/wiki/Wikidata:WikiProject_Ontology

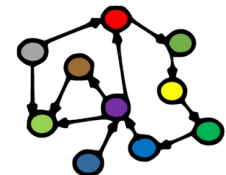
https://www.wikidata.org/wiki/Wikidata:WikiProject_Ontology/Modelling

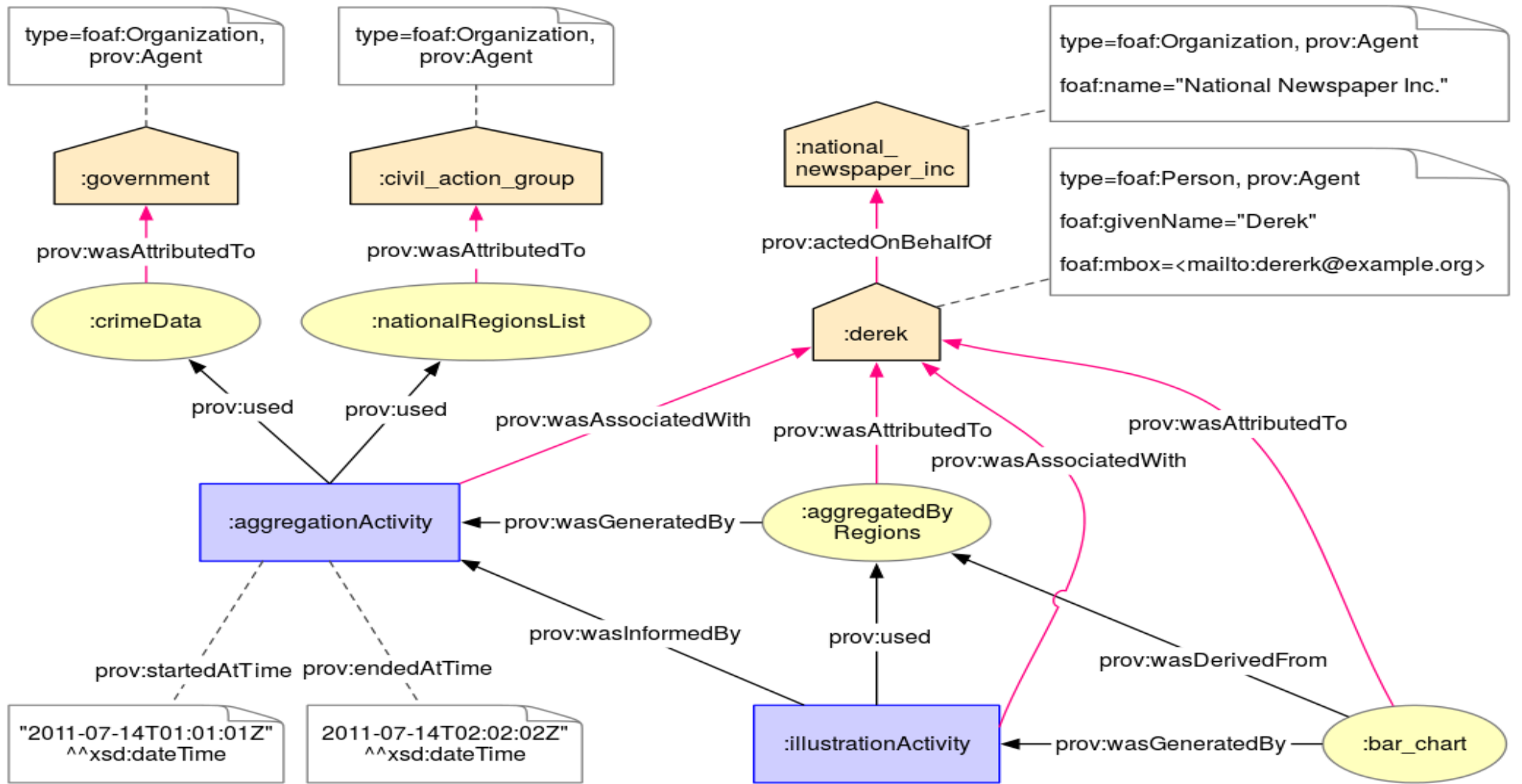
GeoNames ontology

- 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*...

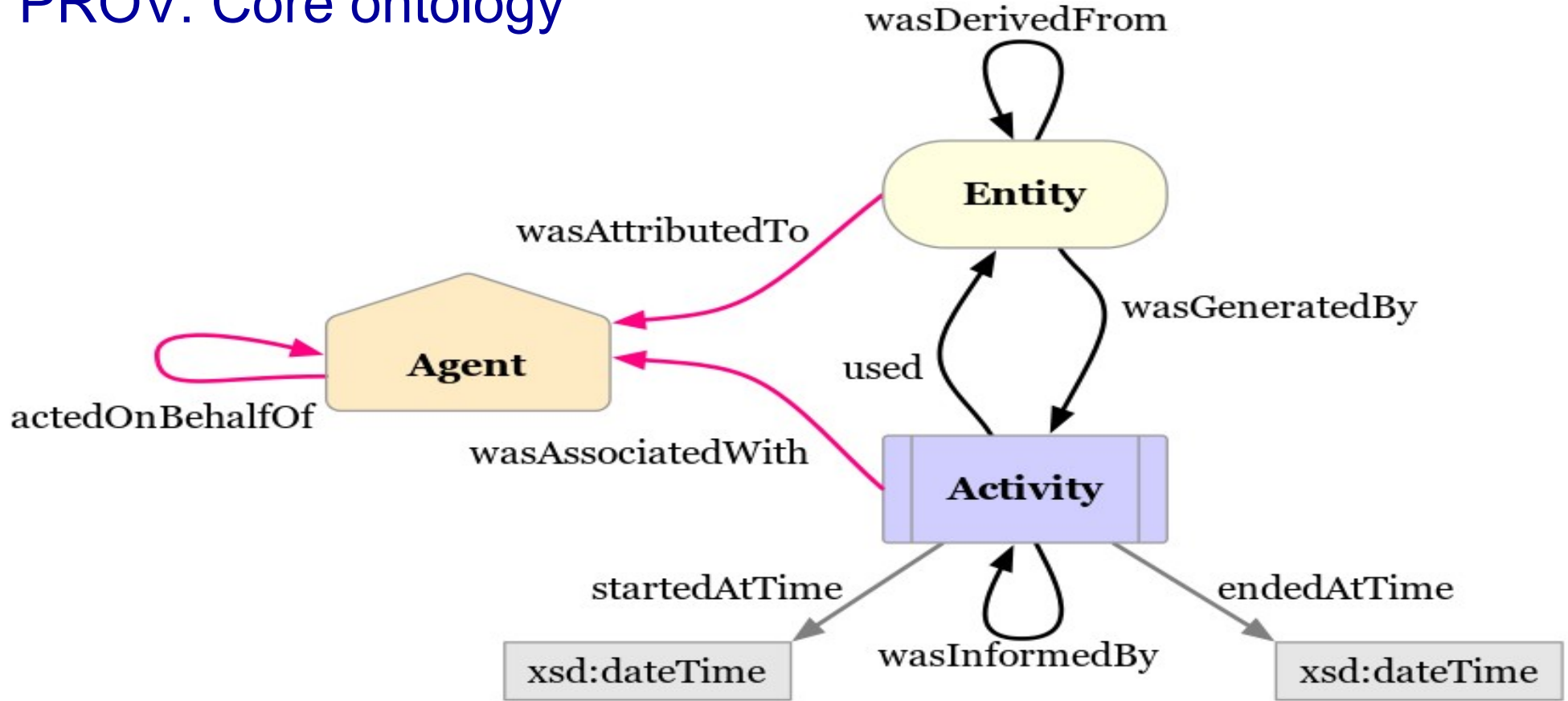
Provenance Interchange (PROV)

- 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*
- PROV: 50 classes, 83 properties, 1 instance
- PROV-O: 30 classes, 49 properties



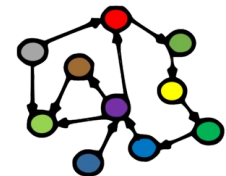


PROV: Core ontology



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



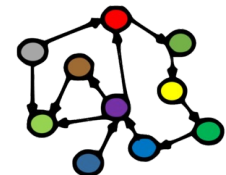
Creative commons (CC)



- 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/>



Next week:
Reasoning about KGs
(Description Logic - DL)