INFO216:
Knowledge Graphs

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

Session S07: Vocabularies

- Themes:
  - semantic vocabularies *(mostly S07-S08)*
    - SKOS, schema.org (← *covered in INFO116*)
    - Several others: DC, FOAF, VCard, geo, Data Cube (qb), VANN, VS, CC, VoID, PROV, Event, Time, Timeline (tl), BIO, SIOC, Bibo, Music (mo)...
  - linked open datasets *(bridge to S09-S10)*
    - DBpedia, Wikipedia, GeoNames
    - Perhaps others *(Facebook OGP, Graph API)*
    - *...some of them have their own vocabularies*

Terms

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

Readings

- Allemang & Hendler (2011):
  Semantic Web for the Working Ontologist.
  - chapters 9, 10 and 13
- Supplementary links in the portal:
  - Linked Open Vocabularies (LOV)
    http://lov.okfn.org/dataset/lov/
  - LOD stats
    http://lodstats.aksw.org

Semantic vocabularies

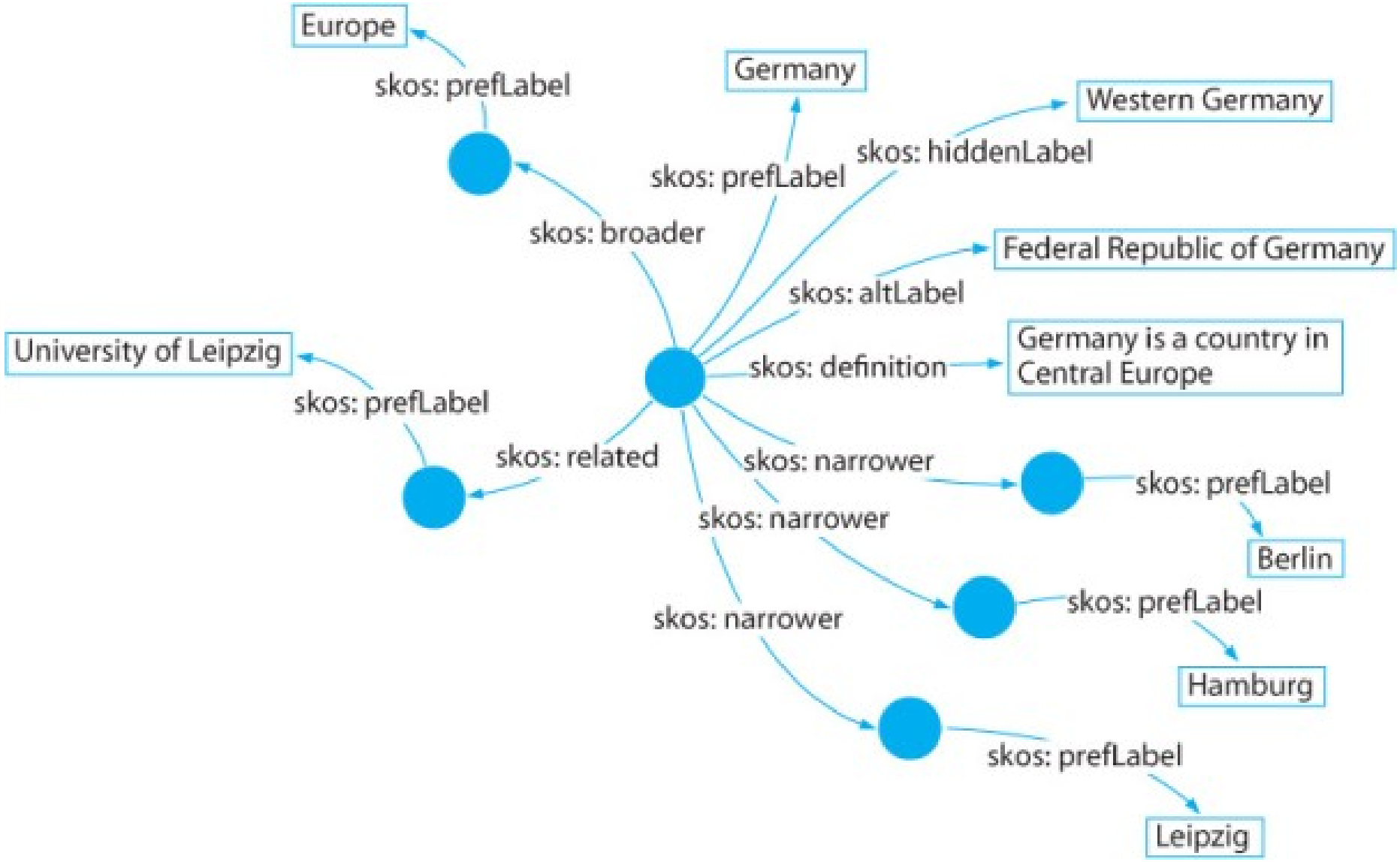
http://lov.okfn.org/dataset/lov/

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*

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*



Source: grips.semantic-web.at

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...

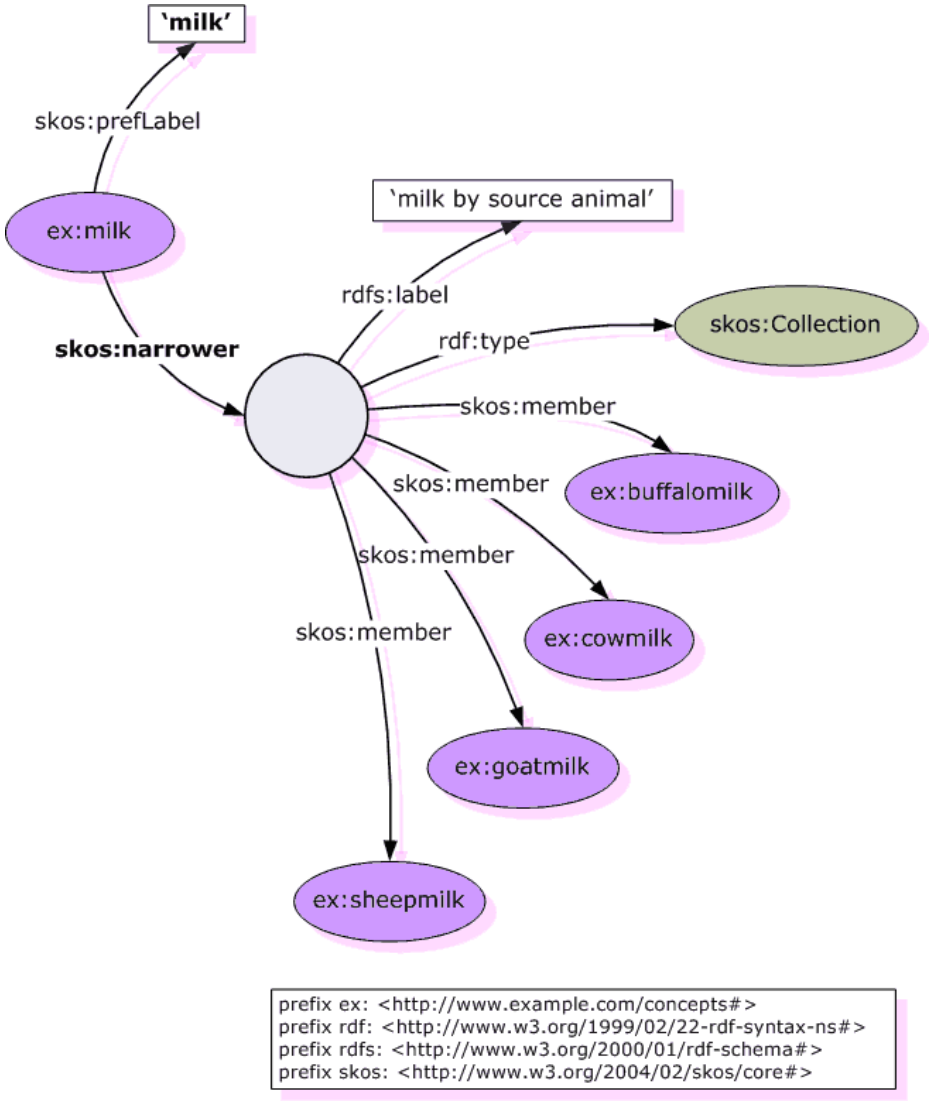
SKOS: concept schemes

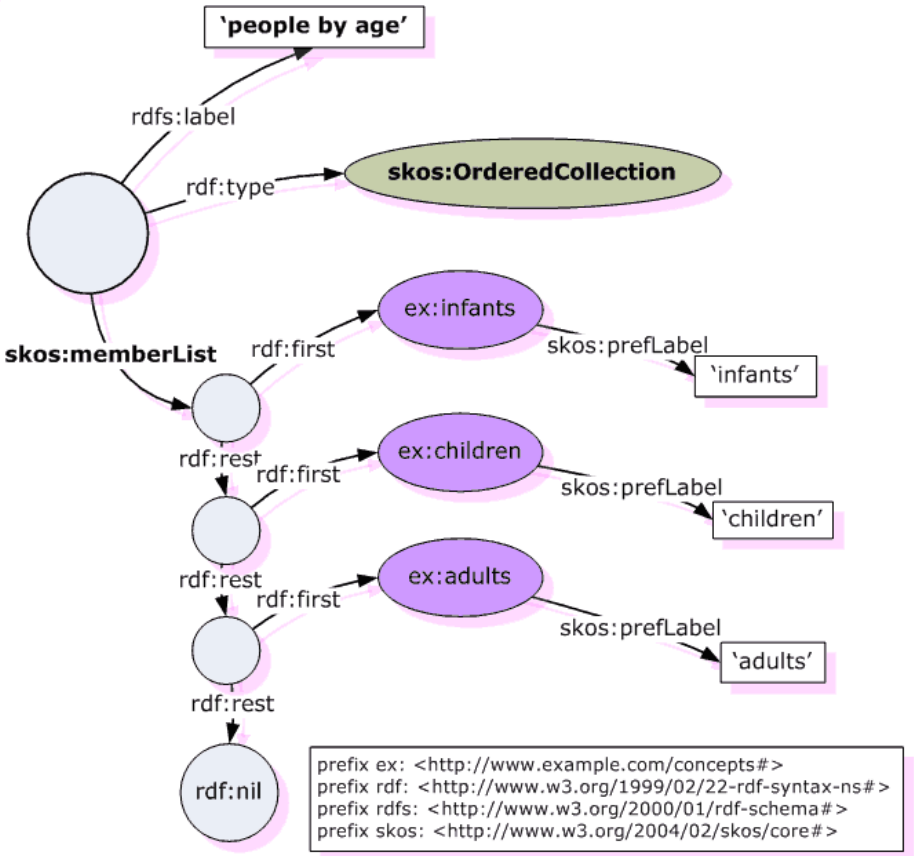
- Class: skos:ConceptScheme
  - skos:inScheme, skos:topConceptOf, skos:hasTopConcept
- Properties that relate concepts in different schemes:
  - skos:mappingRelation, and its “rdfs:subProperties”:
    - skos:closeMatch, skos:exactMatch, skos:relatedMatch, skos:broadMatch, skos:narrowMatch

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)*

schema.org: class/type hierarchy

- Commonly used types (subclasses of Thing):
  - creative works:
    - CreativeWork, Book, Movie, MusicRecording, Recipe, TVSeries...
  - embedded non-text objects:
    - AudioObject, ImageObject, VideoObject
  - Health and medical types, MedicalEntity
  - Person, Organization
  - Event, Place, LocalBusiness, Restaurant...
  - *Product, Offer,* AggregateOffer
  - Review, AggregateRating

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
- *...based on the GoodRelations vocabulary*

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...
- *...also based on GoodRelations*

schema.org: Data types

- Primitive data types (subclasses of DataType):
  - Boolean
  - Date (ISO 8601)
  - DateTime (also ISO 8601)
  - Number (Float, Integer)
  - Text (IRI)
  - 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!*

METADATA
Dublin Core® Metadata Initiative
INNOVATION

Important DCTerms

- dcterms:creator – IRI 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 – IRI for the resource that is responsible for making the dataset available
- dcterms:contributor – IRI the resource that is responsible for making contributions to the dataset
- dcterms:source – IRI 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

VCard

- Representing electronic visiting card in RDF
  - older than the web of data, started with emails etc.
- Classes:
  - Kind: Individual, Organization, Group, Location
- Properties:
  - hasFN, hasGivenName, hasFamilyName, hasAdditionalName, hasHonorificPrefix, hasHonorifixSuffix, hasNickname
  - hasStreetAddress, street-address, locality, region, country-name, postal-code
  - hasGeo, tz, language, hasTelephone, hasEmail
  - hasTitle, hasRole, hasOrganizationName, -Unit...

Geo (WGS84)

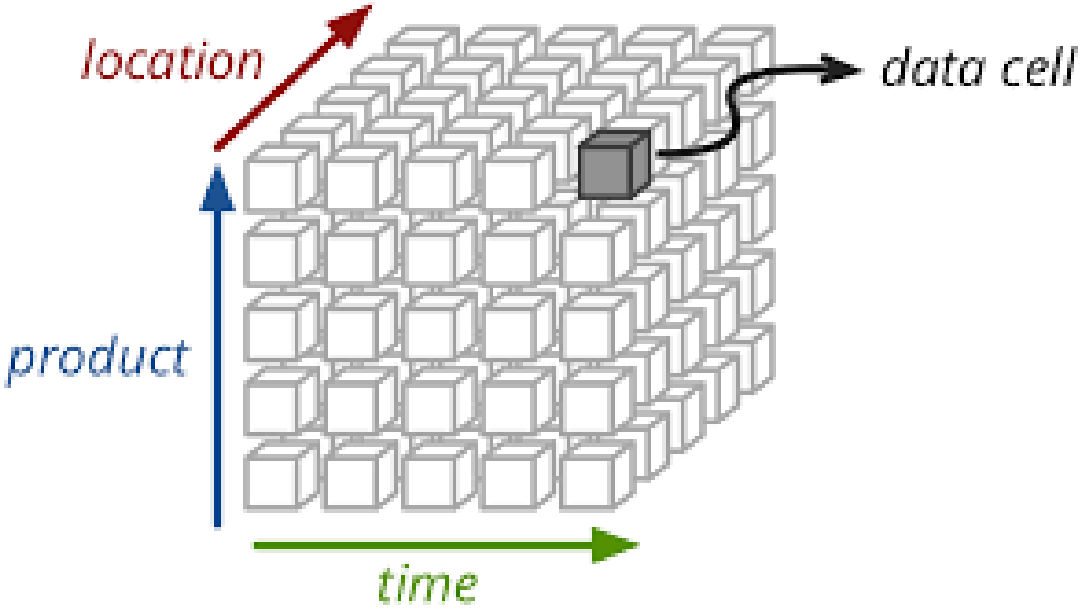
- A vocabulary for representing latitude, longitude and altitude according to WGS84 (World Geodetic Standard)
- @prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
- Classes: SpatialThing, Point
- Properties:
  - lat, long: latitude and longitude in decimal degrees
  - lat_long: comma-separated 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!*

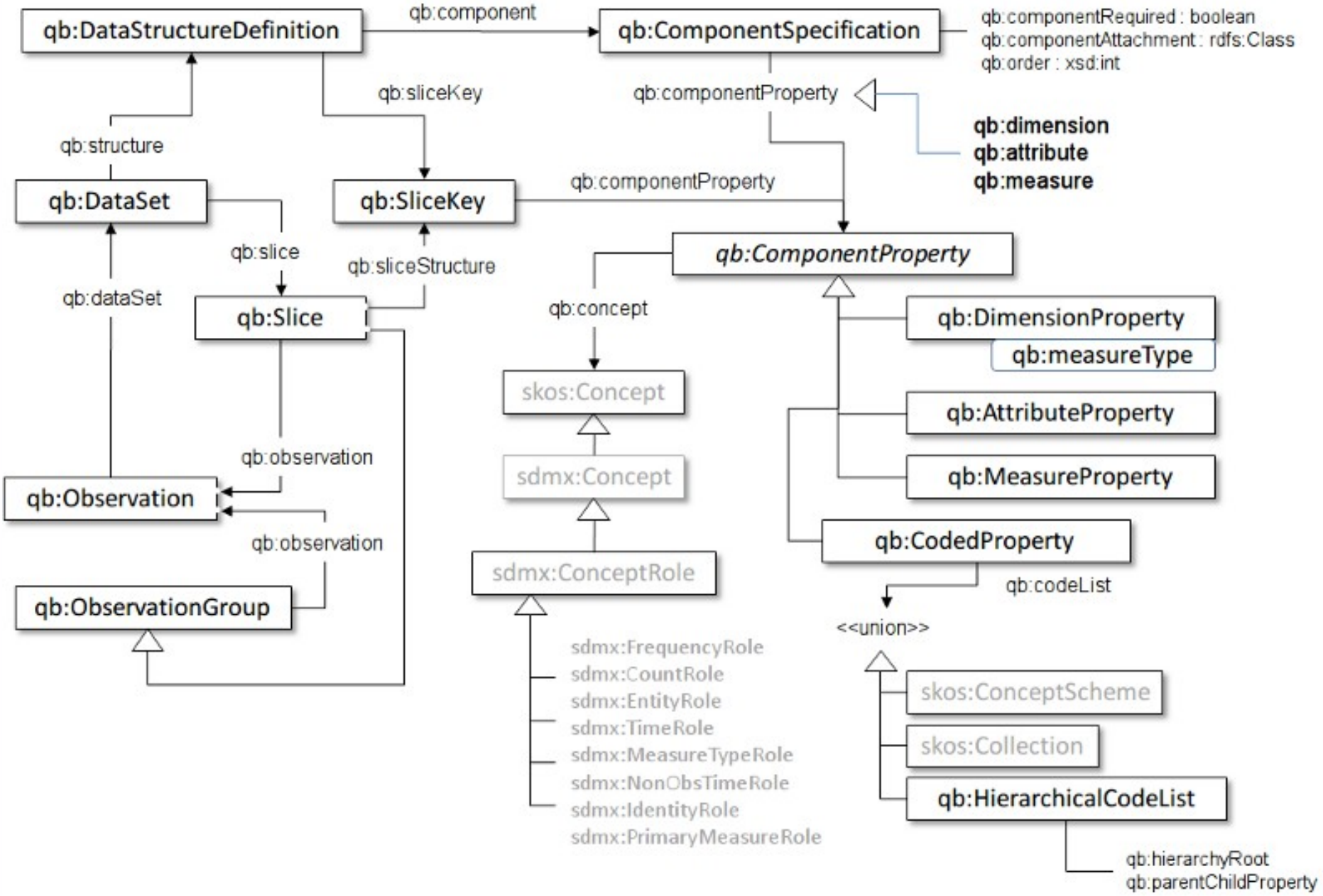
Data cube vocabulary

- To be able to publish multi-dimensional data, such as statistics, on the web … a core foundation which supports extension vocabularies for other aspects of statistical data flows or multi-dimensional data sets
  - @prefix qb: <http://purl.org/linked-data/cube#> .
- A cube is organized according to components:
  - dimension components identify observations uniquely
  - measure components represent the observed phenomenon
  - attribute components qualify and interpret the observed value(s): e.g., units of measure, scaling factors, and metadata such as the observation status

Data cube vocabulary

- Examples:
  - products (by category, type, variant etc.)
  - dimensions are: place and time
  - observations are: sales at places and times
  - attributes are: price, profit, turnover, customer etc.




qb:DataStructureDefinition
qb:component
qb:ComponentSpecification
qb:componentRequired : boolean
qb:componentAttachment : rdfs:Class
qb:order : xsd:int
qb:sliceKey
qb:componentProperty
qb:dimension
qb:attribute
qb:measure
qb:structure
qb:DataSet
qb:SliceKey
qb:componentProperty
qb:slice
qb:sliceStructure
qb:ComponentProperty
qb:dataSet
qb:Slice
qb:concept
qb:DimensionProperty
qb:measureType
skos:Concept
qb:AttributeProperty
qb:observation
sdmx:Concept
qb:MeasureProperty
qb:Observation
qb:observation
qb:CodedProperty
sdmx:ConceptRole
qb:codeList
qb:ObservationGroup
<<union>>
sdmx:FrequencyRole
sdmx:CountRole
sdmx:EntityRole
sdmx:TimeRole
sdmx:MeasureTypeRole
sdmx:NonObsTimeRole
sdmx:IdentityRole
sdmx:PrimaryMeasureRole
skos:ConceptScheme
skos:Collection
qb:HierarchicalCodeList
qb:hierarchyRoot
qb:parentChildProperty

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)

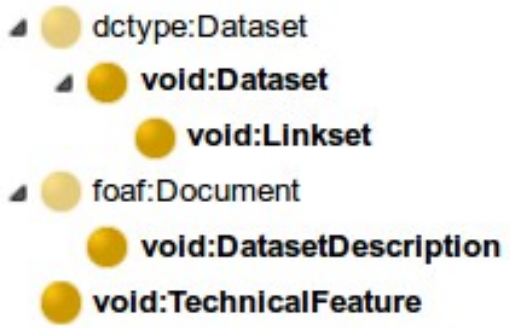
- An RDF vocabulary for describing the status of vocabulary terms on the Web of Data
  @prefix vs: <http://www.w3.org/2003/06/sw-vocab-status/ns#> .
- Main concepts:
  - no classes
  - three properties:
    - term_status (e.g., *unstable*, *testing*, *stable* or *archaic*)
    - moreinfo
    - userdocs

Creative commons (CC)

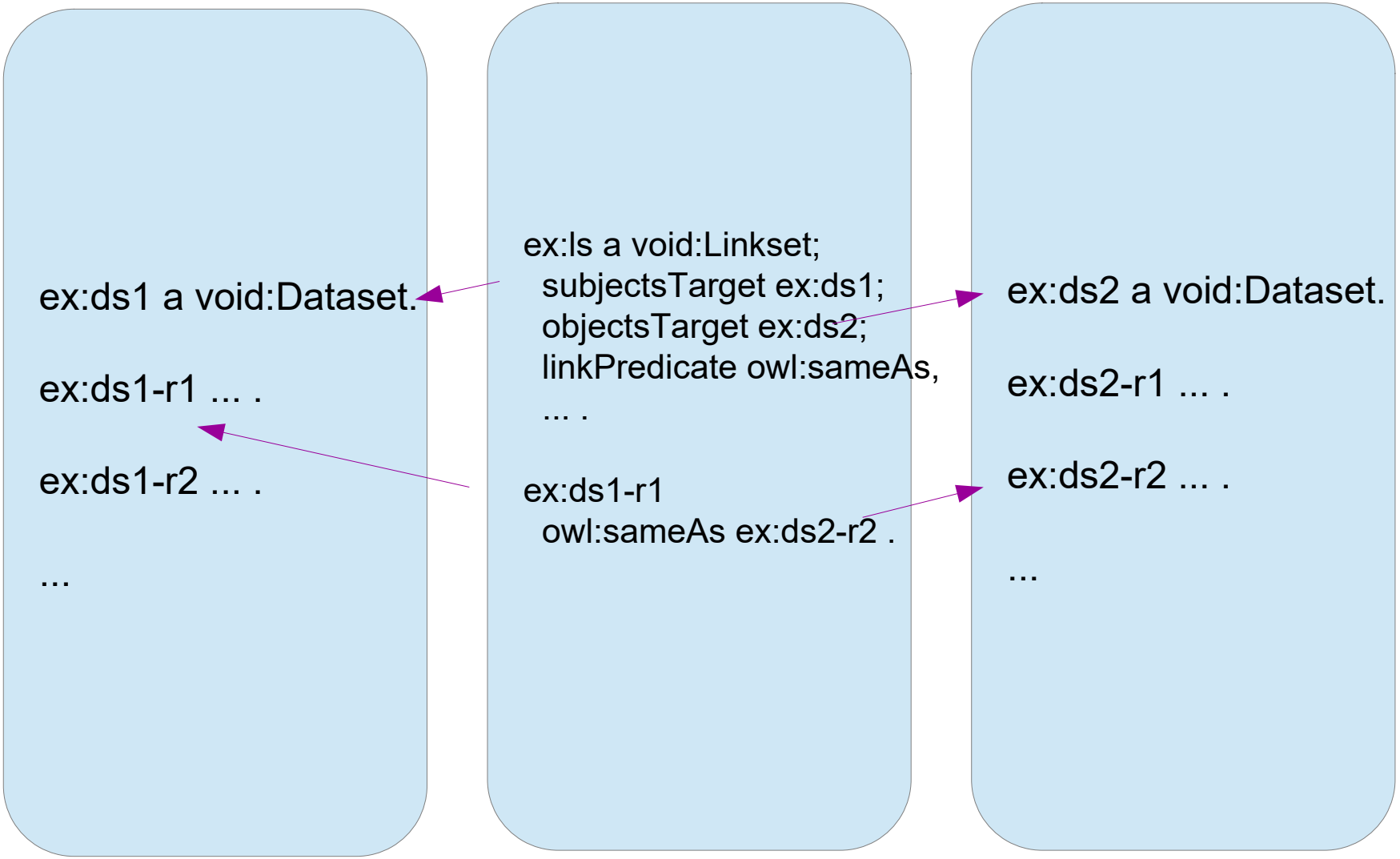
- The Creative Commons Rights Expression Language (CC REL) lets you describe copyright licenses in RDF
- @prefix cc: <http://creativecommons.org/ns> .
- Classes:
  - Work, License, Jurisdiction
  - Permission (Reproduction, Distribution, DerivativeWorks, Sharing)
  - Requirement (Notice, Attribution,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

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



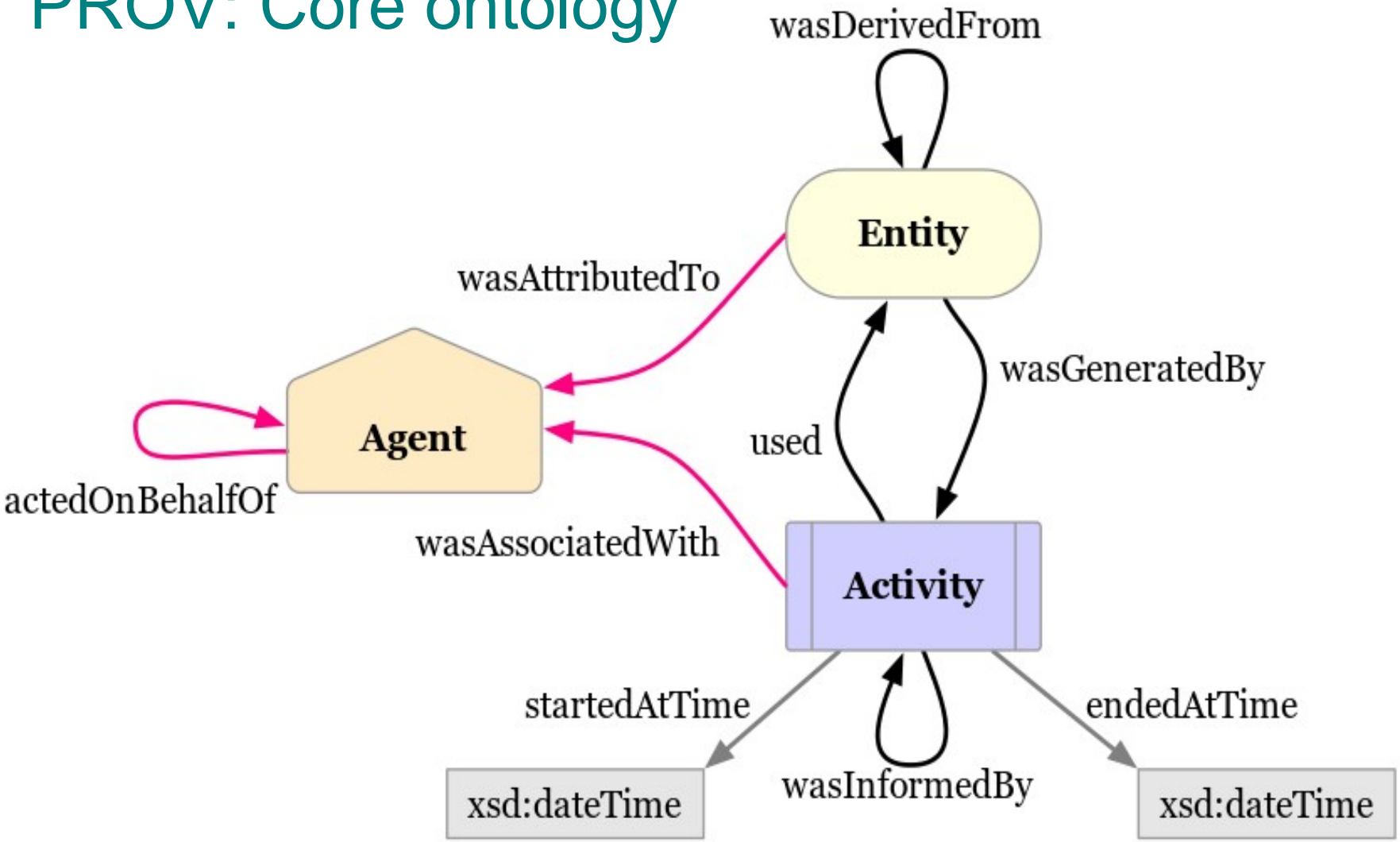
Vocabulary of Interlinked Datasets (VoID)



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

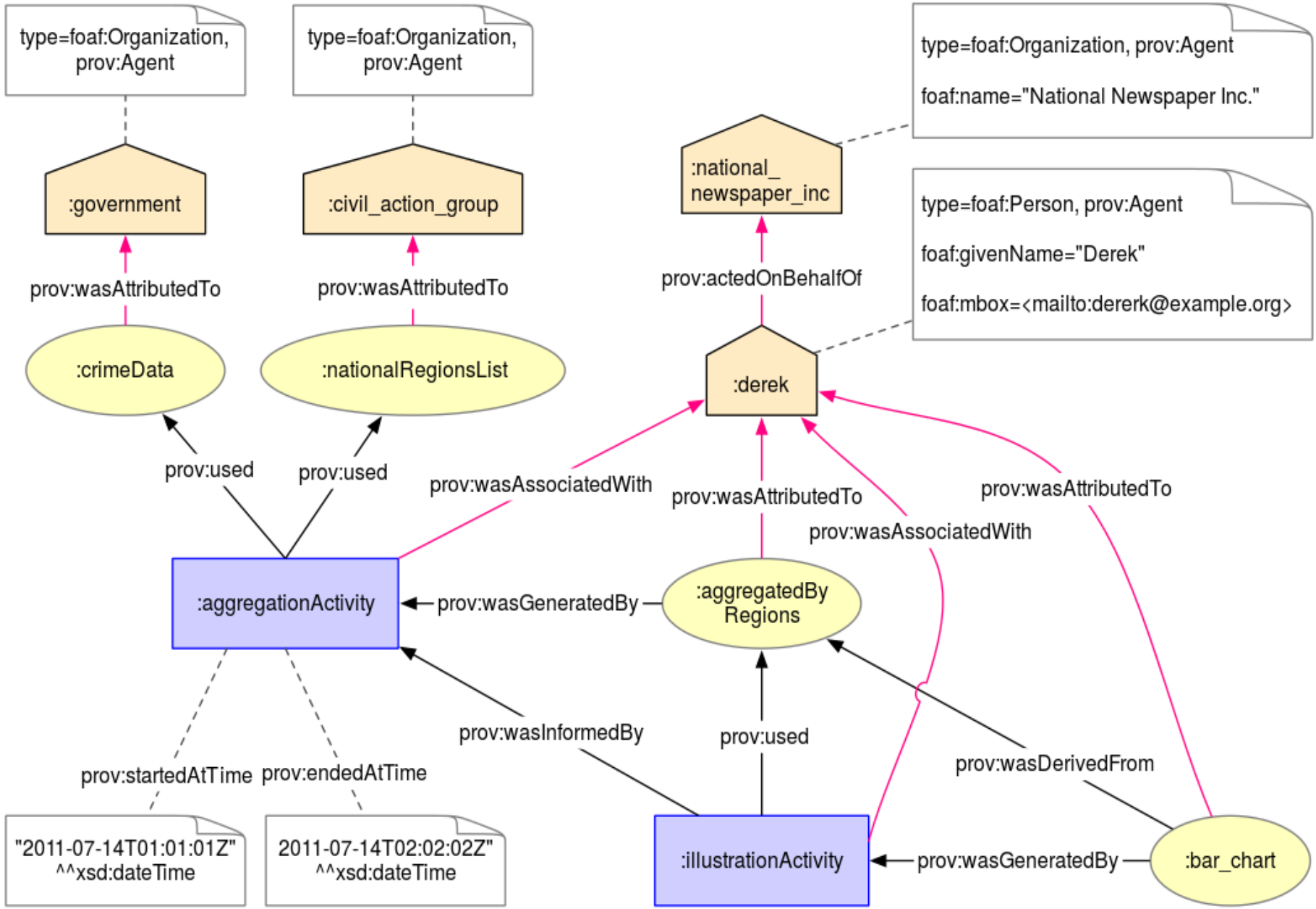
PROV: Core ontology



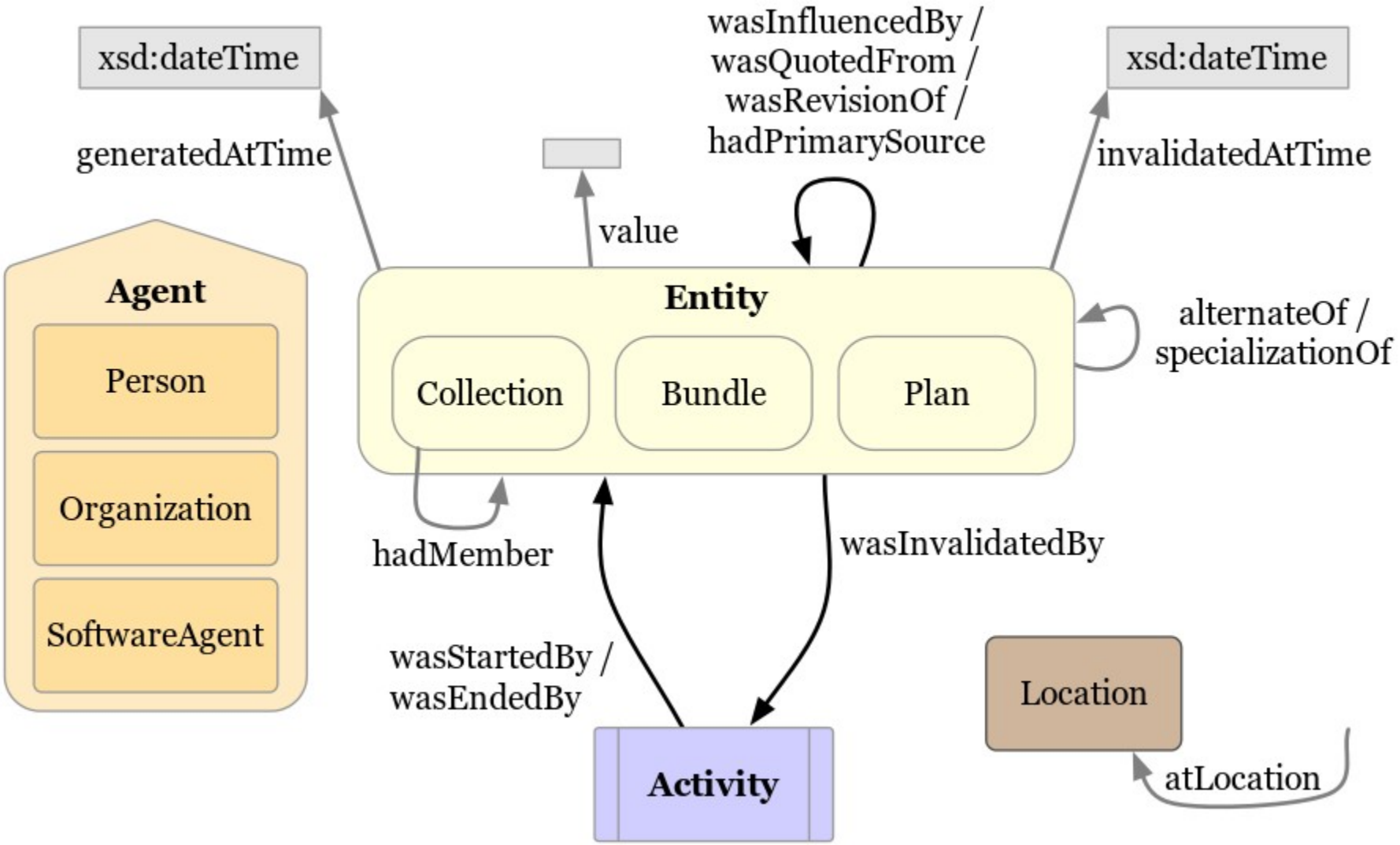
Source: https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/

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



PROV: Qualification Patterns

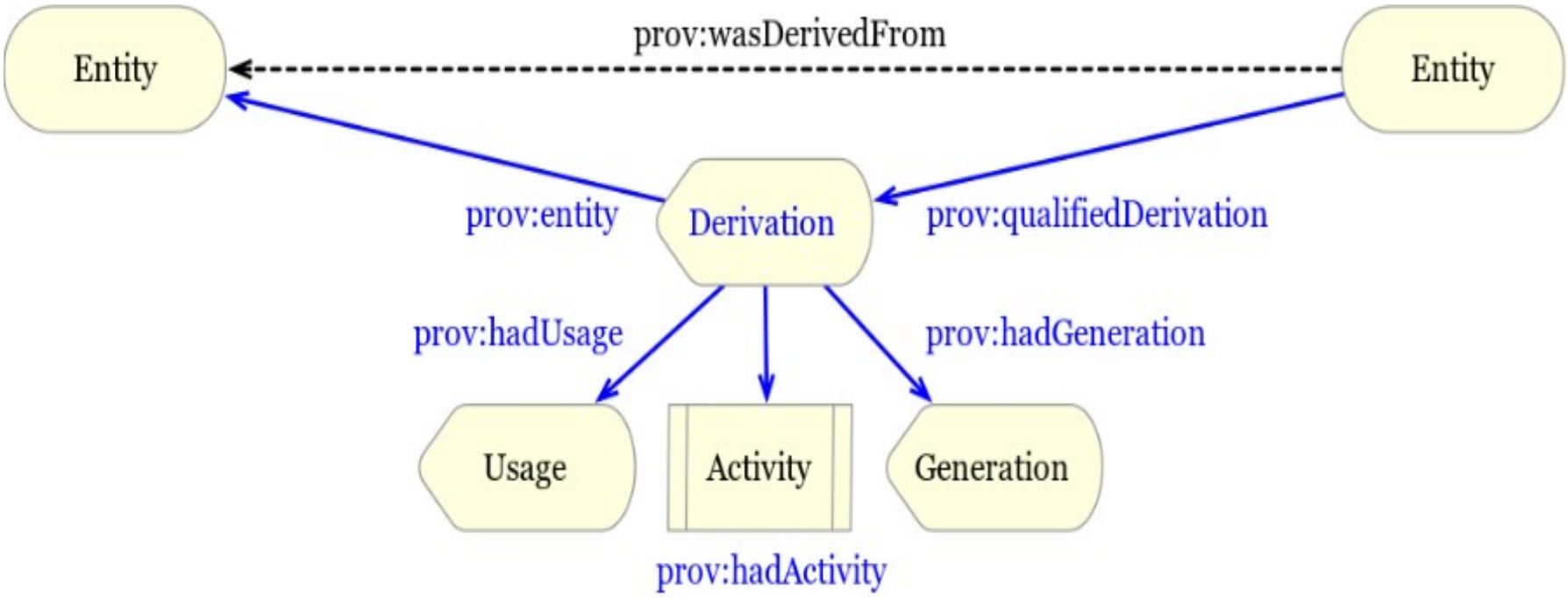
- “My Sweet Lord” was written by George Harrison:

```
ex:MySweetLord      a prov:Entity;
                    prov:wasAttributedTo ex:GeorgeHarrison .
ex:GeorgeHarrison a prov:Agent.
```

- “My Sweet Lord” may have plagiarised The Chiffons “He's So Fine”:

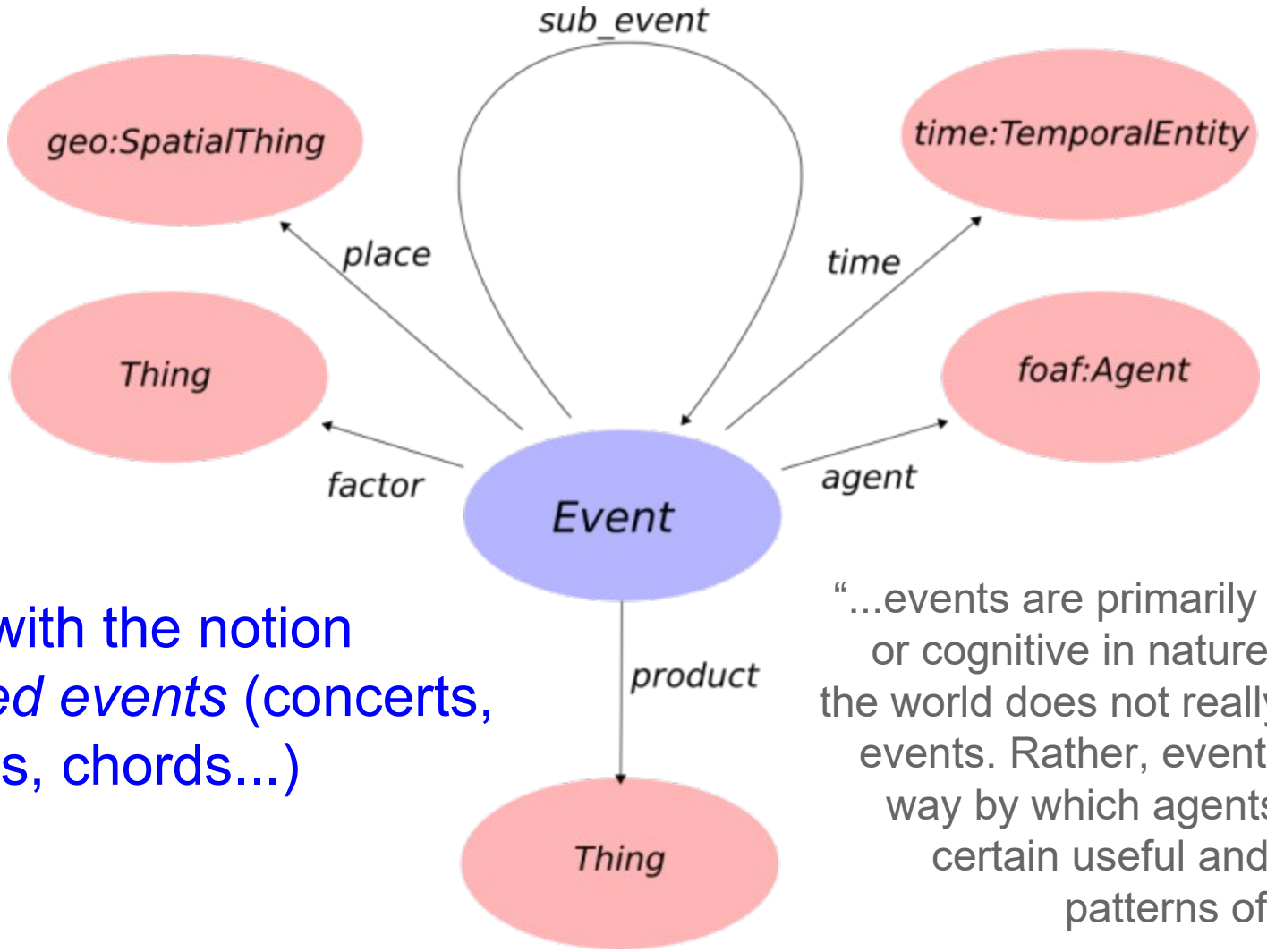
```
ex:MySweetLord      a prov:Entity;
                    prov:wasDerivedFrom ex:HesSoFine .
ex:HesSoFine        a prov:Entity;
                    prov:wasAttributedTo ex:RonaldMack .
ex:RonaldMack       a prov:Agent.
```

PROV: Qualification Patterns



- A useful technique in general
  - whenever we want to add more details about a relationship
  - similar to AssociationClasses and LinkObjects in UML
  - similar to Relationships with Attributes in ER models

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#> .

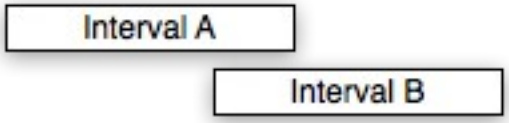
Time ontology in OWL (time, OWL-time)

- 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

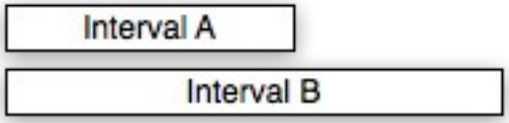
A is before B or
B is after A

A meets B or
B is met by A

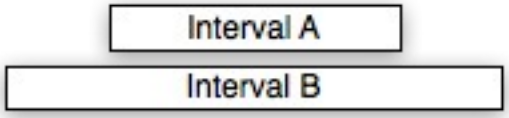
A overlaps with B or
B is overlapped by A



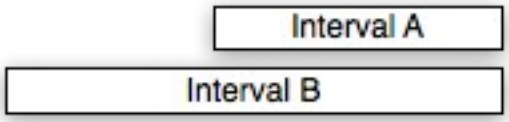
A starts B or
B is started-by A



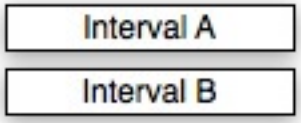
A during B or
B contains A



A finishes B or
B is finished-by A



A and B are cotemporal



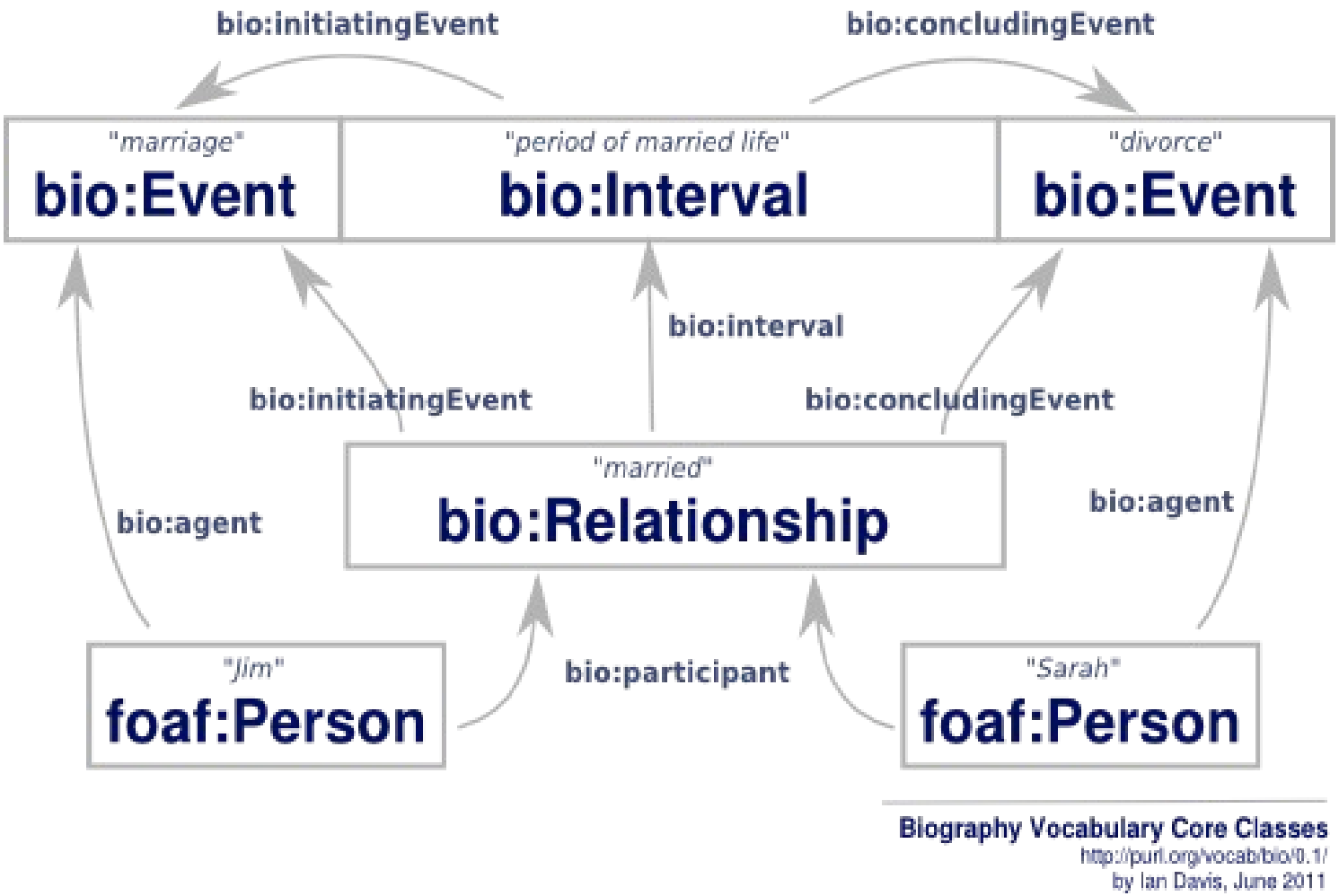
Timeline Ontology (tl)

- 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

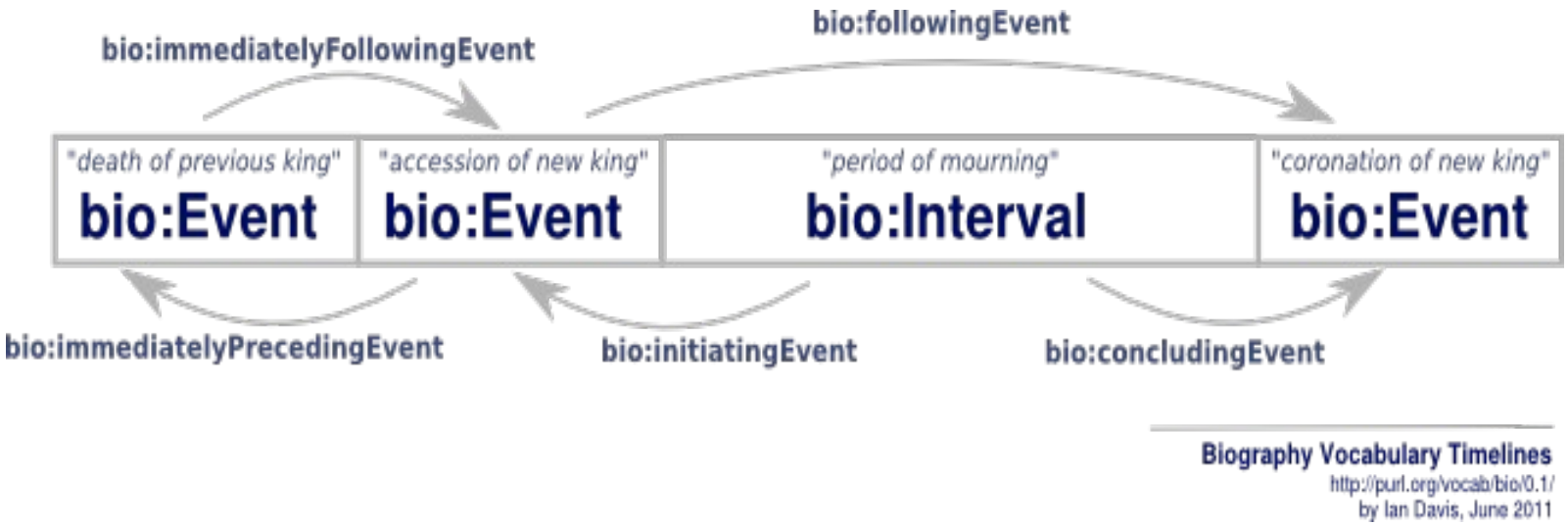
Biographical Information (BIO)

- 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)



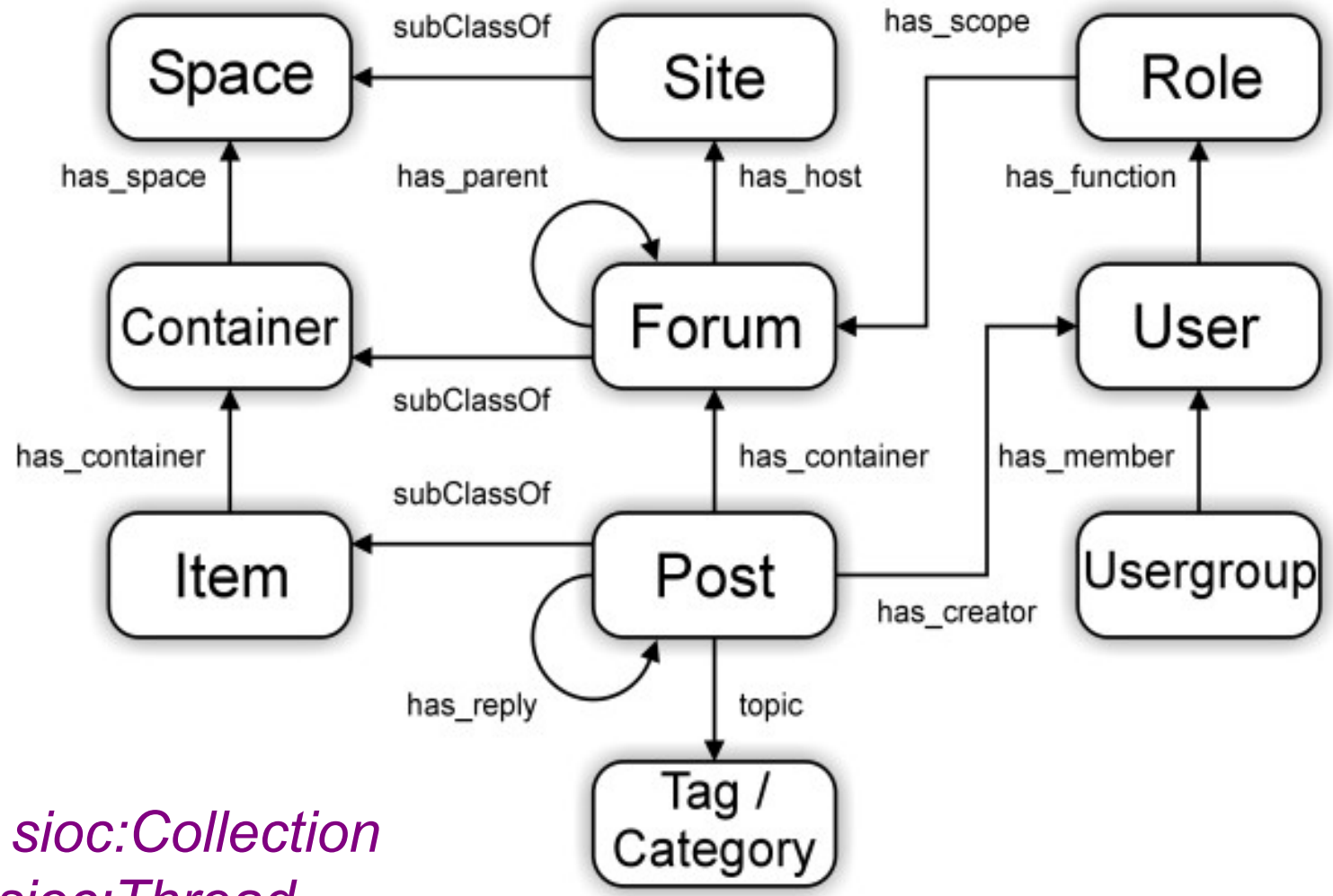
Biographical Information (BIO)



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.

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*

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)

AKT

- Basic concepts to manage scientific projects and knowledge
  - old (2001-2003), pre-RDFS, pre-OWL
  - in RDFS and OWL today
  - offline since 2014, but still much used!
- @prefix akt: <http://www.aktors.org/ontology/portal#>
  - domain no longer dereferencable
- Main concepts:
  - docs (publications), events,
    load (ontology definitions),
    organizations (including people), projects,
    research-areas, techs (technologies)