Lecture 19 – Interoperability

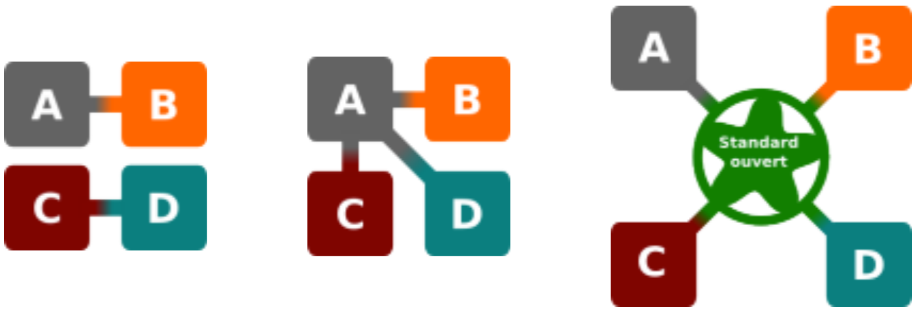
Open Data Management & the Cloud

(Data Science & Scientific Computing / UniTS – DMG)

Interoperability

- Interoperability is a characteristic of a product or system, whose interfaces are completely understood, to work with other products or systems, at present or in the future, in either implementation or access, without any restrictions.
- It requires open standards by definition
  - A dominant (non open) standard means compatibility or de-facto standard
- Open standards can be reached by post-facto solutions

Compatibility De facto standard Interoperability



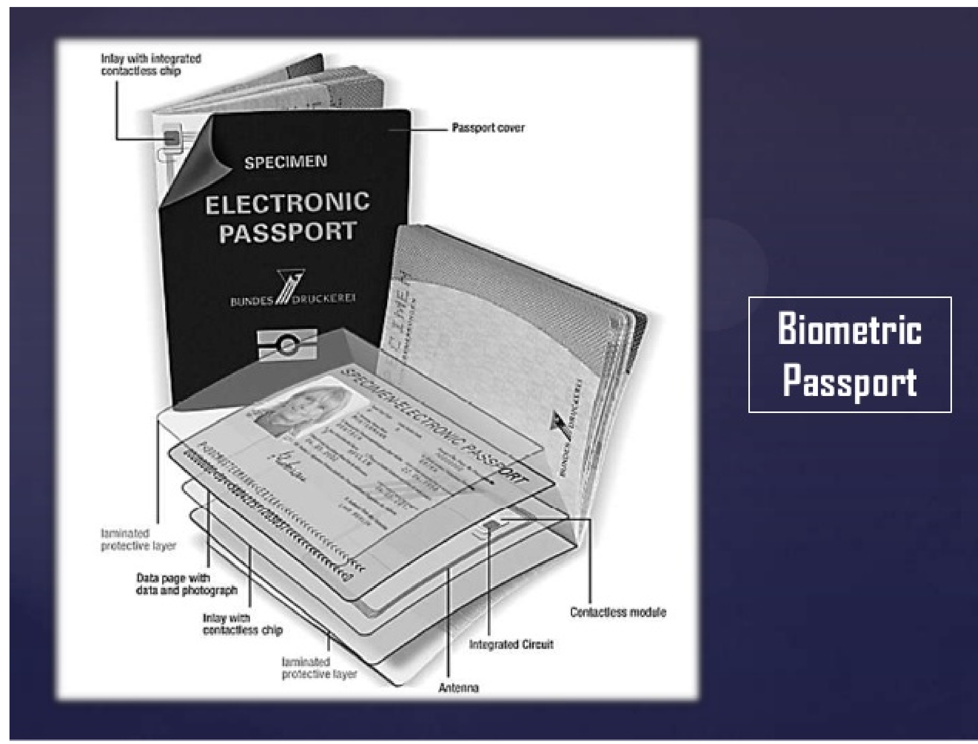
Compatibility

- Power plugs
- Serial, ps/2, usb cables
- Adapters

De-facto standard

- Biometric passport information
  - Contactless smart card
  - Stores information in a shared/agreed format
  - Allows data validation
  - Not an open standard
- Portable Document Format
  - Freely usable
  - Non-modifiable standard

Adobe® PDF



(Open) Interoperability

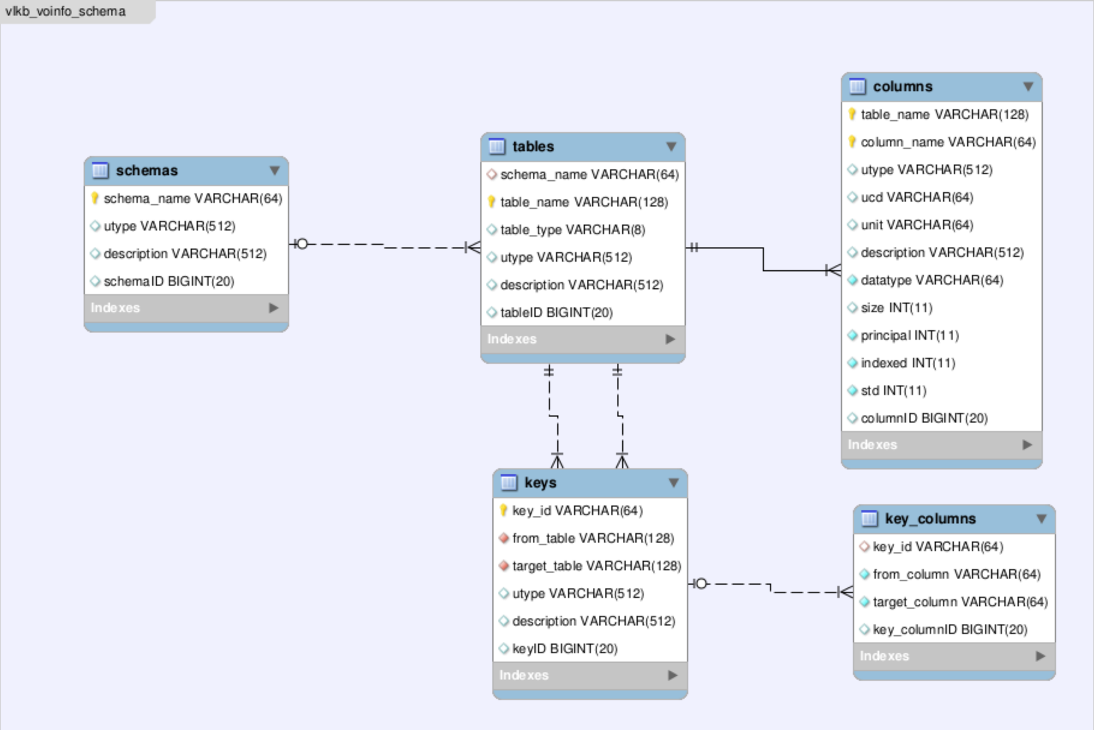
- Service Oriented Architecture (SOA)
  - Network, Service description
    - WSDL, SOAP, ReST
- Virtual Research Environment (VRE)
  - Domain driven
  - Standards organization attached
    - Government can be open itself

Syntactic Interoperability

- Where two or more systems are able to communicate and exchange data. It allows different software components to cooperate, even if the interface and the programming language are different.
- Syntactic interoperability refers to the packaging and transmission mechanisms for data.
- It involves a common data format and common protocol to structure any data so that the manner of processing the information will be interpretable from the structure.
- It also allows detection of syntactic errors, thus allowing receiving systems to request resending of any message that appears to be garbled or incomplete.
- Syntactic interoperability is a prerequisite for semantic interoperability.
- No semantic communication is possible if the syntax is garbled or unable to represent the data.

Syntactic Interoperability (example)

```
<?xml version='1.0'?>
<VOTABLE version="1.3" xmlns="http://www.ivoa.net/xml/VOTable/v1.3">
<!--
 !  VOTable written by STIL version 3.3-1
 !  at 2018-12-03T13:44:12
 !-->
 <RESOURCE>
  <TABLE name="QFT-nofilter-test10.csv" nrows="10">
   <FIELD datatype="int" name="idcss"/>
   <FIELD arraysize="38" datatype="char" name="bands"/>
   <FIELD datatype="short" name="id1100"/>
   ...
   <DATA>
    <TABLEDATA>
     <TR>
      <TD>17812111</TD>
      <TD>1100-500-350-250-160-matches</TD>
      <TD>8190</TD>
      ...
     ...
     </TR>
    </TABLEDATA>
   </DATA>
  </TABLE>
 </RESOURCE>
</VOTABLE>
```



Semantic Interoperability

- The ability of computer systems to exchange data with unambiguous, shared meaning.
- Is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems.
- Semantic interoperability is therefore concerned not just with the packaging of data (syntax), but the simultaneous transmission of the meaning with the data (semantics).
  - This is accomplished by adding data about the data (metadata)
  - Linking each data element to a controlled, shared vocabulary.
  - The meaning of the data is transmitted with the data itself, in one self-describing "information package" that is independent of any information system.
  - It is this shared vocabulary, and its associated links to an ontology, which provides the foundation and capability of machine interpretation, inference, and logic.
- The current internet standard for document markup is XML, which uses "< >" as a data delimiter. The data delimiters convey no meaning to the data other than to structure the data.
  - Without a data dictionary to translate the contents of the delimiters, the data remains meaningless.
- The data exchanged between two or more systems is understandable to each system.

Semantic Interoperability (example)


```
<?xml version='1.0'?>
<VOTABLE version="1.3"
 xmlns="http://www.ivoa.net/xml/VOTable/v1.3">
<!--
 !  VOTable written by STIL version 3.3-1 (uk.ac.starlink.votable.VOTableWriter)
 !  at 2018-12-03T13:59:14
 !-->
<RESOURCE>
<TABLE name="ObsCore" nrows="77">
<DESCRIPTION>
The IVOA-defined obscore table, containing generic metadata for
datasets within this datacenter.
</DESCRIPTION>
<PARAM arraysize="9" datatype="char" name="CoordFlavor" utype="stc:AstroCoordSystem.SpaceFrame.CoordFlavor" value="SPHERICAL"/>
<PARAM arraysize="4" datatype="char" name="CoordRefFrame" utype="stc:AstroCoordSystem.SpaceFrame.CoordRefFrame" value="ICRS"/>
<PARAM arraysize="41" datatype="char" name="URI" utype="stc:DataModel.URI" value="http://www.ivoa.net/xml/STC/stc-v1.30.xsd"/>
<PARAM arraysize="31" datatype="char" name="server" value="http://dc.zah.uni-heidelberg.de"/>
<PARAM arraysize="983" datatype="char" name="query" value="SELECT ivoa.Obscore.dataproduct_type, ivoa.Obscore.dataproduct_subtype, ivoa.Obscore.calib_level,
ivoa.Obscore.obs_collection, ivoa.Obscore.obs_id, ivoa.Obscore.obs_title, ivoa.Obscore.obs_publisher_did, ivoa.Obscore.obs_creator_did,
ivoa.Obscore.access_url, ivoa.Obscore.access_format, ivoa.Obscore.access_estsize, ivoa.Obscore.target_name, ivoa.Obscore.target_class, ivoa.Obscore.s_ra,
ivoa.Obscore.s_dec, ivoa.Obscore.s_fov, ivoa.Obscore.s_region, ivoa.Obscore.s_resolution, ivoa.Obscore.t_min, ivoa.Obscore.t_max, ivoa.Obscore.t_exptime,
ivoa.Obscore.t_resolution, ivoa.Obscore.em_min, ivoa.Obscore.em_max, ivoa.Obscore.em_res_power, ivoa.Obscore.o_ucd, ivoa.Obscore.pol_states,
ivoa.Obscore.facility_name, ivoa.Obscore.instrument_name, ivoa.Obscore.s_xel1, ivoa.Obscore.s_xel2, ivoa.Obscore.t_xel, ivoa.Obscore.em_xel,
ivoa.Obscore.pol_xel, ivoa.Obscore.s_pixel_scale, ivoa.Obscore.em_ucd FROM ivoa.Obscore WHERE ((spoint(RADIANS(16.0), RADIANS(40.0))) @ (s_region)) LIMIT
20000"/>
<PARAM arraysize="48" datatype="char" name="src_res" value="Contains traces from resource __system__/obscore">
<DESCRIPTION>Definition and support code for the ObsCore data model and table.</DESCRIPTION>
</PARAM>
<PARAM arraysize="39" datatype="char" name="src_table" value="Contains traces from table ivoa.ObsCore">
<DESCRIPTION>The IVOA-defined obscore table, containing generic metadata for
datasets within this datacenter.</DESCRIPTION>
</PARAM>
<PARAM arraysize="2" datatype="char" name="QUERY_STATUS" value="OK">
<DESCRIPTION>Query successful</DESCRIPTION>
</PARAM>
<PARAM arraysize="54" datatype="char" name="citation" value="http://dc.zah.uni-heidelberg.de/tableinfo/ivoa.ObsCore">
<DESCRIPTION>For advice on how to cite the resource(s) that contributed to this result, see http://dc.zah.uni-heidelberg.de/tableinfo/ivoa.ObsCore</
DESCRIPTION>
</PARAM>
<FIELD ID="dataproduct_type" arraysize="*" datatype="char" name="dataproduct_type" ucd="meta.id" utype="obscore:obsdataset.dataproducttype">
<DESCRIPTION>High level scientific classification of the data product, taken from an enumeration</DESCRIPTION>
</FIELD>
<FIELD ID="dataproduct_subtype" arraysize="*" datatype="char" name="dataproduct_subtype" ucd="meta.id" utype="obscore:obsdataset.dataproductsubtype">
<DESCRIPTION>Data product specific type</DESCRIPTION>
</FIELD>
<FIELD ID="calib_level" datatype="short" name="calib_level" ucd="meta.code;obs.calib" utype="obscore:obsdataset.caliblevel">
```

Cross-domain Interoperability

- Cross-domain interoperability refers to the ability of systems and organizations to interact and exchange information (inter-operate) among different areas, markets, industries, countries or communities of interest (domains).
- It means seamless communication and activity, despite reliance on different technical environments or frameworks.
- An example of cross-domain interoperability is the exchange of critical information in a disaster situation:
  - all responding organizations can effectively communicate and coordinate their actions to meet their mission objectives.
- Syntactic/Semantic distinction applies as well.

Cross-domain Interoperability

- Bridging domain gaps needs coordination.
- Organizations exist at various levels to build these bridging solutions.
- Domain boundaries exist at different levels
  - Distinction of domain WRT sub-domain
    - Astrophysics
      - Solar Physics
      - Galactic/Extragalactic domains
      - Radio/Optical/High-Energy/Particle Astrophysics
    - Physics
    - Economy
    - Health
    - Agriculture

Cross-domain Interoperability

- RDA – BagIt solution
  - Synctactic interoperability driven by a flexible general exchange format

```
myfirstbag/
|-- data
|   \-- 27613-h
|       \-- images
|           \-- q172.png
|           \-- q172.txt
|-- manifest-md5.txt
|     49afbd86a1ca9f34b677a3f09655eae9 data/27613-h/images/q172.png
|     408ad21d50cef31da4df6d9ed81b01a7 data/27613-h/images/q172.txt
\-- bagit.txt
      BagIt-Version: 0.97
      Tag-File-Character-Encoding: UTF-8
```

Conceptual Interoperability

- Interoperability is a broad term in information technology
- "Levels of Conceptual Interoperability Model" (LCIM)

0 No Interoperability
  - standalone systems

1 Technical Interoperability
  - protocol exists for exchanging data (e.g. at hardware/system level)

2 **Syntactic** Interoperability
  - common data format is applied

3 **Semantic** Interoperability
  - meaning of the data is shared

4 Pragmatic Interoperability
  - common methods and procedures

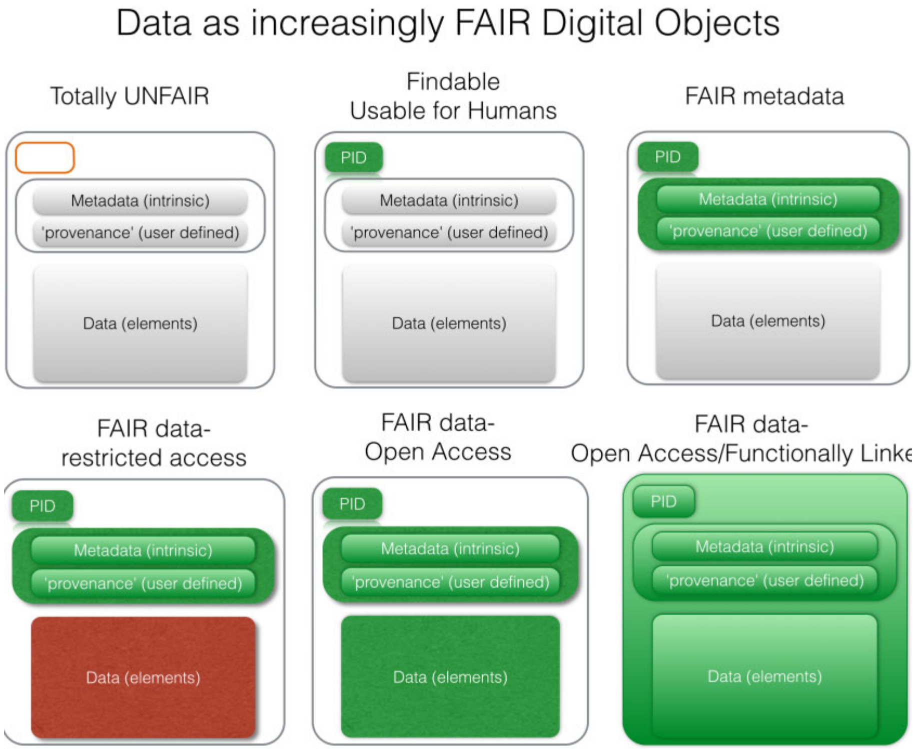
5 Dynamic Interoperability
  - understand system state changes

6 Conceptual Interoperability
  - fully specified, but implementation independent model

FA – I – R Interoperability

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

Metadata
Vocabularies
Annotations
Identifiers
Protocols

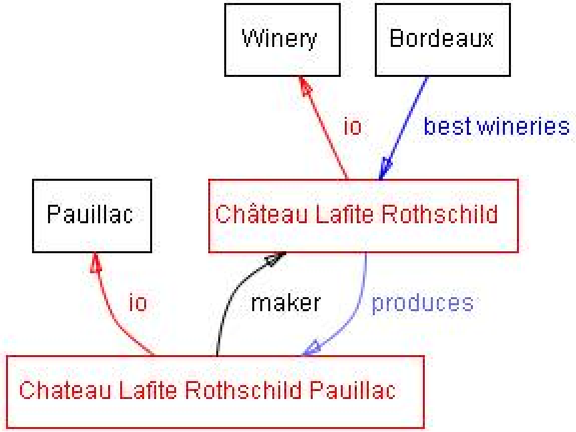


Data Models & Ontologies

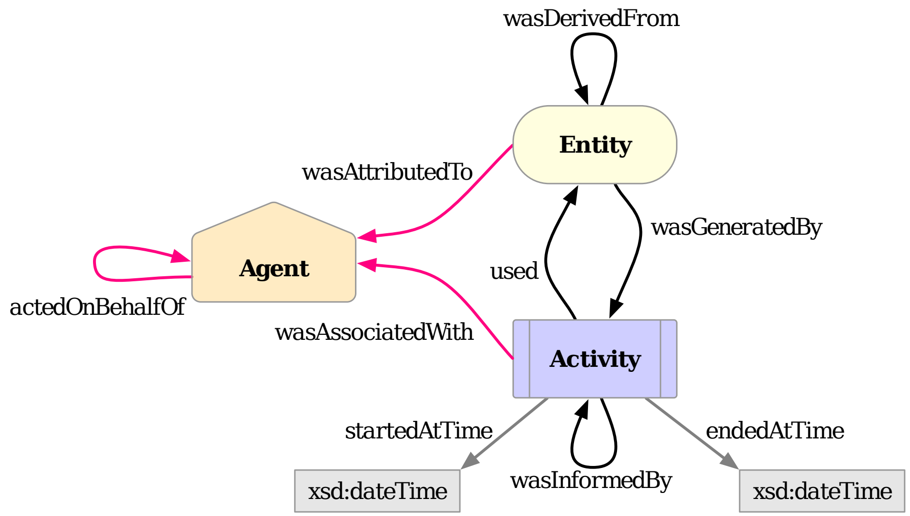
- The purpose of an ontology is to model the business.
  - It is independent from the computer systems.
  - Its purpose is to use formal logic and common terms to describe the business, in a way that both humans and machines can understand.
  - Ontologies use (e.g.) OWL axioms to describe classes and properties that are shared across multiple lines of business so concepts can be defined by their relationships, making them extensible to increasing levels of detail as required.
  - Good ontologies are ‘fractal’ in nature, meaning that the common abstractions create an organizing structure that easily expands to accommodate the complex information management requirements of the business.
- The purpose of a logical model is to describe the structure of the data required for a particular application or service.
  - Typically a logical model shows all the entities, relationships and attributes required for a proposed application.
  - It only includes data relevant to the particular application in question.
  - Ideally logical models are derived from the ontology which ensures consistent meaning and naming across future information systems.

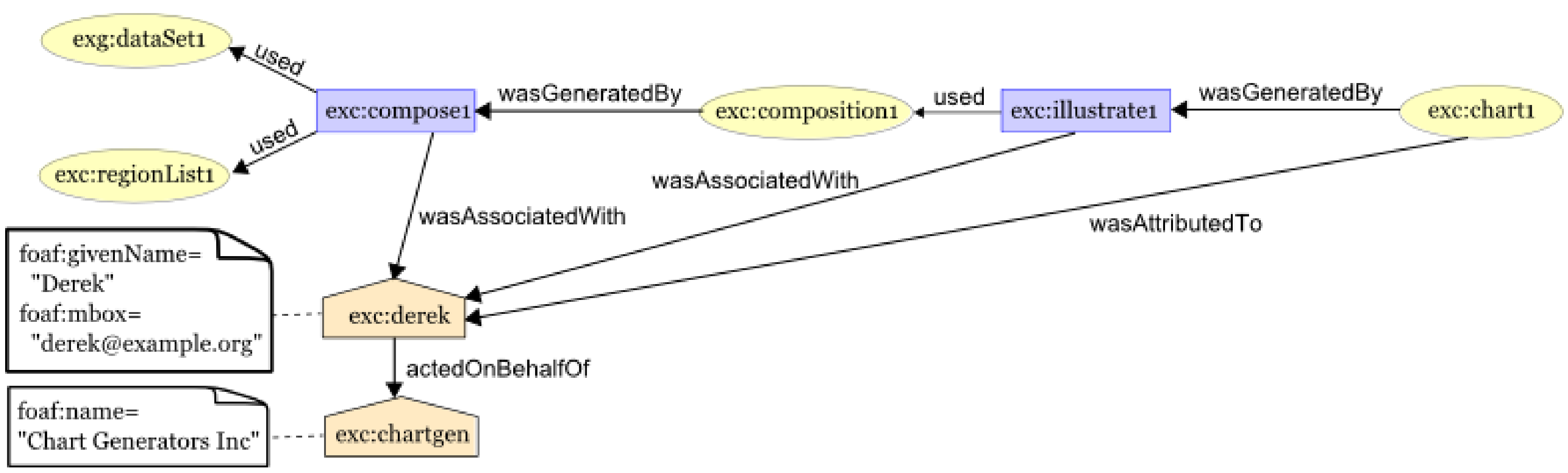
Ontologies

- An ontology is a formal explicit description of concepts in a domain of discourse
  - Classes
    - sometimes called concepts
  - properties of each concept describing various features and attributes of the concept
    - slots/roles/properties
  - and restrictions on slots
    - facets/role restrictions
- An ontology together with a set of individual instances of classes constitutes a knowledge base.
- There is a fine line where the ontology ends and the knowledge base begins.



Ontology (example)





Vocabularies

- On the Semantic Web, vocabularies define the concepts and relationships (also referred to as “terms”) used to describe and represent an area of concern.
- In the driest sense, a "vocabulary" is a context-less list of terms, with no defined interrelationships. "Ontology" is meatier, implying the presence of interrelationships, axioms, classes, etc.
- Vocabularies are used to classify the terms that can be used in a particular application, characterize possible relationships, and define possible constraints on using those terms.
  - Vocabularies can be very complex (with several thousands of terms) or very simple (describing one or two concepts only).
- There is no clear division between what is referred to as “vocabularies” and “ontologies”.
  - The trend is to use the word “ontology” for more complex, and possibly quite formal collection of terms, whereas “vocabulary” is used when such strict formalism is not necessarily used or only in a very loose sense.
    - Vocabularies are the basic building blocks for inference techniques on the Semantic Web.
  - The fundamental difference between an ontology and a controlled vocabulary is the level of abstraction and relationships among concept.
    - A formal ontology is a controlled vocabulary expressed in an ontology representation language.
    - This language has a grammar for using vocabulary terms to express something meaningful within a specified domain of interest.

Vocabularies examples

IVOA Vocabulary: Content levels for VO resources

This is the description of the namespace `http://www.ivoa.net/rdf/voresource/content_level` as of 2016-08-17.

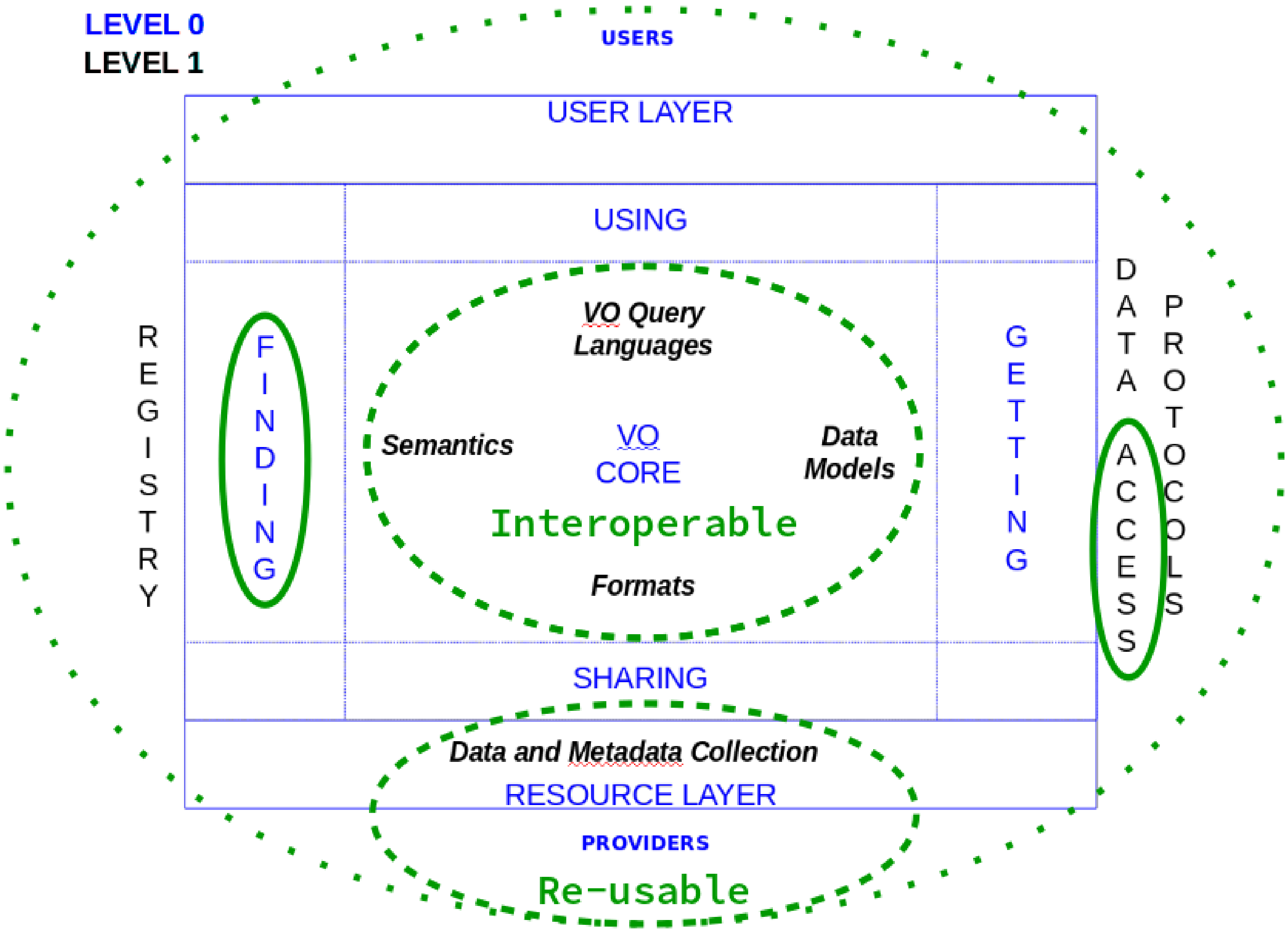
This vocabulary enumerates the intended audiences for resources in the Virtual Observatory. It is designed to enable discovery queries like "only research-level data" or "resources usable in school settings".

Predicate	Label	Description	Parent	Preferred
Research	Research	Resource provides information appropriate for supporting scientific research.		
Amateur	Amateur	Resource provides information of interest to amateur astronomers.		
General	General	Resource provides information appropriate for use in outreach to and education of the general public		

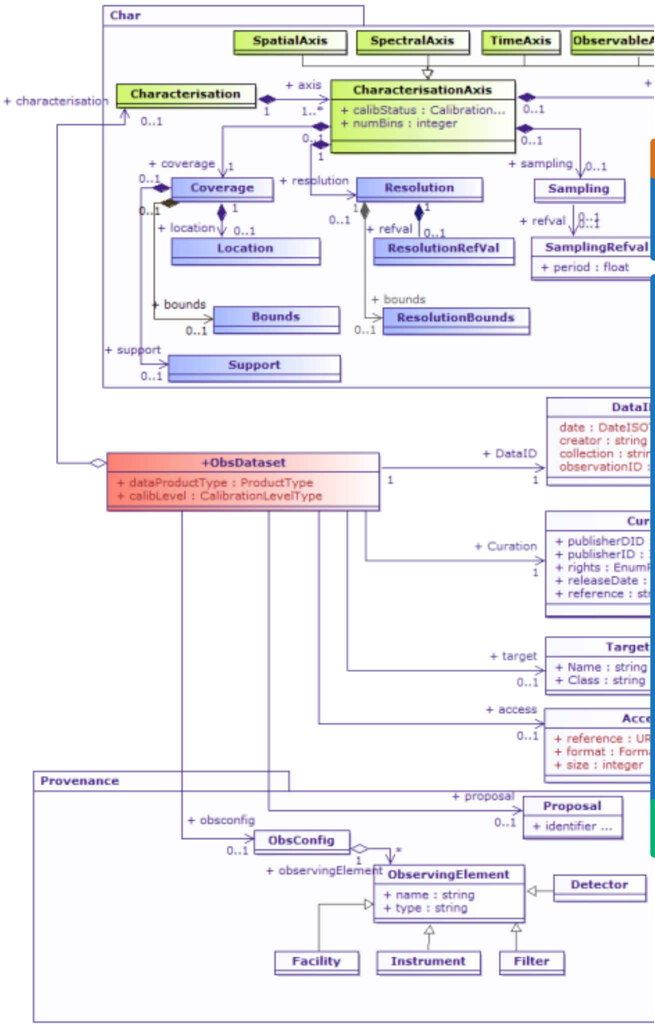
Alternate formats: RDF, Turtle.

```
@base <http://www.ivoa.net/rdf/voresource/content_level>.
@prefix : <#>.

@prefix dc: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/>.

<> a owl:Ontology;
        dc:created "2016-08-17";
        dc:creator [ foaf:name "Ray Plante" ],
    [ foaf:name "Markus Demleitner" ];
        rdfs:label "Content levels for VO resources"@en;
        dc:title "Content levels for VO resources"@en;
        dc:description """This vocabulary enumerates the intended audiences
for resources in the Virtual Observatory.  It is designed to
enable discovery queries like "only research-level data" or
"resources usable in school settings".""".

dc:created a owl:AnnotationProperty.
dc:creator a owl:AnnotationProperty.
dc:title a owl:AnnotationProperty.
dc:description a owl:AnnotationProperty.
<#Research> a rdf:Property;
  rdfs:label "Research";
  rdfs:comment "Resource provides information appropriate for supporting scientific research.".

<#Amateur> a rdf:Property;
  rdfs:label "Amateur";
  rdfs:comment "Resource provides information of interest to amateur astronomers.".

<#General> a rdf:Property;
  rdfs:label "General";
  rdfs:comment "Resource provides information appropriate for use in outreach to and education of the general public".
```

VRE for Astrophysics: IVOA



IVOA: Interoperable (1)



Column Name	*Unit*	*Type*	*Description*
dataproduct_type	unitless	String	Logical data product type (image etc.)
calib_level	unitless	enum integer	Calibration level {0, 1, 2, 3, 4}
obs_collection	unitless	String	Name of the data collection
obs_id	unitless	String	Observation ID
obs_publisher_did	unitless	String	Dataset identifier given by the publisher
access_url	unitless	String	URL used to access (download) dataset
access_format	unitless	String	File content format (see in App. BB.5.2)
access_estsize	kbyte	integer	Estimated size of dataset in kilo bytes
target_name	unitless	String	Astronomical object observed, if any
s_ra	deg	double	Central right ascension, ICRS
s_dec	deg	double	Central declination, ICRS
s_fov	deg	double	Diameter (bounds) of the covered region
s_region	unitless	String	Sky region covered by the data product (expressed in ICRS frame)
s_xel1	unitless	integer	Number of elements along the first spatial axis
s_xel2	unitless	integer	Number of elements along the second spatial axis
s_resolution	arcsec	double	Spatial resolution of data as FWHM
t_min	d	double	Start time in MJD
t_max	d	double	Stop time in MJD
t_exptime	s	double	Total exposure time
t_resolution	s	double	Temporal resolution FWHM
t_xel	unitless	integer	Number of elements along the time axis
em_min	m	double	Start in spectral coordinates
em_max	m	double	Stop in spectral coordinates
em_res_power	unitless	double	Spectral resolving power
em_xel	unitless	integer	Number of elements along the spectral axis
o_ucd	unitless	String	UCD of observable (e.g. phot.flux.density, phot.count, etc.)
pol_states	unitless	String	List of polarization states or NULL if not applicable
pol_xel	unitless	integer	Number of polarization samples
facility_name	unitless	String	Name of the facility used for this observation
instrument_name	unitless	String	Name of the instrument used for this observation

example
Obs. Core Data Model

← identifier (obs_publisher_did)

} dataset access (access_url, access_format, access_estsize)

} spatial (s_ra … s_resolution)

} temporal (t_min … t_xel)

} spectral (em_min … em_xel)

← Unified Content Descriptor (o_ucd)

IVOA: Interoperable (2)

- Unified Content Descriptors
  - Broad terms
- Units
  - standardized
- Data Model link/identifiers
  - “utype”(s)
- Vocabularies
  - General
  - Domain specific
  - Recommendation driven

ucd=”pos.eq.ra;meta.main”
unit=”m.s-1”**

- UCD Tree
  - arith
  - em
  - instr
  - meta
  - obs
    - obs.airMass
    - obs.atmos
    - obs.calib
    - obs.exposure
    - obs.field
    - obs.image
    - obs.observer
    - obs.param
    - obs.proposal
    - obs.sequence
  - phot
  - phys
  - pos
  - spect
    - spect.binSize
    - spect.continuum
    - spect.dopplerParam
    - spect.dopplerVeloc
    - spect.index
    - spect.line
    - spect.resolution
  - src
    - src.calib
    - src.class
    - src.density
    - src.ellipticity
    - src.impactParam
    - src.morph
    - src.net
    - src.orbital
    - src.redshift
    - src.sample
    - src.spType
    - src.var
  - stat
  - time

m	(metre)	g	(gram)	J	(joule)	Wb	(weber)
s	(second)	rad	(radian)	W	(watt)	T	(tesla)
A	(ampere)	sr	(steradian)	C	(coulomb)	H	(henry)
K	(kelvin)	Hz	(hertz)	V	(volt)	lm	(lumen)
mol	(mole)	N	(newton)	S	(siemens)	lx	(lux)
cd	(candela)	Pa	(pascal)	F	(farad)	Ohm	(ohm)

min	(minute of time)	deg	(degree of angle)	Jy	(jansky)
h	(hour of time)	arcmin	(arcminute)	pc	(parsec)
d	(day)	arcsec	(arcsecond)	eV	(electron volt)
a, yr	(year)	mas	(milliarcsecond)	AU	(astronomical unit)
u	(atomic mass)				

Datalink core ontology

This is the description of the namespace http://www.ivoa.net/rdf/datalink/core/core as of 201

Terms in this vocabulary are intended for use in the semantics column in the output from the DataLinl

As specified in DataLink-1.0, terms from the vocabulary may be used in the Datainl output using onl http://www.ivoa.net/rdf/datalink/core#word).

Alternate formats: RDF TTL

Predicate	Parent	Label	Comment
#this		the data itself	the primary (as opposed to related) data of the
#progenitor		Progenitor	data resources that were used to create this da
#derivation		Derivation	data resources that are derived from this datas
#auxiliary		Auxiliary	auxiliary resources
#weight	#auxiliary	Weight map	resource with array(s) containing weighting v
#error	#auxiliary	Error map	resource with array(s) containing error values
#noise	#auxiliary	Noise map	resource with array(s) containing noise values
#calibration		Calibration data	resource used to calibrate the primary data
#bias	#calibration	Bias calibration data	used to subtract the detector offset level
#dark	#calibration	Dark calibration data	used to subtract the accumulated detector darl
#flat	#calibration	Flat field calibration data	used to calibrate variations in detector sensiti
#preview		Preview	low fidelity but easily viewed representation o
#preview-image	#preview	Image preview	preview of the data as a 2-dimensional image
#preview-plot	#preview	Plot preview	preview of the data as a plot (e.g. spectrum or
#proc		Processing	server-side data processing result
#cutout	#proc	Cutout	a subsection of the primary data

Interoperability (example)

- Vizier
  - http://vizier.u-strasbg.fr/viz-bin/VizieR
  - catalogue search by UCD
  - Column descriptions
- Datalink semantics in the response
  - GAVO (TOPCAT → TAP → ObsTAP example)
    - SAMP (interoperations)
- Registry
  - Content Level (TOPCAT RegTAP)
    - Usage reflected in vocabulary update
      - Cfr.: http://ivoa.net/rdf/voresource/content_level/2016-08-17/content_level.html