Update on Vocabularies

Vocabulary management and vocabulary content

Ongoing process to move controlled-value lists into Semantics.

- Ongoing process to move controlled-value lists into Semantics.
- Vocabularies will evolve in Semantics using VEPs.
 - New terms will appear, some will be deprecated or replaced. With possible impact on interoperability.
 - Need for referring to specific versions of a vocabulary?

- Ongoing process to move controlled-value lists into Semantics.
- Vocabularies will evolve in Semantics using VEPs.
 - New terms will appear, some will be deprecated or replaced. With possible impact on interoperability.
 - Need for referring to specific versions of a vocabulary?
- Versioning of vocabularies are currently based on the released date.
 - Except for the UCD lists, which is using a "X.Y" versioning scheme.

- Ongoing process to move controlled-value lists into Semantics.
- Vocabularies will evolve in Semantics using VEPs.
 - New terms will appear, some will be deprecated or replaced. With possible impact on interoperability.
 - Need for referring to specific versions of a vocabulary?
- Versioning of vocabularies are currently based on the released date.
 - Except for the UCD lists, which is using a "X.Y" versioning scheme.
- Semantic interoperability and FAIR principles:
 - Ontologies (hence vocabularies) should be versioned;
 - URIs should be used instead of terms (less implicit).

- Ongoing process to move controlled-value lists into Semantics.
- Vocabularies will evolve in Semantics using VEPs.
 - New terms will appear, some will be deprecated or replaced. With possible impact on interoperability.
 - Need for referring to specific versions of a vocabulary?
- Versioning of vocabularies are currently based on the released date.
 - Except for the UCD lists, which is using a "X.Y" versioning scheme.
- Semantic interoperability and FAIR principles:
 - Ontologies (hence vocabularies) should be versioned;
 - URIs should be used instead of terms (less implicit).
- Proposal: Enforce using full URIs when possible (hence including version).
 - If "term" only is used (instead of full "URI"): this would mean the latest version of the vocabulary should be used.

Vocabulary versions and full URIs - example

• https://voparis-ns.obspm.fr/rdf/epn/2.0/product-type redirects to latest version with link to previous one.

PADC Vocabulary: Product Types for EPNcore metadata

This is the description of the vocabulary http://voparis-ns.obspm.fr/rdf/epn/2.0/product-type as of 2024-03-28.

The previous version is available at http://voparis-ns.obspm.fr/rdf/epn/2.0/product-type/2023-11-11.

This vocabulary is not yet approved by the IVOA. This means that terms can still disappear without prior notice.

The EPNcore product-type parameter describes the high level scientific organization of the data product linked by the access_url parameter, or directly included in the table (in which case the value is 'ci' for catalogue_item). EPNCore currently defines several types listed below. The data provider must select the type most adapted to his data. In complex situations (e. g., when a file contains several data products), several types can be used to describe the same granule by using a hash-separated-list — although using several granules to describe the file content may be a better solution. In EPN-TAP these types are identified by a 2-characters ID, so that multivalued queries are unambiguous.

```
<> a owl:Ontology;
    owl:versionIRI <http://voparis-ns.obspm.fr/rdf/epn/2.0/spatial-frame-type/2024-05-07> ;
    owl:priorVersion < http://voparis-ns.obspm.fr/rdf/epn/2.0/spatial-frame-type/2023-11-11>
    dc:created "2024-05-07";
    vann:preferredNamespacePrefix "epnsftyp";
    dc:creator [ foaf:name "Erard, S." ],
    [ foaf:name "Cecconi, B." ];
    dc:license <http://creativecommons.org/publicdomain/zero/1.0/>;
    rdfs:label "EPNcore Spatial Frame Type"@en;
    dc:title "EPNcore Spatial Frame Type"@en;
    dc:description """Provides the "flavor" of the coordinate system, which defines the
nature of the spatial coordinates (c1,c2,c3) in the EPNCore table and queries,
and the way they are defined. A value is always required (use "none" if not
applicable, although "body" is found in older services and may be OK). This may
be different from the coordinate system associated to / included in the data
themselves. The reference frame itself is defined by the
spatial_coordinate_description parameter, and the frame center can be specified
using the spatial_origin parameter in case of ambiguity."";
    ivoasem:vocflavour "SKOS".
```

- Currently, the "Vocabularies" git repository keeps the vocabulary source in two places:
 - 1 CSV file per vocabulary for the list of terms, the hierarchy and the relations
 - 1 configuration file containing the metadata for all vocabularies

- Currently, the "Vocabularies" git repository keeps the vocabulary source in two places:
 - 1 CSV file per vocabulary for the list of terms, the hierarchy and the relations
 - 1 configuration file containing the metadata for all vocabularies
- The "convert.py" script builds the content of the vocabulary pages (https://ivoa.net/ rdf/), including HTML, RDF, Turtle and desise outputs.

- Currently, the "Vocabularies" git repository keeps the vocabulary source in two places:
 - 1 CSV file per vocabulary for the list of terms, the hierarchy and the relations
 - 1 configuration file containing the metadata for all vocabularies
- The "convert.py" script builds the content of the vocabulary pages (https://ivoa.net/
 rdf/), including HTML, RDF, Turtle and desise outputs.
- In some recent cases (e.g., the ObsFacility vocabulary, or the draft UCD vocabulary), storing the source as a RDF file would allow to store more information, used for different purposes:
 - e.g., for ObsFacility, the RDF file contains all the known aliases (with a lang tag).

- Currently, the "Vocabularies" git repository keeps the vocabulary source in two places:
 - 1 CSV file per vocabulary for the list of terms, the hierarchy and the relations
 - 1 configuration file containing the metadata for all vocabularies
- The "convert.py" script builds the content of the vocabulary pages (https://ivoa.net/
 rdf/), including HTML, RDF, Turtle and desise outputs.
- In some recent cases (e.g., the ObsFacility vocabulary, or the draft UCD vocabulary), storing the source as a RDF file would allow to store more information, used for different purposes:
 - e.g., for ObsFacility, the RDF file contains all the known aliases (with a lang tag).
- Modification of the convert.py script to build (https://ivoa.net/rdf/) is already done (my fork of repo).

Vocabulary management vocabulary API early discussion

• Currently, the vocabularies are accessible as static html (which is good, because it's simple).

Vocabulary management vocabulary API early discussion

- Currently, the vocabularies are accessible as static html (which is good, because it's simple).
- However, if we want to build to tools helping research teams to use the IVOA vocabulary terms is open science context, we probably need a search interface.
 - use-case data publication: data steward / documentalist fills in a DMP for a science team, including our vocabulary terms (i.e.: IVOA-UAT, RefFrame, ObsFacility...), so that they can be inserted in Datacite metadata, in linked-data metadata...
 - use-case FAIR assessment: having IVOA vocabulary terms used in DOI landing pages metadata (using URIs), makes dataset more easily finable and interoperable (e.g.: using standard RefFrame or ObsFacility terms)
 - use-case refined search: search for data from a specific ObsFacility (e.g.: LOFAR), enabling extension of the query to parts of the facility (individual LOFAR stations throughout Europe)

Vocabulary management vocabulary API early discussion

- Currently, the vocabularies are accessible as static html (which is good, because it's simple).
- However, if we want to build to tools helping research teams to use the IVOA vocabulary terms is open science context, we probably need a search interface.
 - use-case data publication: data steward / documentalist fills in a DMP for a science team, including our vocabulary terms (i.e.: IVOA-UAT, RefFrame, ObsFacility...), so that they can be inserted in Datacite metadata, in linked-data metadata...
 - use-case FAIR assessment: having IVOA vocabulary terms used in DOI landing pages metadata (using URIs), makes dataset more easily finable and interoperable (e.g.: using standard RefFrame or ObsFacility terms)
 - use-case refined search: search for data from a specific ObsFacility (e.g.: LOFAR), enabling extension of the query to parts of the facility (individual LOFAR stations throughout Europe)
- need for an programmatic interface (API), example:
 - search for a term: https://api.[...]/search?q=magnitude&ontology=ivoa-uat (ontoportal API style)
 - need for capability to get the metadata for a term (e.g.: to get the its relations) without loading the full vocabulary.

IVOA Semantics (ivoasem) new mappings

proposed modifications:

IVOA Semantics (ivoasem) new mappings

- proposed modifications:
 - "deprecated" property usage is defined the following way:
 <term> ivoasem:deprecated ":___".

A more standard way to say that is:

<term> a owl:DeprecatedClass.

Non breaking change could be to declare this in the ivoasem RDF file

owl:DeprecatedClass owl:equivalentClass [rdf:type owl:Restriction; owl:onProperty ivoasem:deprecated; owl:hasValue ":__"].

IVOA Semantics (ivoasem) new mappings

- proposed modifications:
 - "deprecated" property usage is defined the following way:
 <term> ivoasem:deprecated ":__".

A more standard way to say that is:

<term> a owl:DeprecatedClass.

Non breaking change could be to declare this in the ivoasem RDF file

owl:DeprecatedClass owl:equivalentClass [rdf:type owl:Restriction; owl:onProperty ivoasem:deprecated; owl:hasValue ":__"].

"useInstead" property indicates what term should be used when a term is deprecated.
 We could add the following statement in RDF to make sure external libraries understand what to do:

ivoasem:useInstead owl:equivalentProperty dct:isReplacedBy.

- 3 lists of terms are in a state to be released (in preliminary state):
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/product-type/
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/spatial-frame-type/
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/target-class/

- 3 lists of terms are in a state to be released (in preliminary state):
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/product-type/
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/spatial-frame-type/
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/target-class/
- 1 list of terms from an EPN-TAP extension is being prepared:
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/small-body-dynamical-class/

- 3 lists of terms are in a state to be released (in preliminary state):
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/product-type/
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/spatial-frame-type/
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/target-class/
- 1 list of terms from an EPN-TAP extension is being prepared:
 - https://voparis-ns.obspm.fr/rdf/epn/2.0/small-body-dynamical-class/
- not pushed yet to ivoa-std/Vocabularies repo:
 - need to push updates of convert.py script first, to include new relations
 - need to check relations with existing IVOA vocabularies

A proposal

The list of UCD words is a special vocabulary.
 The UCD words can be used by themselves, but most of the cases are using composed UCD words to express concepts.

A proposal

- The list of UCD words is a special vocabulary.
 The UCD words can be used by themselves, but most of the cases are using composed UCD words to express concepts.
- An vocabulary of UCD words is not difficult to prepare, but needs to keep the composition rule types
 of each word.

A proposal

- The list of UCD words is a special vocabulary.
 The UCD words can be used by themselves, but most of the cases are using composed UCD words to express concepts.
- An vocabulary of UCD words is not difficult to prepare, but needs to keep the composition rule types
 of each word.
- Proposal
 - define UCD classes by their composition rules:
 - P = Primary word
 - S = Secondary word
 - -Q = P or S
 - C, V and E are Q with special meaning (Color, Vector and Photometry)
 - add a "has_ucd" property with "domain" = UCD words
 - then declare UCD words as individuals of a UCD class, e.g.:

```
ucd:arith.diff a ucd:S ;
    rdfs:label "arith.diff"@en ;
    rdfs:comment "Difference between two quantities described by the same UCD"@en .
```

A proposal

Note the declaration of the UCD classes would look like this:

```
ucd:C rdfs:label "Colour Index"@en ;
   rdfs:comment "A colour index, and can be followed by two successive word describing a part of the electromagnetic spectrum"@en ;
   rdfs:subClassOf ucd:Q .
ucd:V rdfs:label "Vector"@en ;
   rdfs:comment "Such a word can be followed by another describing the axis or reference frame in which the measurement is done"@en ;
   rdfs:subClassOf ucd:Q .
ucd:E rdfs:label "Photometric Quantity"@en ;
   rdfs:comment "A photometric quantity, and can be followed by a word describing a part of the electromagnetic spectrum"@en ;
   rdfs:subClassOf ucd:Q .
ucd:P a owl:Class;
   rdfs:label "Primary"@en ;
   rdfs:comment "The word can only be used as ``primary'' or first word."@en .
ucd:S a owl:Class;
   rdfs:label "Secondary"@en ;
   rdfs:comment "The word cannot be used as the first word to describe a single quantity"@en .
ucd:Q a owl:Class;
   rdfs:label "Primary or Secondary"@en ;
   rdfs:comment "The word can be used indifferently as first or secondary word."@en ;
    owl:equivalentClass [ a owl:Class ;
            owl:unionOf ( ucd:P ucd:S ) ] .
```

• Examples:

```
<col1> has_ucd ucd:pos.ra .
<col2> has_ucd ucd:pos.dec .
```

Examples:

```
<col1> has_ucd ucd:pos.ra .
<col2> has_ucd ucd:pos.dec .
```

- Impact:
 - need to revise VocInVO document, because we don't have (yet) a vocabulary defined as a series of classes and individuals.
 - check how this can be exposed into desise.

Examples:

```
<col1> has_ucd ucd:pos.ra .
<col2> has_ucd ucd:pos.dec .
```

- Impact:
 - need to revise VocInVO document, because we don't have (yet) a vocabulary defined as a series of classes and individuals.
 - check how this can be exposed into desise.
- Other options:
 - define a UCD vocabulary as a list of Classes (each word is a class, but that doesn't fit so well with RDF), and UCD composition flag would be set with a property.

Examples:

```
<col1> has_ucd ucd:pos.ra .
<col2> has_ucd ucd:pos.dec .
```

- Impact:
 - need to revise VocInVO document, because we don't have (yet) a vocabulary defined as a series of classes and individuals.
 - check how this can be exposed into desise.
- Other options:
 - define a UCD vocabulary as a list of Classes (each word is a class, but that doesn't fit so well with RDF), and UCD composition flag would be set with a property.
- Not sure we can do:

How to describe composite UCDs (e.g.: "phot.count;em.IR")? we may be able to write:

<col3> has_ucd (ucd:phot.count ucd:em.IR) . # this is an ordered list
but I need to check this...

Southern spring clean-up?

Current RefFrame vocabulary is limited to a set of terms:
 AZ_EL, BODY, ECLIPTIC, EQUATORIAL, FK4, FK5, GALACTIC, GALACTIC_I,
 GENERIC_GALACTIC, ICRS, SUPER_GALACTIC, UNKNOWN
 plus a series of (already) deprecated terms:
 barycentric, ecl_FK4, ecl_FK5, eq_FK4, eq_FK5, galactic, geo_app,
 supergalactic, xy

Southern spring clean-up?

- Current RefFrame vocabulary is limited to a set of terms:
 AZ_EL, BODY, ECLIPTIC, EQUATORIAL, FK4, FK5, GALACTIC, GALACTIC_I,
 GENERIC_GALACTIC, ICRS, SUPER_GALACTIC, UNKNOWN
 plus a series of (already) deprecated terms:
 barycentric, ecl_FK4, ecl_FK5, eq_FK4, eq_FK5, galactic, geo_app,
 supergalactic, xy
- Question:

since we start a new vocabulary, do we need to keep those deprecated terms?

Southern spring clean-up?

- Current RefFrame vocabulary is limited to a set of terms:
 AZ_EL, BODY, ECLIPTIC, EQUATORIAL, FK4, FK5, GALACTIC, GALACTIC_I,
 GENERIC_GALACTIC, ICRS, SUPER_GALACTIC, UNKNOWN
 plus a series of (already) deprecated terms:
 barycentric, ecl_FK4, ecl_FK5, eq_FK4, eq_FK5, galactic, geo_app,
 supergalactic, xy
- Question:
 since we start a new vocabulary, do we need to keep those deprecated terms?
- Special question: the geo_app term comes from very far away back in time (even before votable: Sébastien traced it back in astrores and before). Shouldn't we get rid of this one?

Inputs from helio and planetary sciences

 Many additional RefFrame terms can be found in STC 1.33, for solar and planetary frames.

Inputs from helio and planetary sciences

- Many additional RefFrame terms can be found in STC 1.33, for solar and planetary frames.
- Heliophysics worked hard to come up with a list of RefFrame terms (but not ready yet). About unique 100 RefFrames identified.
 - All proposed terms are listed in a few selected source of information: online tools (for coordinate transforms) and lists in curated data models.

Inputs from helio and planetary sciences

- Many additional RefFrame terms can be found in STC 1.33, for solar and planetary frames.
- Heliophysics worked hard to come up with a list of RefFrame terms (but not ready yet). About unique 100 RefFrames identified.
 All proposed terms are listed in a few selected source of information: online tools (for coordinate transforms) and lists in curated data models.
- Planetary Sciences:
 - A list of planetary Coordinate Reference Systems have been adopted by the OGC ecosystem (Earth sciences). See here: http://voparis-vespa-crs.obspm.fr:8080/web/
 - This list includes references frames and projections.
 - Importing the reference frames into RefFrame should be easy (24 frames).

Almost there!

- Long standing action: gathering a list of terms for observation facilities.
- Required for interoperability and discoverability of dataset based on telescope name or spacecraft name.
- Semantic reminder:
 Observation Facility = the concept bearing the location of the observation
 Instrument = the concept bearing the instrumental capabilities (fov, filter...)

Almost there!

- Long standing action: gathering a list of terms for observation facilities.
- Required for interoperability and discoverability of dataset based on telescope name or spacecraft name.
- Semantic reminder:
 Observation Facility = the concept bearing the location of the observation
 Instrument = the concept bearing the instrumental capabilities (fov, filter...)
- Current scope:
 - a reference list of terms for interoperability.
 - a name resolver to help finding the right term to be used.
 - hierarchy (has_part, is_part_of)

Almost there!

- Long standing action: gathering a list of terms for observation facilities.
- Required for interoperability and discoverability of dataset based on telescope name or spacecraft name.
- Semantic reminder:
 Observation Facility = the concept bearing the location of the observation
 Instrument = the concept bearing the instrumental capabilities (fov, filter...)
- Current scope:
 - a reference list of terms for interoperability.
 - a name resolver to help finding the right term to be used.
 - hierarchy (has_part, is_part_of)
- Not in scope:
 - modelling the observation facility (classification, metadata, relation to instruments)
 - instruments

Almost there!

• Sources: SPASE, PDS, ObsCode, NSSDC, NAIF, AAS, etc, and Wikidata Wikidata has a lot of identifiers already mapped together: great starting point. But very difficult to curate not being part of wikidata community.

Almost there!

• Sources: SPASE, PDS, ObsCode, NSSDC, NAIF, AAS, etc, and Wikidata Wikidata has a lot of identifiers already mapped together: great starting point. But very difficult to curate not being part of wikidata community.

Process:

- Export from Wikidata and merge additional terms and aliases and produce a curated list of terms with aliases.
- list of terms + external identifiers + relations => IVOA Vocabulary
- list of terms + known aliases / ids => Name resolver

Almost there!

• Sources: SPASE, PDS, ObsCode, NSSDC, NAIF, AAS, etc, and Wikidata Wikidata has a lot of identifiers already mapped together: great starting point. But very difficult to curate not being part of wikidata community.

Process:

- Export from Wikidata and merge additional terms and aliases and produce a curated list of terms with aliases.
- list of terms + external identifiers + relations => IVOA Vocabulary
- list of terms + known aliases / ids => Name resolver

• Plan:

- first version with terms having an ObsCode or a SPASE id or a PDS id (almost there)
- decisions to be made on the accepted terms, curation of label and description
- second version making sure that all VOResource/ObsCore/EPNcore observation facility values can be mapped to a term in the vocabulary.
- then: updates through VEP.

Example

```
obs:gemini-north a owl:Class ;
    rdfs:label "Gemini North"@en ;
    rdfs:comment "northern hemisphere facility of the Gemini Observatory (wikiata:Q6140627) (obscode:T15)"@en ;
    skos:altLabel "Q6140627",
        "T15",
        "frederick c. gillett gemini telescope"@en,
        "gemini north observatory"@en,
        "gemini north telescope"@en,
        "the frederick c. gillett gemini telescope"@en ;
    skos:exactMatch < <a href="http://www.wikidata.org/entity/06140627">http://www.wikidata.org/entity/06140627</a>>,
         <https://minorplanetcenter.net/iau/lists/ObsCodesF.html#T15>,
         <<u>urn:nαsα:pds:context:facility:observatory.gemini north-maunakea</u>> .
obs:gemini-south a owl:Class ;
    rdfs:label "Gemini South"@en ;
    rdfs:comment "southern hemisphere facility of the Gemini Observatory (wikiata:Q19673584) (obscode:I11)"@en ;
    skos:altLabel "I11",
        "Q19673584",
        "gemini south observatory"@en,
        "gemini south telescope"@en,
         "gems"@en ;
    skos:exactMatch < http://www.wikidata.org/entity/019673584>,
         <https://minorplanetcenter.net/iau/lists/ObsCodesF.html#I11>,
         <urn:nasa:pds:context:facilitu:observatorv.gemini-south> .
```

- Why:
 - EPN-TAP has a "map_projection" keyword, not constrained.
 - HiPS should contain metadata tracing the original projection of the dataset (complain from planetary science geek)
- Need for a vocabulary?

- Why:
 - EPN-TAP has a "map_projection" keyword, not constrained.
 - HiPS should contain metadata tracing the original projection of the dataset (complain from planetary science geek)
- Need for a vocabulary?
- Recent work with OGC: large list of projections available (well defined, but we
 just need the terms).

- Why:
 - EPN-TAP has a "map_projection" keyword, not constrained.
 - HiPS should contain metadata tracing the original projection of the dataset (complain from planetary science geek)
- Need for a vocabulary?
- Recent work with OGC: large list of projections available (well defined, but we
 just need the terms).
- FITS WCS document also has list of projections.

- Why:
 - EPN-TAP has a "map_projection" keyword, not constrained.
 - HiPS should contain metadata tracing the original projection of the dataset (complain from planetary science geek)
- Need for a vocabulary?
- Recent work with OGC: large list of projections available (well defined, but we
 just need the terms).
- FITS WCS document also has list of projections.
- Let's start this.

Processing-level Vocabulary new proposal

- Why:
 - Obscore has a "calibration-level" list of terms.
 - EPN-TAP has a "processing-level" keyword, with a constrained list (different from obscore)

Processing-level Vocabulary new proposal

- Why:
 - Obscore has a "calibration-level" list of terms.
 - EPN-TAP has a "processing-level" keyword, with a constrained list (different from obscore)
- "Processing-level" is semantically wider than "calibration-level":
 - "calibrated" is a processing level
 - "derived" is not really a calibration level (but it is a processing level)

Processing-level Vocabulary new proposal

- Why:
 - Obscore has a "calibration-level" list of terms.
 - EPN-TAP has a "processing-level" keyword, with a constrained list (different from obscore)
- "Processing-level" is semantically wider than "calibration-level":
 - "calibrated" is a processing level
 - "derived" is not really a calibration level (but it is a processing level)
- Proposal by Laurent Michel with obscore terms.
 EPNcore terms to be added and mapped.
 Other terms: CODMAC, PDS3, PDS4... should also be mapped