



Open Information Linking for Environmental science research infrastructures (OIL-E)

Paul Martin and Zhiming Zhao

For ENVRiplus (Horizon 2020 project #654182)

The ENVRIplus project

ENVRIplus (<http://envriplus.eu/>) builds upon the work of the original ENVRI project:

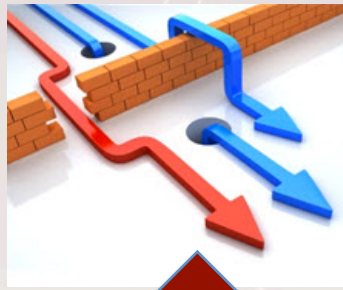
- Providing **shared solutions for science and society**.
- Defining **common operations for environmental research infrastructures (RIs)**.

The **Data for Science** theme addresses the need for **interoperable services**:

- Addressing six topics of *identification and citation, curation, cataloguing, processing, optimisation and provenance*.
- Guided by a common *reference model (ENVRI RM)*, *semantic linking framework (OIL-E)* and *architecture design*.



Challenge 1: support
system level of
sciences



Challenge 2: share
solutions to common
problems



Challenge 3: Interface
virtual research
environment(s)



Challenge 4: re-use technologies
(e.g. from e-Infrastructures)





Requires common controlled vocabulary.

Challenge 1: support system level of sciences

Challenge 2: share solutions to common problems

Challenge 3: Interface virtual research environment(s)

Requires functional federation of infrastructures.

Requires standard APIs.



Challenge 4: re-use technologies (e.g. from e-Infrastructures)

Requires smart services.





OIL-E describes RIs and their components to produce a semantic landscape of RIs.

Challenge 1: support system level of science

Challenge 2: share solutions to common problems

Challenge 3: Interface virtual research environment(s)

ENVRI RM can be used to help standardise RI descriptions and identify compatible interfaces and points of overlap.

OIL-E can be used to identify the APIs (and API standards) used by RIs, VREs and e-Is.



Challenge 4: re-use technologies (e.g. from e-Infrastructures)

OIL-E will provide a knowledge layer for use by services that associates technology profiles with different workflows/activities.



clouds

ENVRI Reference Model

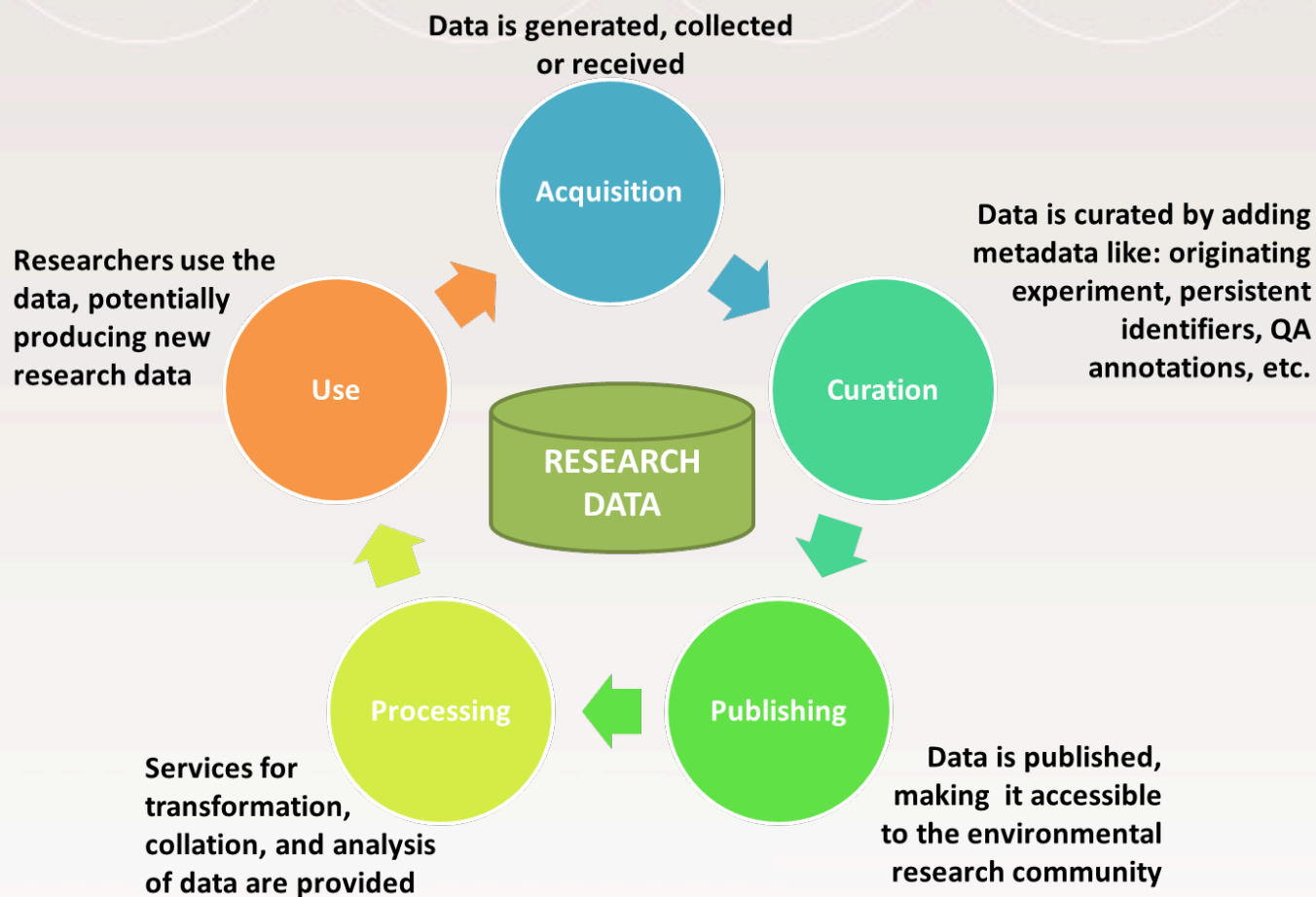
The ENVRI Reference Model (<http://envri.eu/rm/>) provides a standard set of **archetypes** for different classes of actor, information object, behaviour, etc. commonly found within environmental science RIs.

- The reference model (**ENVRI RM**) places these archetypes in the context of the **research data lifecycle**, identifying the critical elements of *data acquisition, curation, publishing, processing and use*.

By referring to ENVRI RM, RI architects can:

- Identify which elements are most important to them.
- Determine any gaps within their own (planned) infrastructure.
- Compare against other RI specifications—in particular looking at *how* other RIs solved the same problems and what technologies they used.

ENVRI research data lifecycle



ENVRI RM viewpoints

The main actors in a research infrastructure and their behaviours.

Science
viewpoint

The data entities found in research infrastructures and their state transitions.

Information
viewpoint

The computing elements necessary to support research infrastructure operations.

Computational
viewpoint

The distribution of computing elements on underlying resources and their communication channels.

Engineering
viewpoint

systems &
their environment

The technologies and standards used to implement the systems of the infrastructure.

Technology
viewpoint



Common vocabularies

- There are a large number of controlled vocabularies and vocabulary management systems (VMSs) being used by RIs.
- For example in the ecosystem domain:
 - EnvThes (<http://vocabs.ceh.ac.uk/evn/tbl/envthes.evn>).
 - GEMET (<http://www.eionet.europa.eu/gemet>).
 - BioPortal (<https://bioportal.bioontology.org/>).
- Also plenty of general metadata schemes of importance, such as:
 - ISO 19115 (and 19139) w/ INSPIRE (<http://inspire-geoportal.ec.europa.eu/>)
 - DCAT-AP (https://joinup.ec.europa.eu/asset/dcat_application_profile/description)
 - CERIF (specifically classification schemes;
<http://www.eurocris.org/cerif/main-features-cerif>)
- Many of these vocabularies can be (and in practice, are) used to describe various objects of concern to RIs.

Common vocabularies

ENVRI RM provides **standardised vocabulary** for defining:

- The major **actors** in an RI and the **behaviours** they engage in.
- The main **information objects** generated and used within an RI.
- The main **services and resources** needed to support RI **operations**.

By promoting a common set of terms for RI entities, ENVRI RM helps normalise and streamline discourse between different RI initiatives.

Open Information Linking

Open Information Linking for Environmental science research infrastructures (OIL-E) (<http://www.oil-e.net/>) is intended to provide a **framework for semantic linking** between different RI standards and vocabularies.

- OIL-E uses the archetypes of ENVRI RM to produce an **upper ontology** for RI specifications.
- OIL-E provides a **linking model** for describing the overlaps between the different metadata schemes used by RIs to describe their resources.
- OIL-E provides links to **semantic mappings** that used to convert between schemes.

Using OIL-E, the ENVRIplus project will build a **knowledge base** describing:

- The **semantic landscape** of environmental science RIs in Europe.
- Information about *metadata schemes, ontologies, thesauri* and other controlled vocabularies used by RIs.
- How to navigate the semantic bottlenecks facing the establishment of an open science commons in Europe and beyond.

The **ENVRI Reference Model** describes environmental science research infrastructures based on input from RI communities



Requirements

Documentation

Current RI technologies

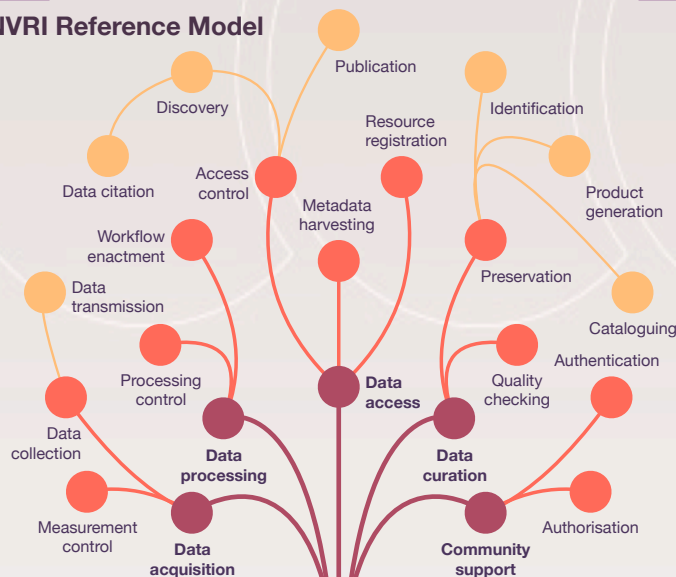
Community standards
(ISO 19115/19139, CERIF, EnvThes...)

Resource catalogues
(RI metadata for datasets, resources, sites...)

Semantic Web technologies
(RDF, OWL 2, SPARQL...)

Open Information Linking for Environmental science research infrastructures is intended to provide a framework for linking the different standards, vocabularies and metadata schemes used by communities

ENVRI Reference Model



The **ENVRI Reference Model** defines standard archetypes (actors, services, behaviours, etc.) and a standard vocabulary for describing RI architecture

Multi-view ODP model
(<http://envri.eu/rm>)

Specialising RI design for different stakeholders:
scientists, data curators, software engineers, etc.

Common vocabulary

A standard nomenclature for the actors, services,
behaviours, information objects and other RI
archetypes.

Research data lifecycle

Navigating the evolution of research data through
acquisition, curation, publishing, processing and use.

OIL-E ontologies
(<http://www.oil-e.net/ontology/>)

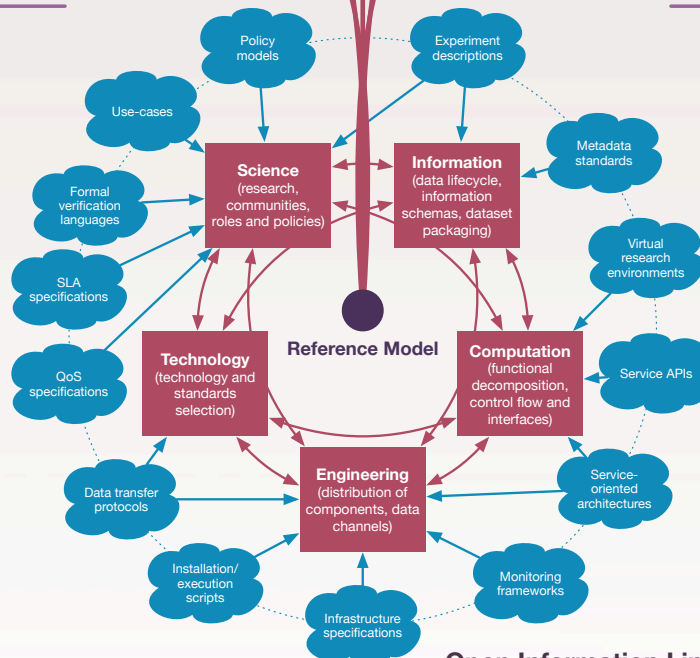
An upper ontology for RI design and
architecture, serving as a basis for knowledge.

**ENVRIplus
Knowledge Base**

Providing a map of environmental RIs and their
technologies, and supporting comparison and
analysis of RI designs.

Semantic landscape

Interlinking RIs' semantic contexts; the
standards, vocabularies and metadata schemes
used and the mappings available between them.



**Open Information Linking
for Environmental science RIs**

Semantic linking in ENVRIplus

Open Information Linking for Environmental science RIs uses the Reference Model as an upper ontology for linking different standards, building a semantic landscape for European environmental research

The ENVRI Reference Model describes environmental science research infrastructures based on input from RI communities



Requirements

Documentation

Current RI technologies

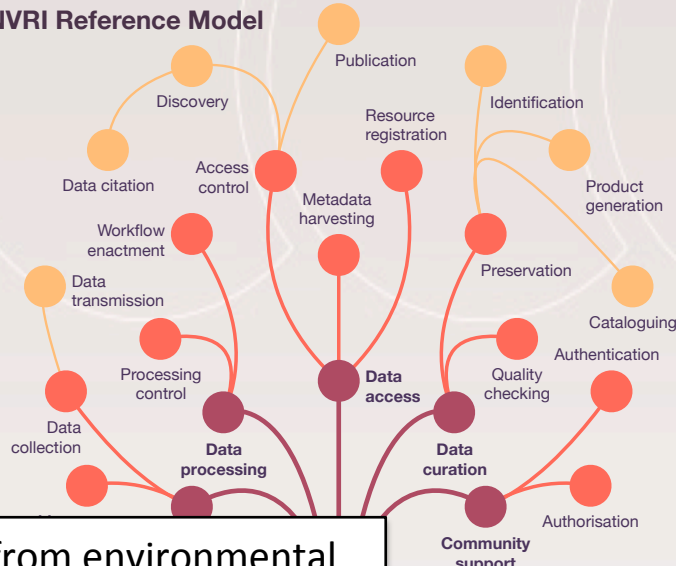
We collected information from environmental RIs and communities: requirements, technologies and the current state of the art.

Resource catalogues
(RI metadata for datasets, resources, sites...)

Semantic Web technologies
(RDF, OWL 2, SPARQL...)

Open Information Linking for Environmental science research infrastructures is intended to provide a framework for linking the different standards, vocabularies and metadata schemes used by communities

ENVRI Reference Model



The ENVRI Reference Model defines standard archetypes (actors, services, behaviours, etc.) and a standard vocabulary for describing RI architecture

Multi-view ODP model
(<http://envri.eu/rm>)

Specialising RI design for different stakeholders: scientists, data curators, software engineers, etc.

Common vocabulary

A standard nomenclature for the actors, services, behaviours, information objects and other RI archetypes.

Research data lifecycle

Navigating the evolution of research data through acquisition, curation, publishing, processing and use.

OIL-E ontologies
(<http://www.oil-e.net/ontology/>)

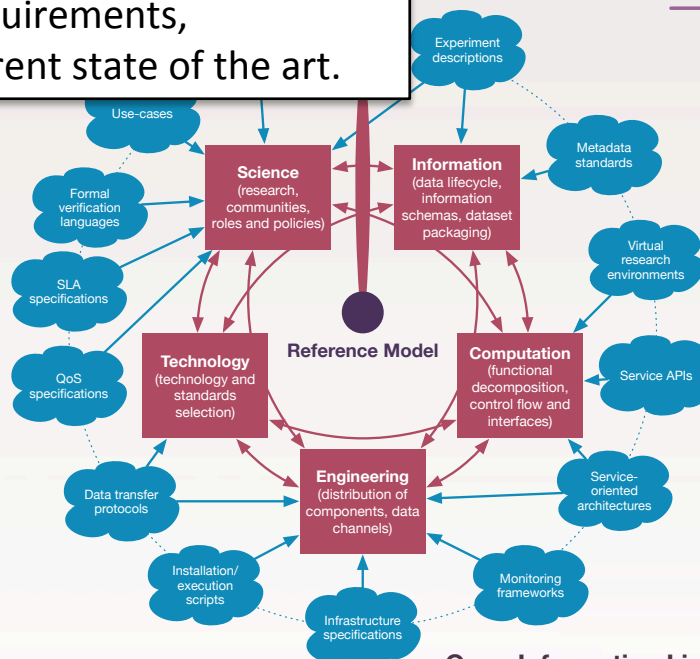
An upper ontology for RI design and architecture, serving as a basis for knowledge.

ENVRIplus Knowledge Base

Providing a map of environmental RIs and their technologies, and supporting comparison and analysis of RI designs.

Semantic landscape

Interlinking RIs' semantic contexts; the standards, vocabularies and metadata schemes used and the mappings available between them.



Open Information Linking
for Environmental science RIs

Semantic linking in ENVRIplus

Open Information Linking for Environmental science RIs uses the Reference Model as an upper ontology for linking different standards, building a semantic landscape for European environmental research

The ENVRI Reference Model describes environmental science research infrastructures based on input from RI communities

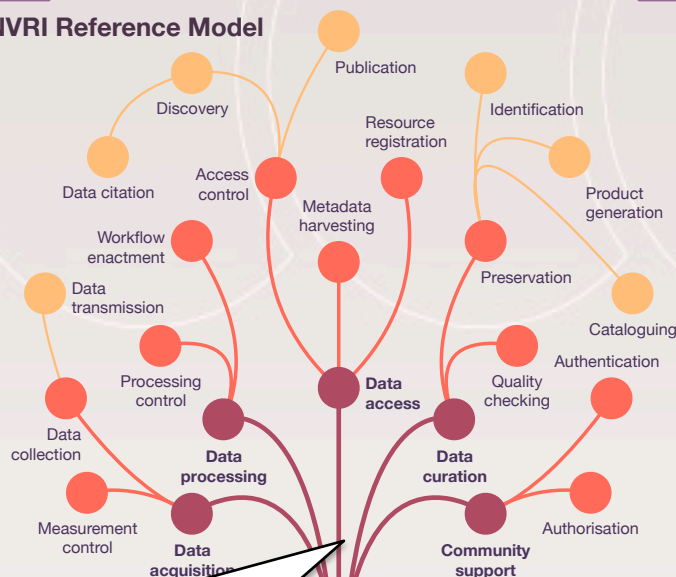


Requirements

Documentation

Current RI technologies

ENVRI Reference Model



The ENVRI Reference Model defines standard archetypes (actors, services, behaviours, etc.) and a standard vocabulary for describing RI architecture

Multi-view ODP model
(<http://envri.eu/rm>)

Specialising RI design for different stakeholders: scientists, data curators, software engineers, etc.

Common vocabulary

A standard nomenclature for the actors, services, behaviours, information objects and other RI archetypes.

Research data lifecycle

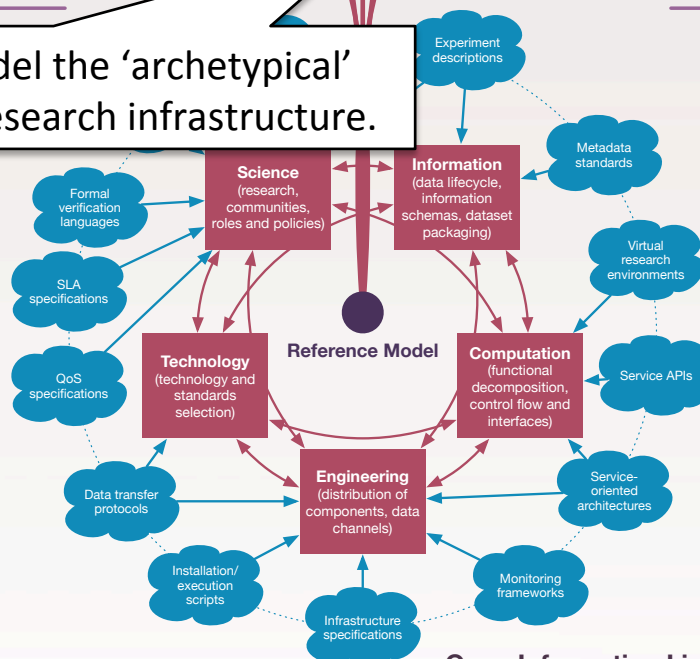
Navigating the evolution of research data through acquisition, curation, publishing, processing and use.

We used RM-ODP to model the 'archetypical' environmental science research infrastructure.

Resource catalogues
(RI metadata for datasets, resources, sites...)

Semantic Web technologies
(RDF, OWL 2, SPARQL...)

Open Information Linking for Environmental science research infrastructures is intended to provide a framework for linking the different standards, vocabularies and metadata schemes used by communities



Open Information Linking
for Environmental science RIs

OIL-E ontologies
(<http://www.oil-e.net/ontology/>)

An upper ontology for RI design and architecture, serving as a basis for knowledge.

ENVRIplus
Knowledge Base

Providing a map of environmental RIs and their technologies, and supporting comparison and analysis of RI designs.

Semantic landscape

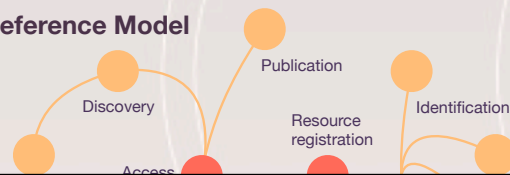
Interlinking RIs' semantic contexts; the standards, vocabularies and metadata schemes used and the mappings available between them.

Semantic linking in ENVRIplus

Open Information Linking for Environmental science RIs uses the Reference Model as an upper ontology for linking different standards, building a semantic landscape for European environmental research

The **ENVRI Reference Model** describes environmental science research infrastructures based on input from RI communities

ENVRI Reference Model

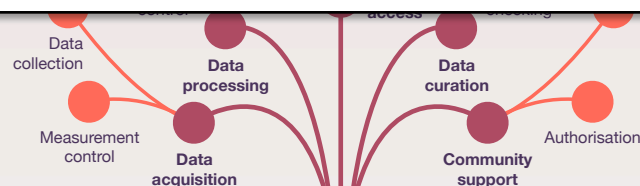


The **ENVRI Reference Model** defines standard archetypes (actors, services, behaviours, etc.) and a standard vocabulary for describing RI architecture



This resulted in a standard reference model (**ENVRI RM**) with a common vocabulary for describing various kinds of component and activity used in RIs.

Requ
Docu
Cul
technologies



Multi-view ODP model
(<http://envri.eu/rm>)

Specialising RI design for different stakeholders: scientists, data curators, software engineers, etc.

Common vocabulary

A standard nomenclature for the actors, services, behaviours, information objects and other RI archetypes.

Research data lifecycle

Navigating the evolution of research data through acquisition, curation, publishing, processing and use.

OIL-E ontologies
(<http://www.oil-e.net/ontology/>)

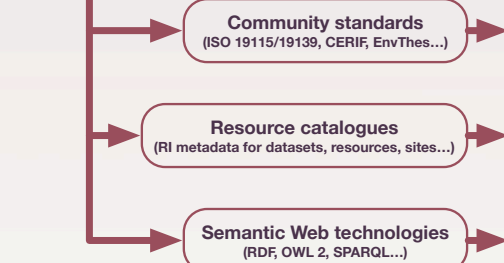
An upper ontology for RI design and architecture, serving as a basis for knowledge.

ENVRIplus Knowledge Base

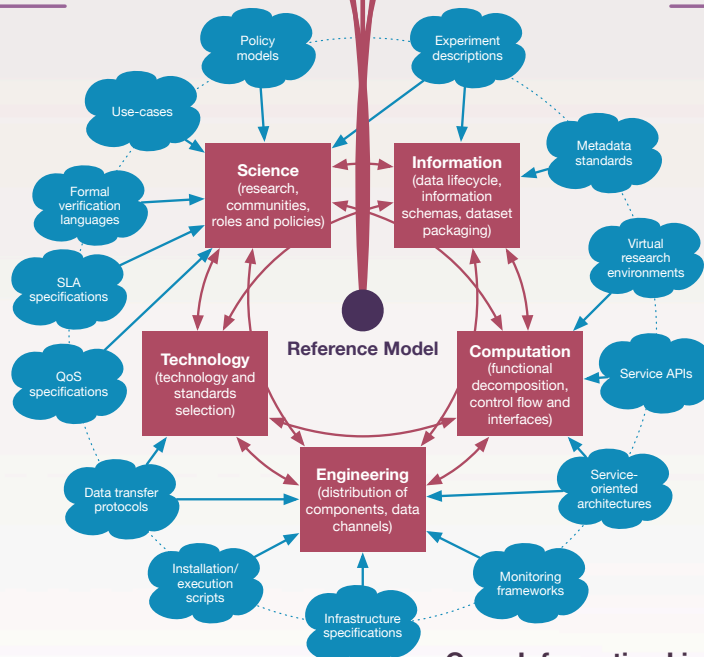
Providing a map of environmental RIs and their technologies, and supporting comparison and analysis of RI designs.

Semantic landscape

Interlinking RIs' semantic contexts; the standards, vocabularies and metadata schemes used and the mappings available between them.



Open Information Linking for Environmental science research infrastructures is intended to provide a framework for linking the different standards, vocabularies and metadata schemes used by communities



Open Information Linking
for Environmental science RIs

Semantic linking in ENVRIplus

Open Information Linking for Environmental science RIs uses the Reference Model as an upper ontology for linking different standards, building a semantic landscape for European environmental research

The **ENVRI Reference Model** describes environmental science research infrastructures based on input from RI communities

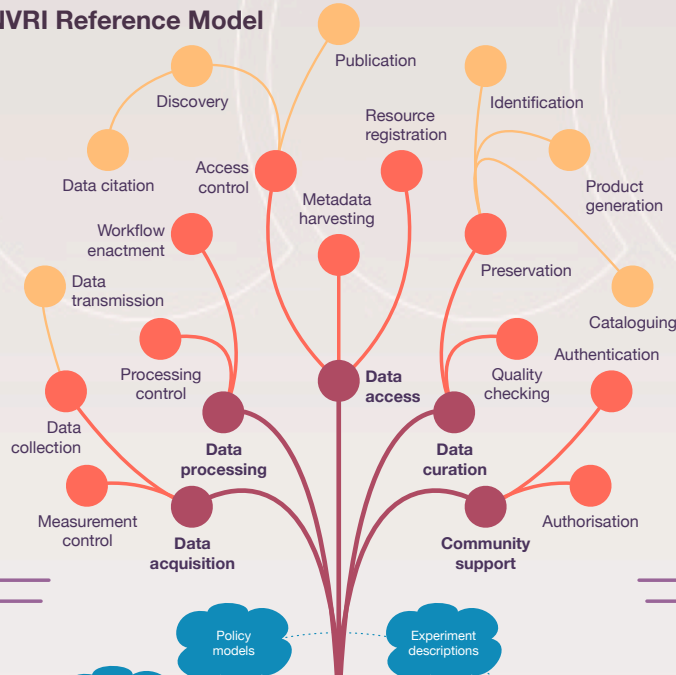


Requirements

Documentation

Current RI technologies

ENVRI Reference Model



The **ENVRI Reference Model** defines standard archetypes (actors, services, behaviours, etc.) and a standard vocabulary for describing RI architecture

Multi-view ODP model
(<http://envri.eu/rm>)

Specialising RI design for different stakeholders: scientists, data curators, software engineers, etc.

Common vocabulary

A standard nomenclature for the actors, services, behaviours, information objects and other RI archetypes.

Research data lifecycle

Navigating the evolution of research data through acquisition, curation, publishing, processing and use.

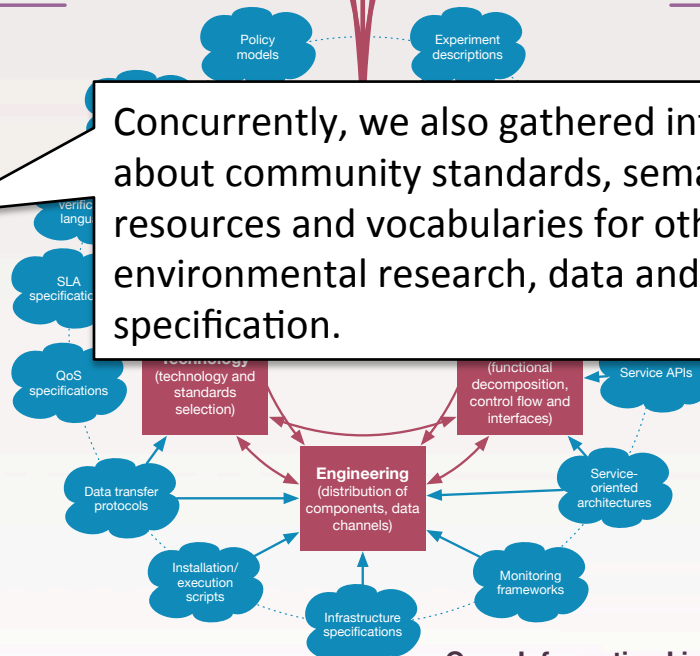
Community standards
(ISO 19115/19139, CERIF, EnvThes...)

Resource catalogues
(RI metadata for datasets, resources, sites...)

Semantic Web technologies
(RDF, OWL 2, SPARQL...)

Open Information Linking for Environmental science research infrastructures is intended to provide a framework for linking the different standards, vocabularies and metadata schemes used by communities

Concurrently, we also gathered information about community standards, semantic resources and vocabularies for other aspects of environmental research, data and process specification.



Open Information Linking
for Environmental science RIs

Ontologies
(ontology/)

An upper ontology for RI design and architecture, serving as a basis for knowledge.

Base

Providing a map of environmental RIs and their technologies, and supporting comparison and analysis of RI designs.

Semantic landscape

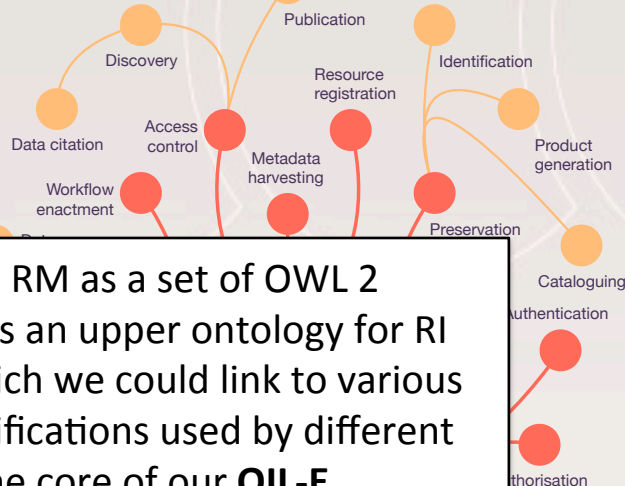
Interlinking RIs' semantic contexts; the standards, vocabularies and metadata schemes used and the mappings available between them.

Semantic linking in ENVRIplus

Open Information Linking for Environmental science RIs uses the Reference Model as an upper ontology for linking different standards, building a semantic landscape for European environmental research

The ENVRI Reference Model describes environmental science research infrastructures based on input from RI communities

ENVRI Reference Model



We captured ENVRI RM as a set of OWL 2 ontologies, to use as an upper ontology for RI architecture via which we could link to various standards and specifications used by different RI entities. This is the core of our **OIL-E** framework.

Research Infrastructures (RIs) and communities

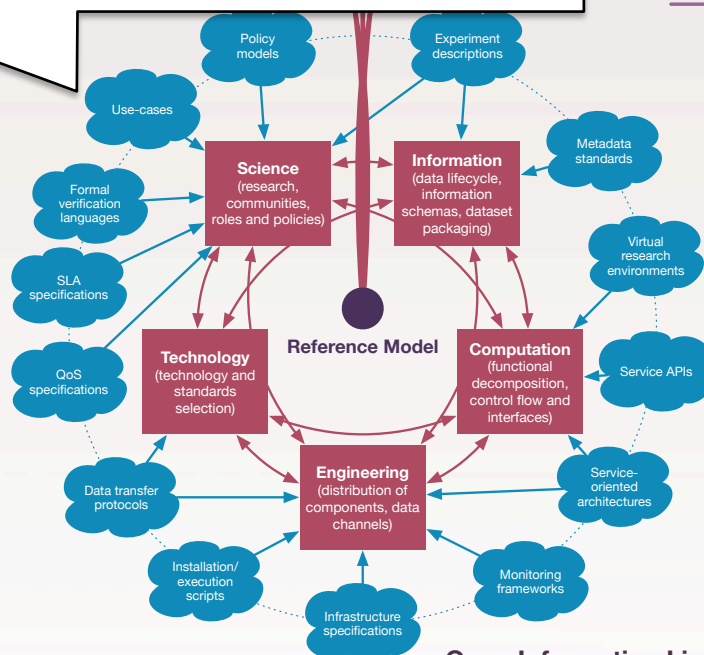
Requirements

Community standards
(ISO 19115/19139, CERIF, EnvThes...)

Resource catalogues
(RI metadata for datasets, resources, sites...)

Semantic Web technologies
(RDF, OWL 2, SPARQL...)

Open Information Linking for Environmental science research infrastructures is intended to provide a framework for linking the different standards, vocabularies and metadata schemes used by communities



Open Information Linking
for Environmental science RIs

The ENVRI Reference Model defines standard archetypes (actors, services, behaviours, etc.) and a standard vocabulary for describing RI architecture

Multi-view ODP model
(<http://envri.eu/rm>)

Specialising RI design for different stakeholders:
scientists, data curators, software engineers, etc.

Common
vocabulary

A standard nomenclature for the actors, services,
behaviours, information objects and other RI
archetypes.

Research data
lifecycle

Navigating the evolution of research data through
acquisition, curation, publishing, processing and use.

OIL-E ontologies
(<http://www.oil-e.net/ontology/>)

An upper ontology for RI design and
architecture, serving as a basis for knowledge.

ENVRIplus
Knowledge Base

Providing a map of environmental RIs and their
technologies, and supporting comparison and
analysis of RI designs.

Semantic landscape

Interlinking RIs' semantic contexts; the
standards, vocabularies and metadata schemes
used and the mappings available between them.

Open Information Linking for Environmental science RIs
uses the Reference Model as an upper ontology for
linking different standards, building a semantic
landscape for European environmental research

Semantic linking in ENVRIplus

The **ENVRI Reference Model** describes environmental science research infrastructures based on input from RI communities

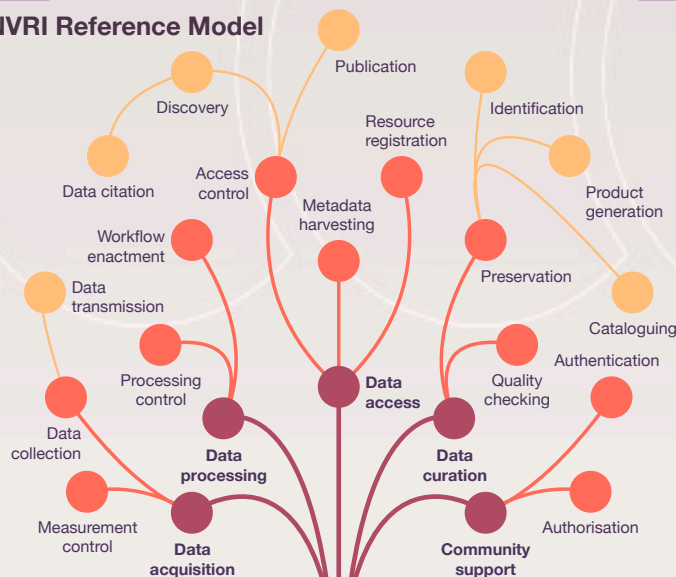


Requirements

Documentation

Current RI technologies

ENVRI Reference Model



The **ENVRI Reference Model** defines standard archetypes (actors, services, behaviours, etc.) and a standard vocabulary for describing RI architecture

Multi-view ODP model
(<http://envri.eu/rm>)

Specialising RI design for different stakeholders: scientists, data curators, software engineers, etc.

Common vocabulary

A standard nomenclature for the actors, services, behaviours, information objects and other RI archetypes.

Research data lifecycle

Navigating the evolution of research data through acquisition, curation, publishing, processing and use.

OIL-E ontologies
(<http://www.oil-e.net/ontology/>)

An upper ontology for RI design and architecture, serving as a basis for knowledge.

ENVRIplus Knowledge Base

Providing a map of environmental RIs and their technologies, and supporting comparison and analysis of RI designs.

Semantic landscape

Interlinking RIs' semantic contexts; the standards, vocabularies and metadata schemes used and the mappings available between them.

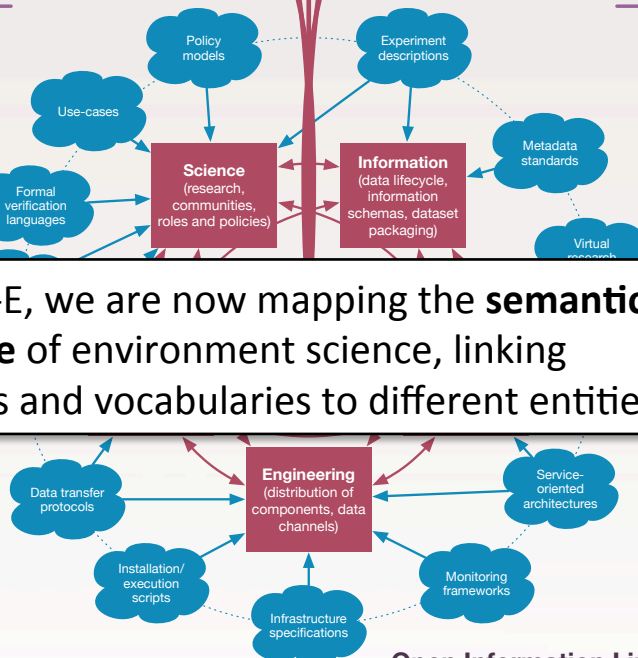
Community standards
(ISO 19115/19139, CERIF, EnvThes...)

Resource catalogues
(RI metadata for datasets, resources, sites...)

Semantic Web technologies
(RDF, OWL 2, SPARQL...)

With OIL-E, we are now mapping the **semantic landscape** of environment science, linking standards and vocabularies to different entities.

Open Information Linking for Environmental science research infrastructures is intended to provide a framework for linking the different standards, vocabularies and metadata schemes used by environmental research communities



Open Information Linking
for Environmental science RIs

Semantic linking
in ENVRIplus

Open Information Linking for Environmental science RIs uses the Reference Model as an upper ontology for linking different standards, building a semantic landscape for European environmental research

The **ENVRI Reference Model** describes environmental science research infrastructures based on input from RI communities

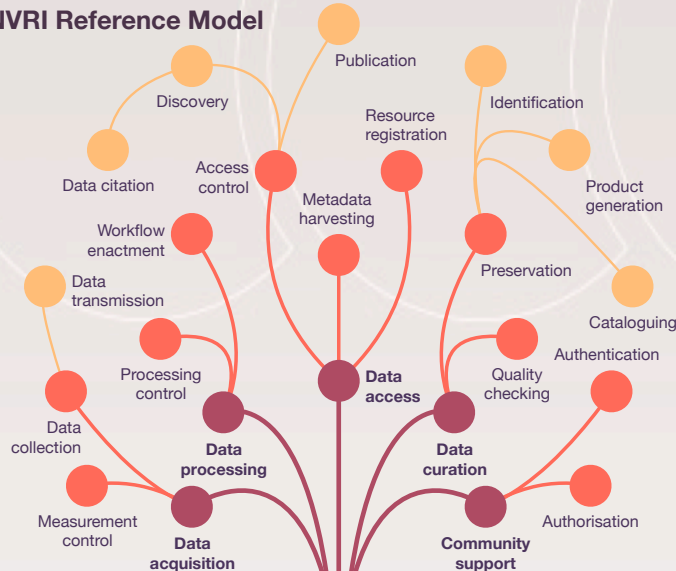


Requirements

Documentation

Current RI technologies

ENVRI Reference Model



The **ENVRI Reference Model** defines standard archetypes (actors, services, behaviours, etc.) and a standard vocabulary for describing RI architecture

Multi-view ODP model
(<http://envri.eu/rm>)

Specialising RI design for different stakeholders: scientists, data curators, software engineers, etc.

Common vocabulary

A standard nomenclature for the actors, services, behaviours, information objects and other RI archetypes.

Research data lifecycle

Navigating the evolution of research data through acquisition, curation, publishing, processing and use.

OIL-E ontologies
(<http://www.oil-e.net/ontology/>)

An upper ontology for RI design and architecture, serving as a basis for knowledge.

ENVRIplus Knowledge Base

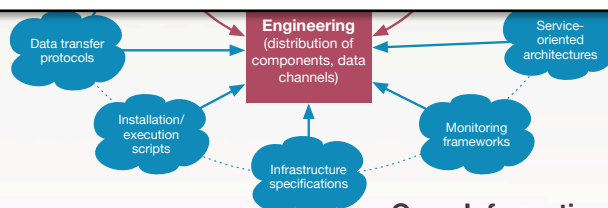
Providing a map of environmental RIs and their technologies, and supporting comparison and analysis of RI designs.

Semantic landscape

Interlinking RIs' semantic contexts; the standards, vocabularies and metadata schemes used and the mappings available between them.

In ENVRIplus we are building a **knowledge base** to contain all the OIL-E data, and provide interfaces for interacting with RI descriptions and linked external data.

Open Information Linking for Environmental science research infrastructures is intended to provide a framework for linking the different standards, vocabularies and metadata schemes used in environmental research communities



Open Information Linking
for Environmental science RIs

Semantic linking in ENVRIplus

Open Information Linking for Environmental science RIs uses the Reference Model as an upper ontology for linking different standards, building a semantic landscape for European environmental research

OIL-E ontologies

The ENVRI RM ontology provides the core of the OIL-E framework.

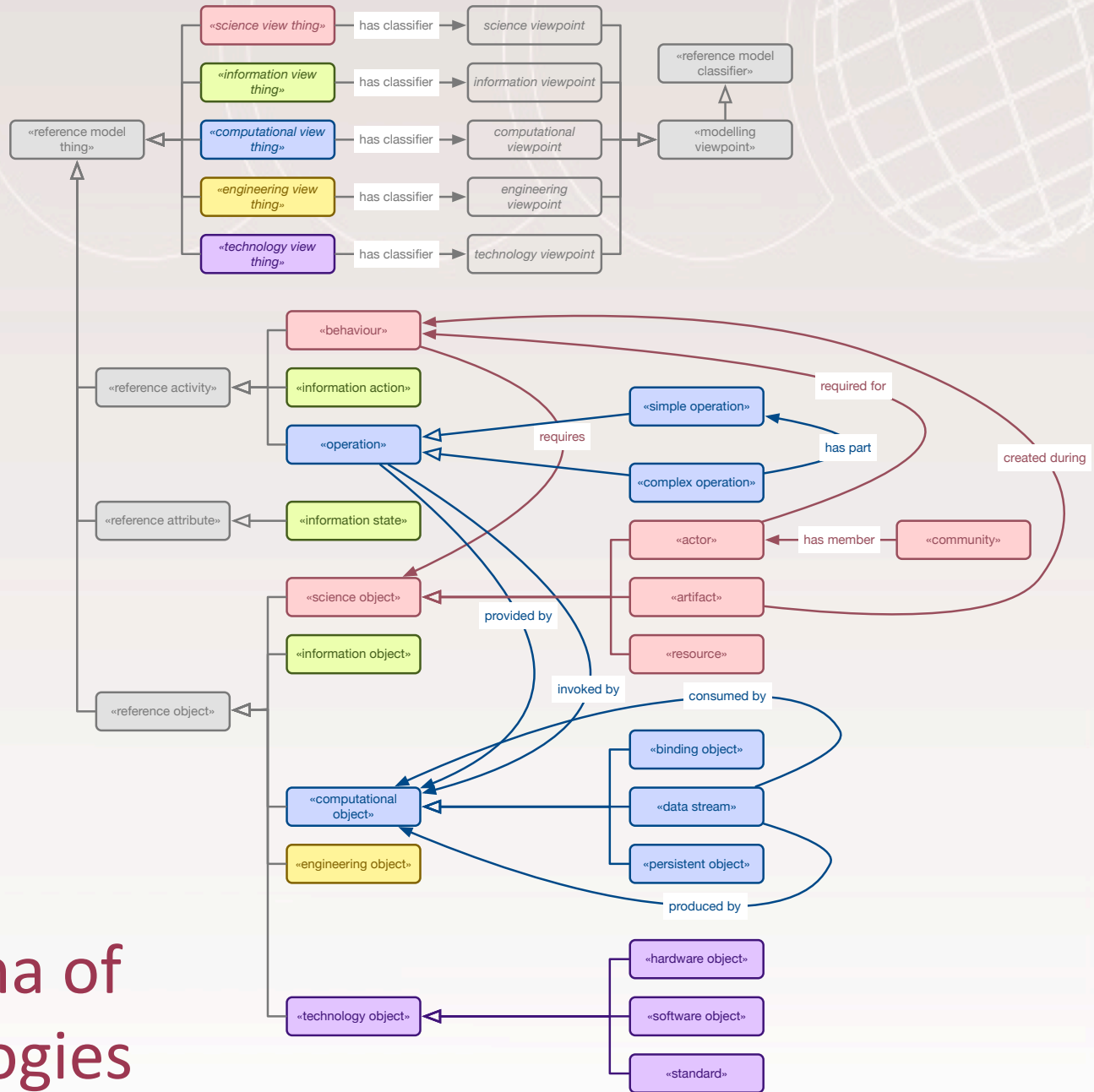
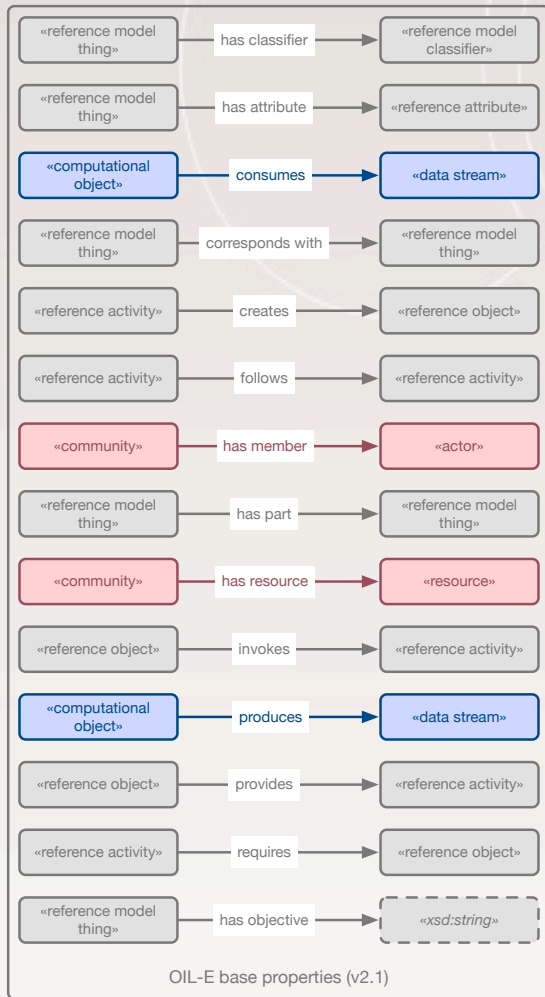
- Ontologies available at: <http://www.oil-e.net/ontology/>

The current working version of the ontology is split across three OWL 2 files:

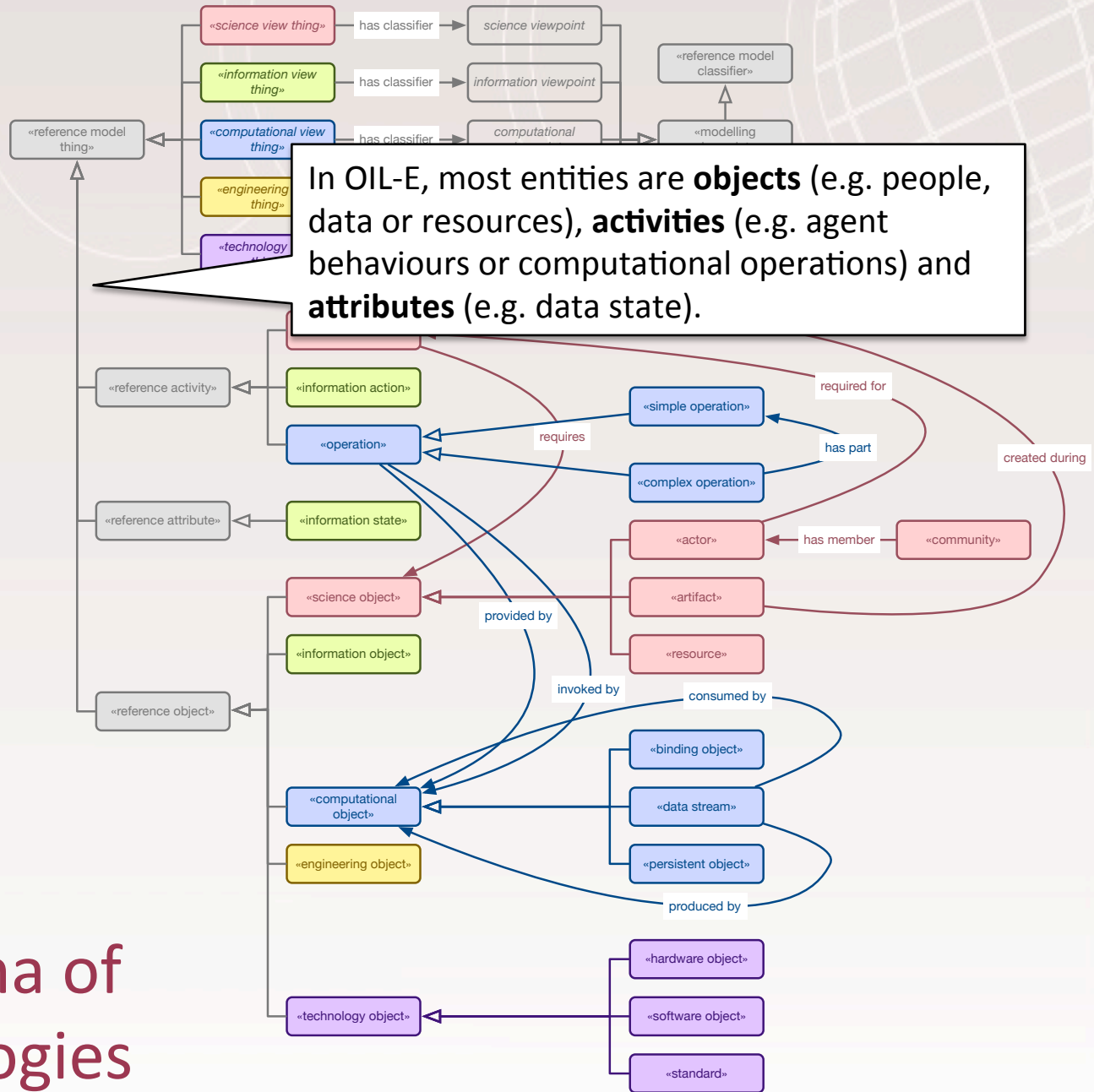
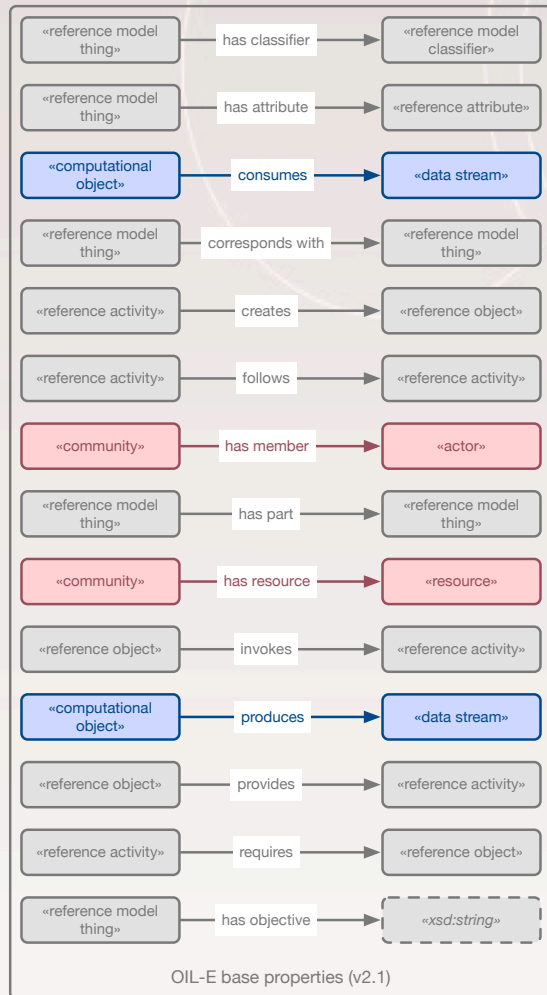
- **oil-base** (<http://www.oil-e.net/ontology/rm-core.owl>)
- **envri-rm** (<http://www.oil-e.net/ontology/rm-archetypes.owl>)
- **rm-correspondences** (<http://www.oil-e.net/ontology/rm-correspondences.owl>)

Online visualisation of **envri-rm** (using WebVOWL):

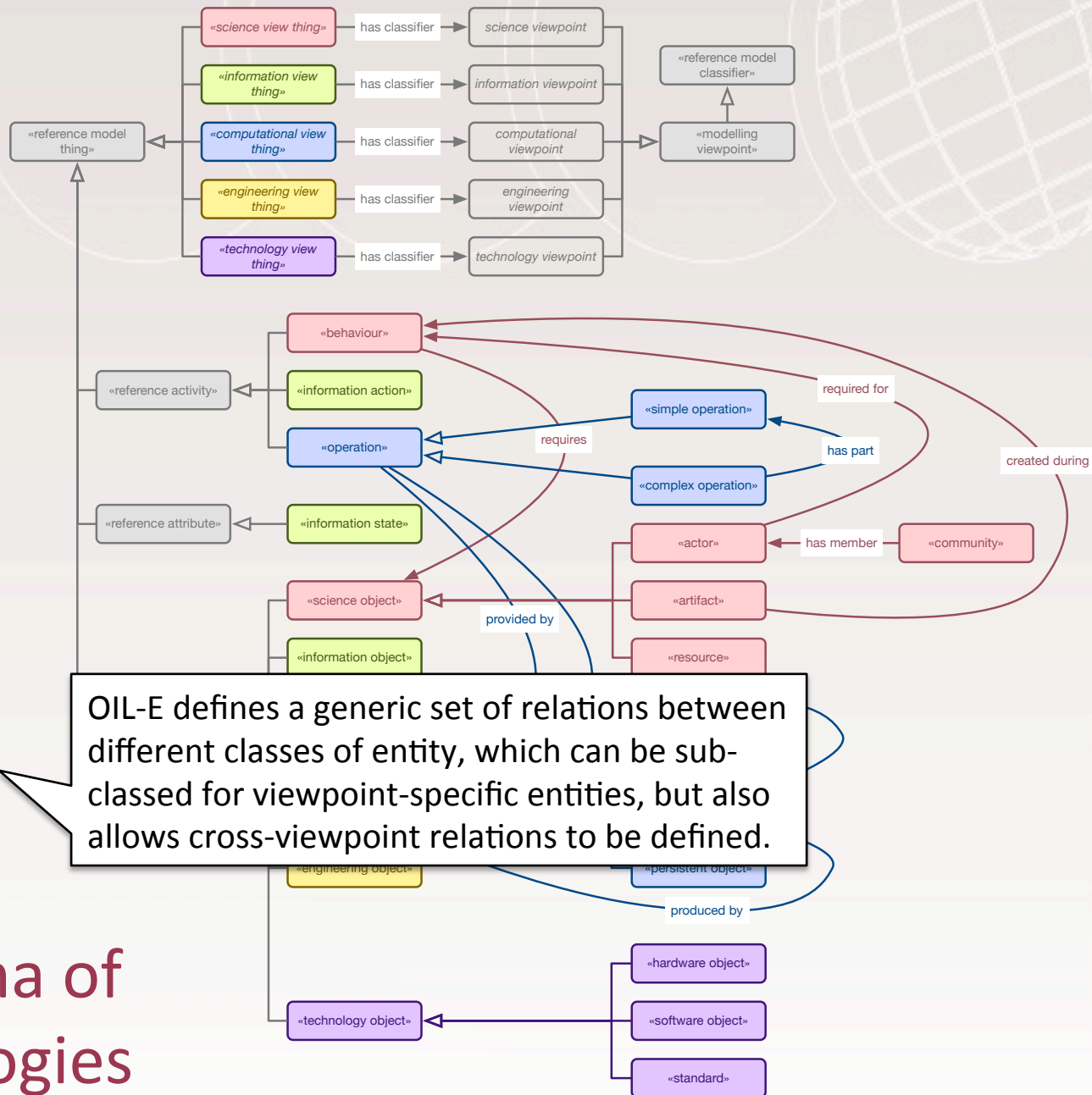
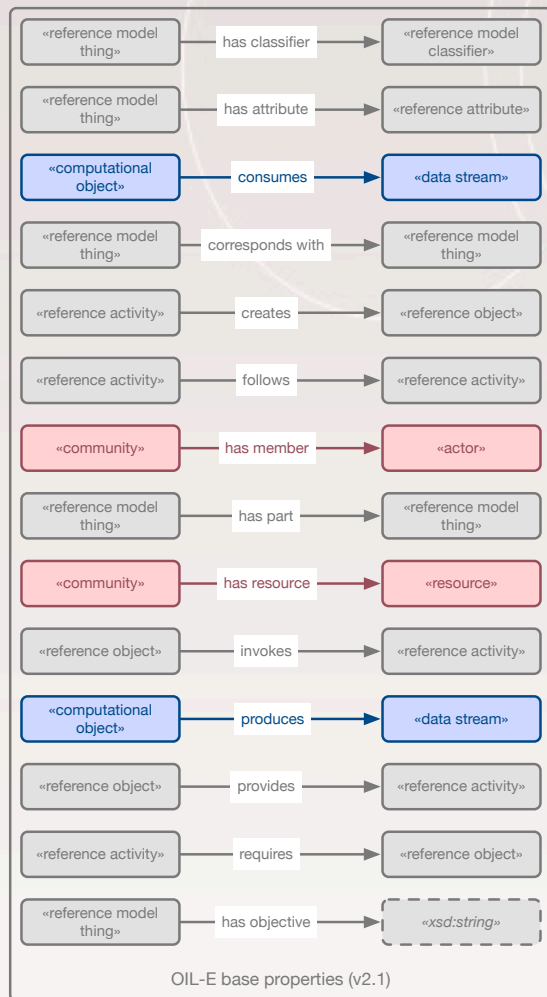
- <http://visualdataweb.de/webvowl/#iri=http://www.oil-e.net/ontology/envri-rm.owl>



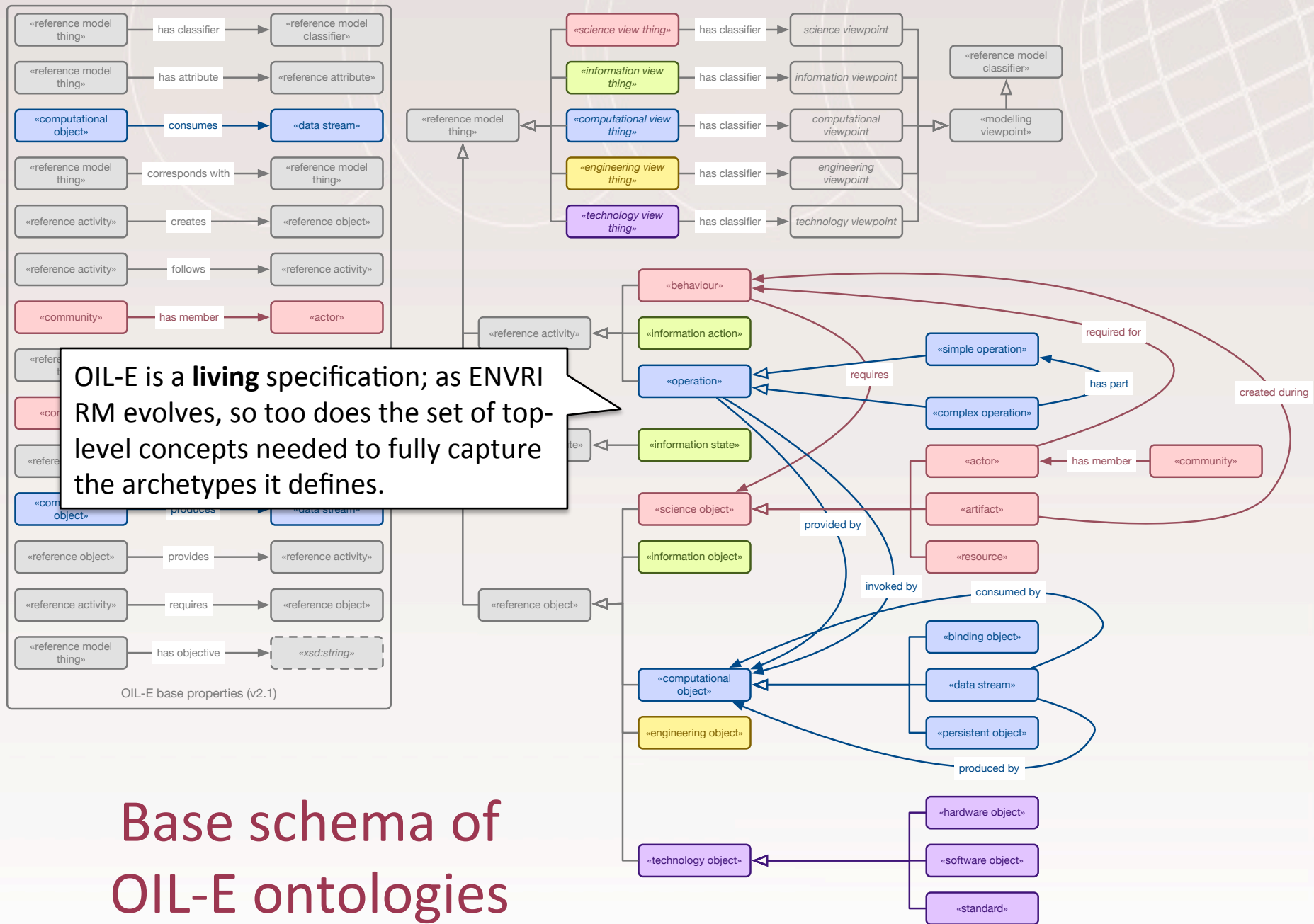
Base schema of OIL-E ontologies



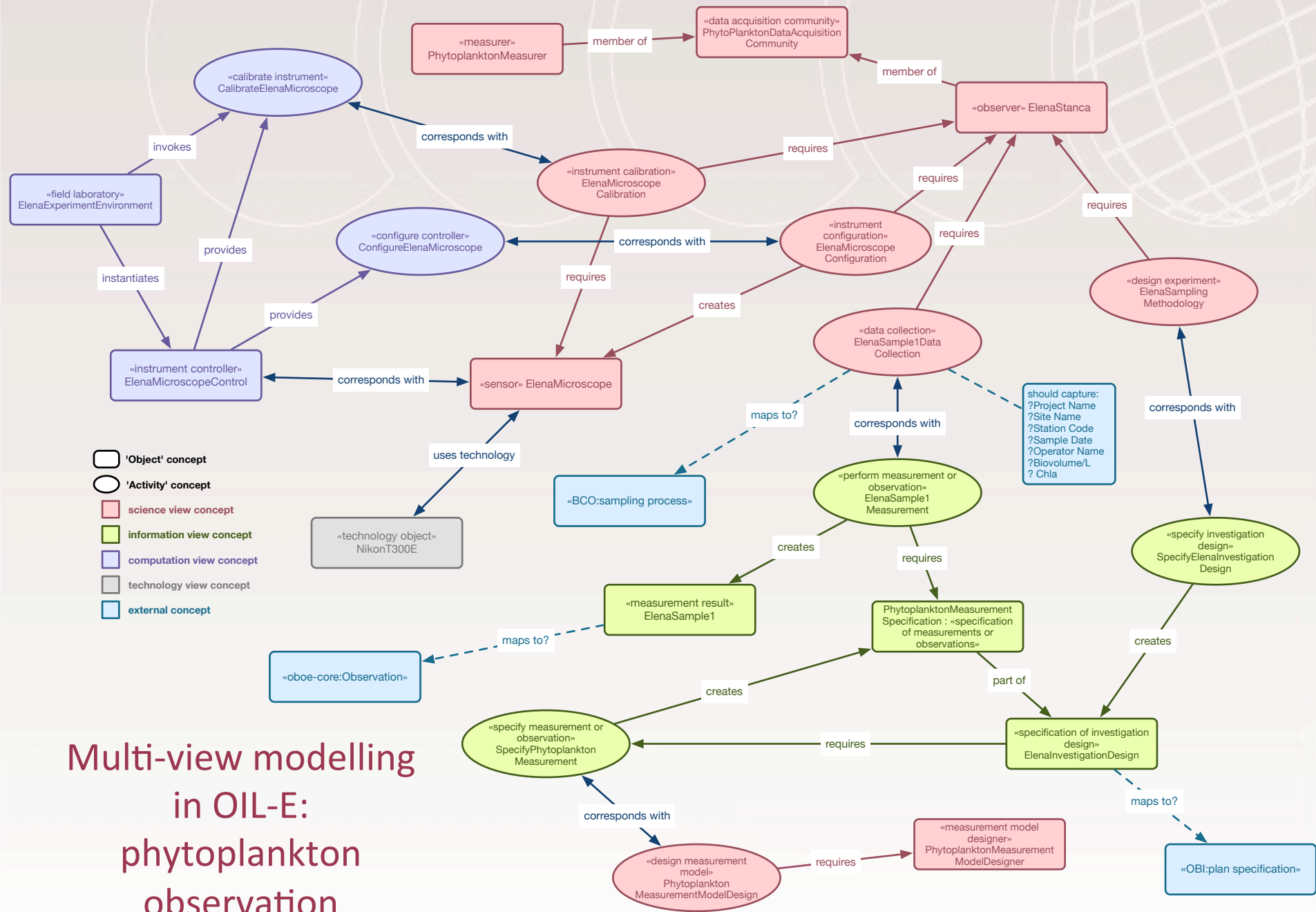
Base schema of
OIL-E ontologies



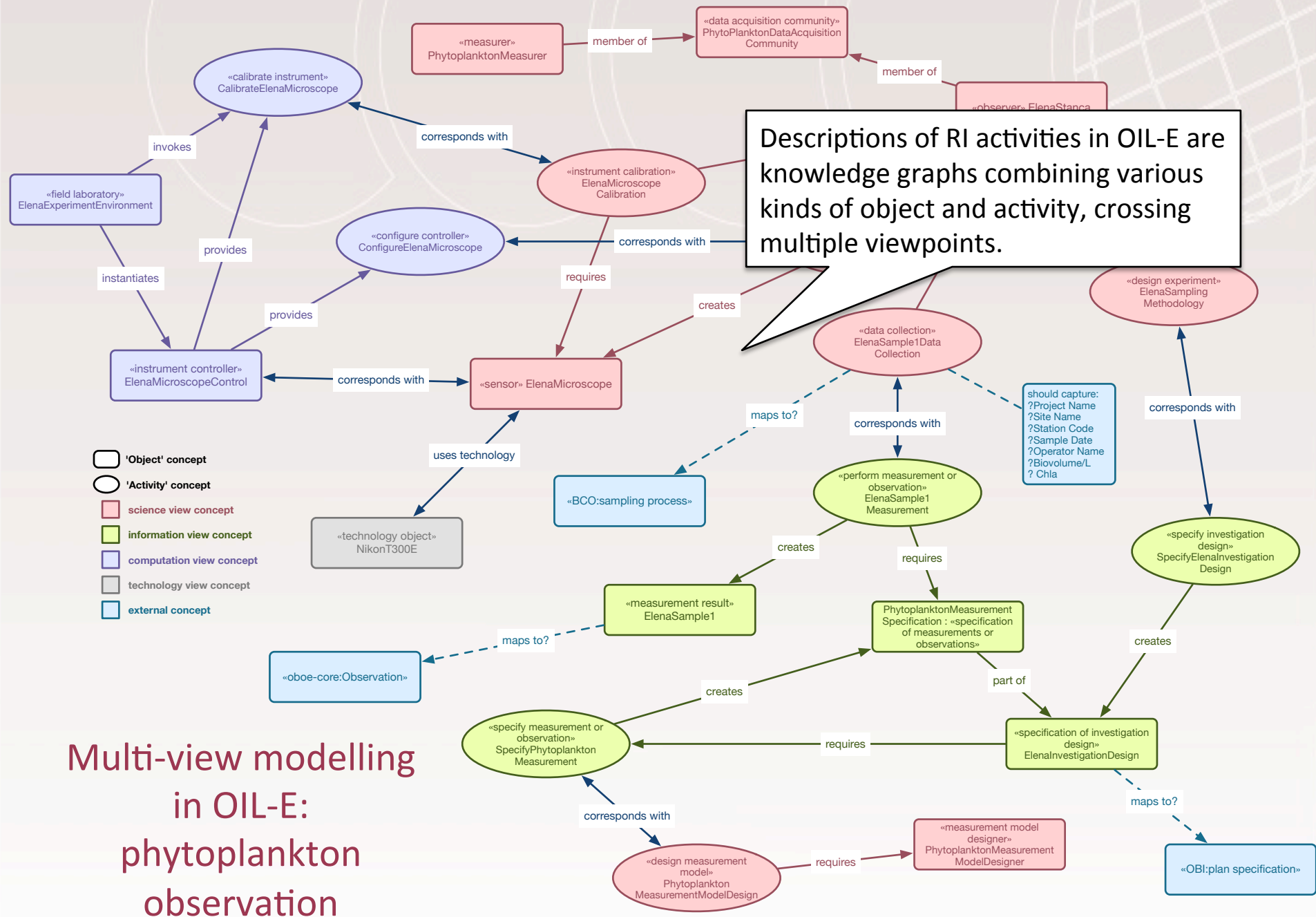
Base schema of
OIL-E ontologies



Base schema of OIL-E ontologies



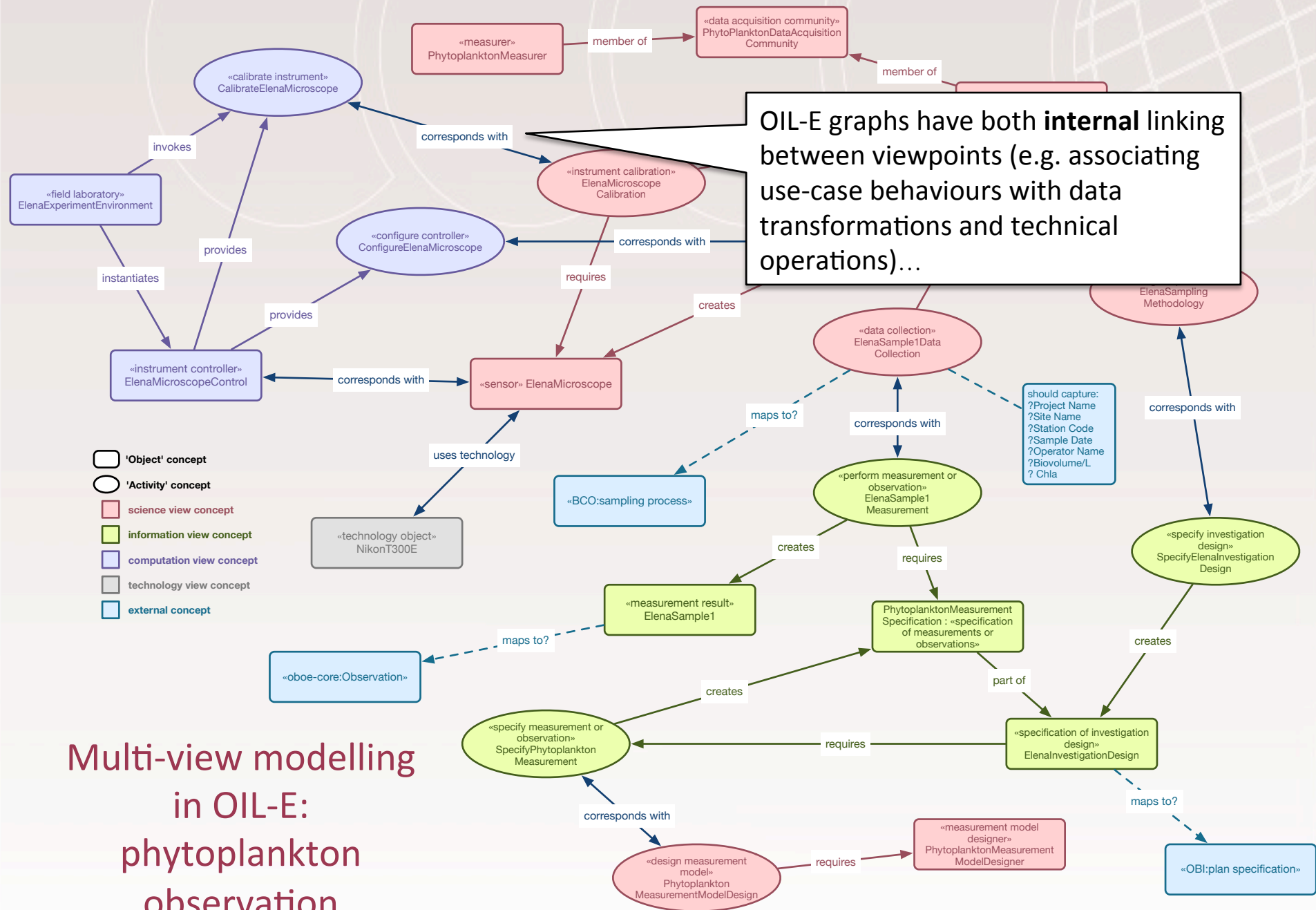
Multi-view modelling in OIL-E: phytoplankton observation

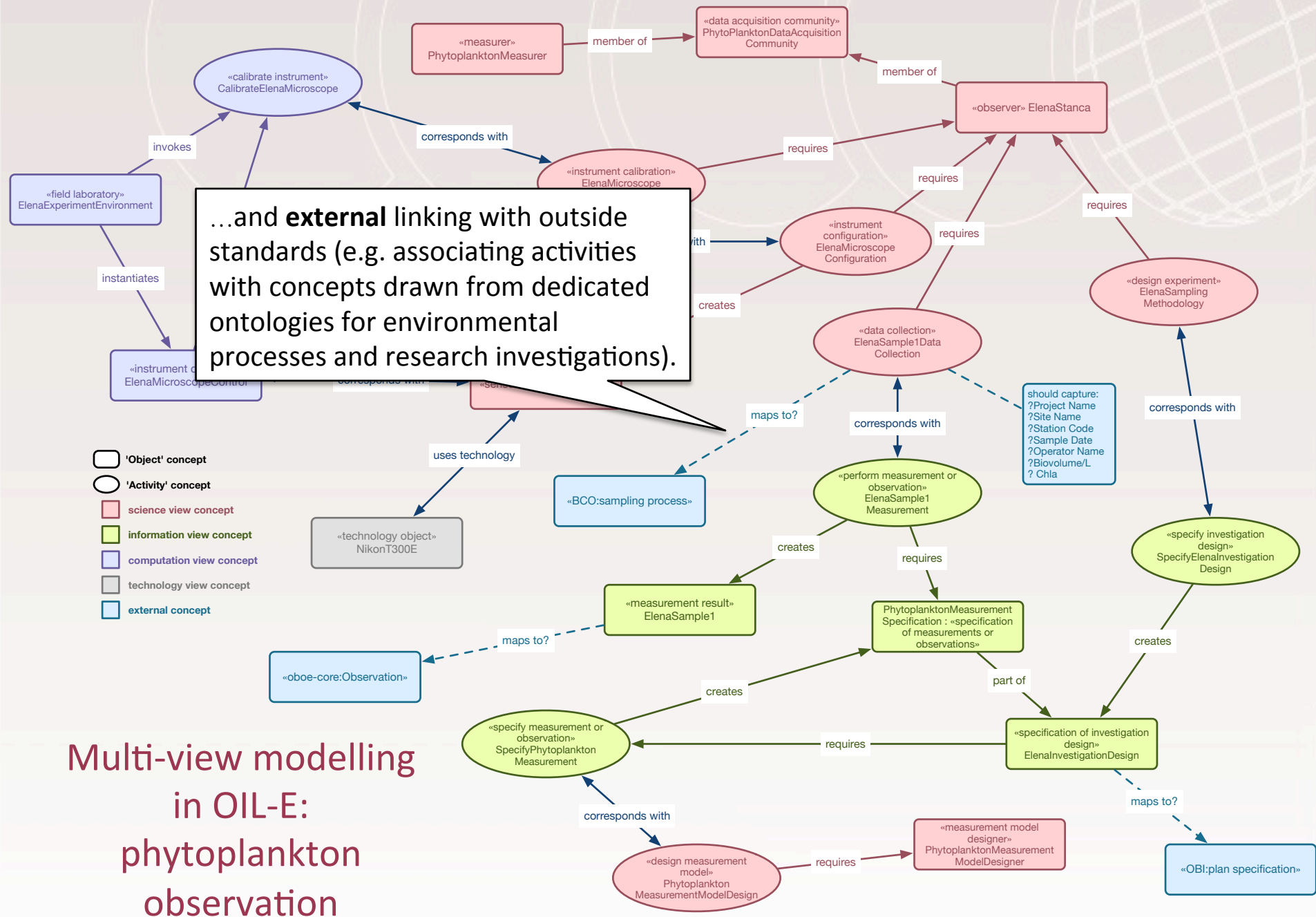


OIL-E graphs have both **internal** linking between viewpoints (e.g. associating use-case behaviours with data transformations and technical operations)...

- 'Object' concept
- 'Activity' concept
- science view concept
- information view concept
- computation view concept
- technology view concept
- external concept

Multi-view modelling in OIL-E: phytoplankton observation





Multi-view modelling in OIL-E: phytoplankton observation

Linking with external standards: OIL-E and CERIF

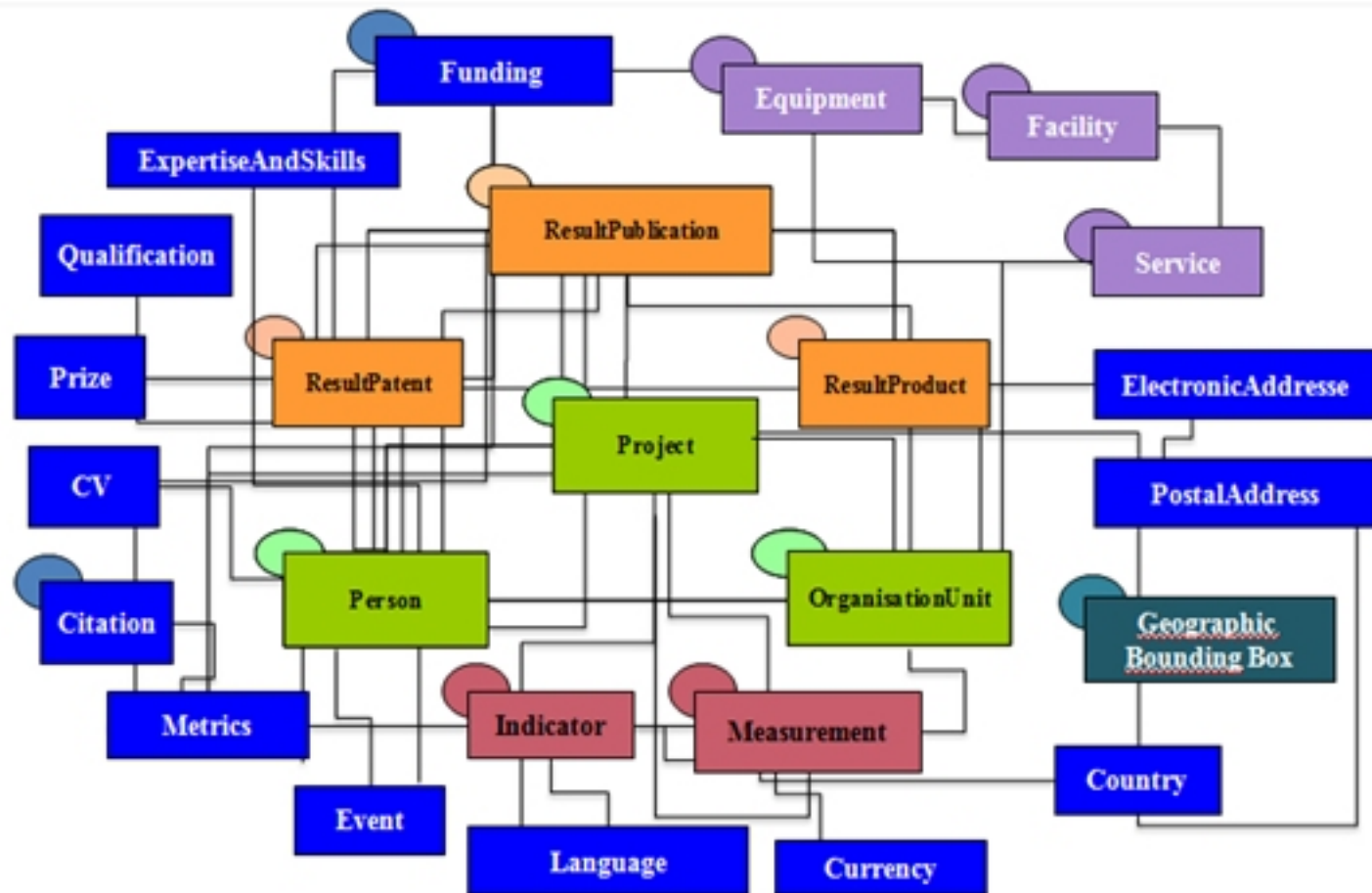
CERIF is a European-supported standard for research object representation in RDBMS form.

- CERIF also has an OWL specification, for allowing mappings of ontological data into CERIF databases (and vice versa).
- Relations between objects are modelled as objects themselves, with semantics added via classification.

The most natural way to leverage ENVRI RM in the context of CERIF is to use it as a set of classification schemes for CERIF concepts and relations.

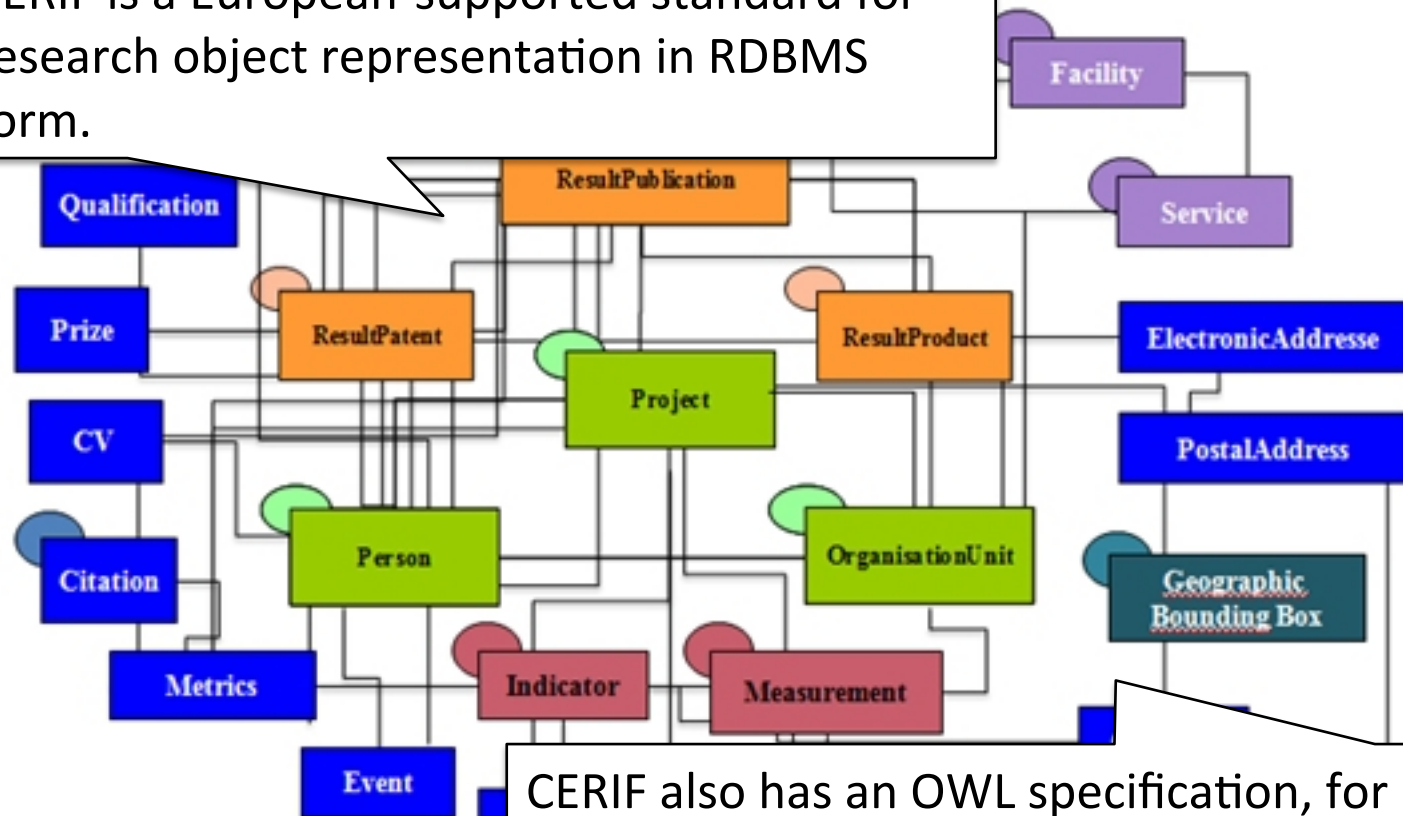
- A formal mapping has been defined in the context of the VRE4EIC project, and is being encoded in X3ML format using 3M (www.ics.forth.gr/isl/3M/).

Linking with external standards: OIL-E and CERIF



Linking with external standards: OIL-E and CERIF

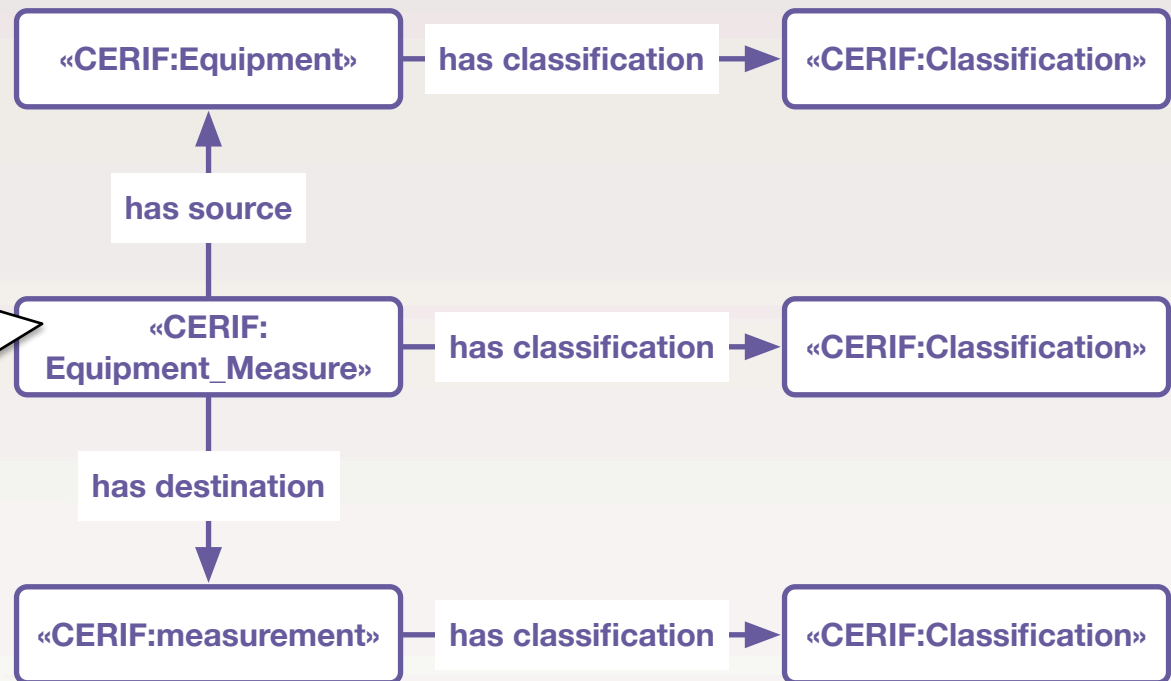
CERIF is a European-supported standard for research object representation in RDBMS form.



CERIF also has an OWL specification, for allowing mappings of ontological data into CERIF databases (and vice versa).

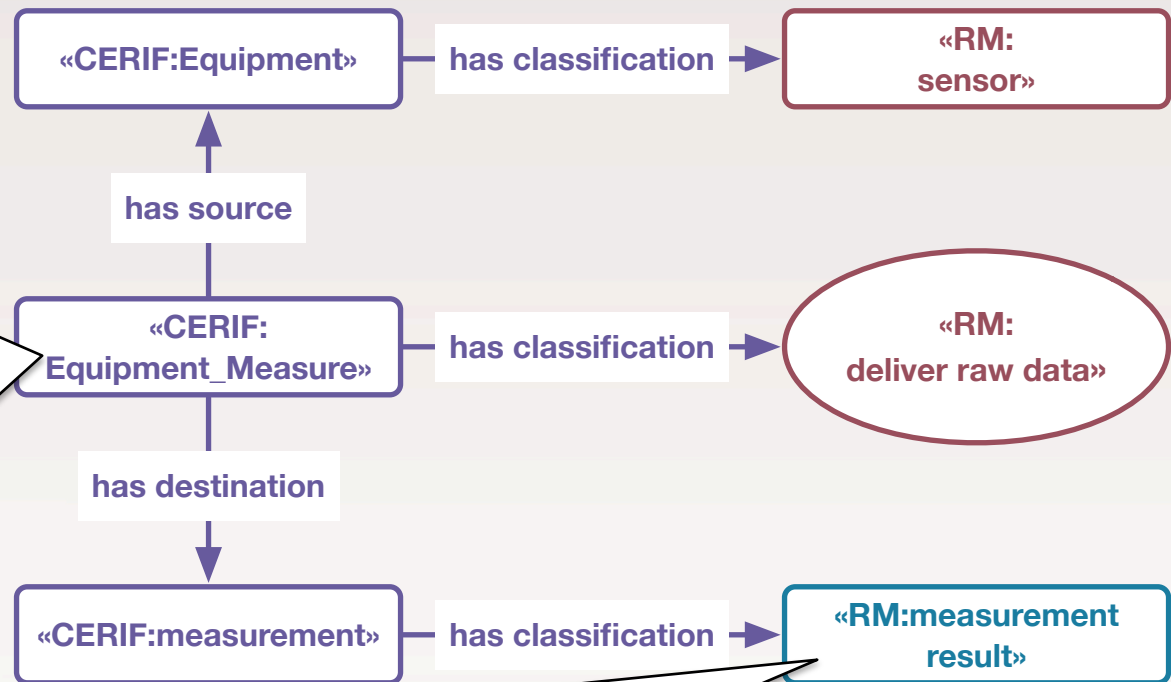
Linking with external standards: OIL-E and CERIF

Relations between objects are modelled as objects themselves, with semantics added via classification.



Linking with external standards: OIL-E and CERIF

The most natural way to leverage ENVRI RM in the context of CERIF is to use it as a set of classification schemes for CERIF concepts and relations.



A formal mapping has been defined in the context of the VRE4EIC project, and is being encoded in X3ML format using 3M (www.ics.forth.gr/isl/3M/).

ENVRI+ Knowledge Base

Current prototype based on Apache Jena Fuseki (<https://jena.apache.org/documentation/fuseki2/>).

- Contains full OIL-E ontologies.
- Currently contains a sample of RI data for experimentation and analysis.
- Provides SPARQL endpoint (<http://oil-e.vlan400.uvalight.net/rm/sparql?>).

Applies automated reasoning based on OIL-E ontologies.

- Allows clients to navigate the hierarchy of ENVRI RM archetypes to infer required characteristics and interactions.

Provides the core of a developing ecosystem of knowledge driven services as part of the ENVRIplus common services suite.

- Augmenting cataloguing, optimisation, RI design, requirements analysis, etc.

ENVRI+ Knowledge Base

Want to support a range of knowledge driven services:

- From **problems** to **processes** across different **domains**.

Need to support a range of queries from RI architects/engineers and from upstream services regarding:

- Metadata about RIs by domain and organisation.
- Services offered by RIs, the technology used and their requirements.
- What RIs offer same/similar services.
- How RIs make use of particular technologies (including standards) for their own (different?) services.

From these basic building blocks, will build a number of knowledge-driven services regarding:

- Interaction with provenance databases.
- Quality checking of annotations.
- Recommendation generation based on requirements.
- Cross-RI catalogue search.
- Interoperability analysis (gap analysis and identification of complementary solutions).
- Scenario analysis (link use-cases to RI specifications).