



ENVRI
FAIR

FAIR the smart way: Introducing the ENVRI Knowledge Base

Markus Stocker with Barbara Magagna, Doron Goldfarb, Xiaofeng Liao et al.

TIB Leibniz Information Centre for Science and Technology



ENVRI-FAIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068



Overview

- Part I: Introduction to the ENVRI KB
- Q & A: Questions on Part I
- Part II: Technical details
- Q & A: Questions on Part II

Part I

Introduction to the ENVRI KB



Motivation

- ENVRI creates, maintains and publishes substantial information



Motivation

- ENVRI creates, maintains and publishes substantial information
- There is of course all the scientific information



Motivation

- ENVRI creates, maintains and publishes substantial information
- There is of course all the scientific information
- But setting all that aside here, there is information about
 - Research infrastructures, instruments, systems, staff, etc.
 - Generated data, data collections, data products, data analysis
 - Infrastructure R&D, recommendations, best practices
 - Infrastructure assessments, in particular on FAIR-ness
 - ... just to name some entities



Motivation

- Information is buried in all kinds of resources (websites, deliverables, etc.)
- Not easy to find, not interoperable, difficult to reuse
- As *un-FAIR* as your worst nightmare



Knowledge base

- The solution: Let's build a knowledge base!
- Integrate or at least link all (or most) information sources
- Provide a single entry point, search to ENVRI-published information



Knowledge base

- Obviously not a new idea
- ENVRI-FAIR Redmine and ENVRI Wiki are knowledge bases
- ENVRI has been discussing Knowledge Base since at least ENVRIplus
- Laying the foundations for the original aims in ENVRI the first project
- Indeed knowledge bases are popular in other projects
- THOR had a Knowledge Hub (project-thor.readme.io)
- FREYA has its PID Forum (pidforum.org)
- ... just to name a few



Knowledge base

- Common to most of them is *design for humans* (experts)
- Information is encoded in natural language text, images, tables, etc.
- Information is not (re)usable for machines
- Information is not efficiently (re)usable for humans



ENVRI Knowledge Base

- ENVRI had a different vision for its Knowledge Base
- Information in the ENVRI KB should be machine processable
- The vision is certainly half a decade old
- Grounded in ENVRI RM based comprehensive descriptions of RIs
- The vision never quite materialized
- Truth be told, we are still not there
- But ENVRI-FAIR gave it another try, perhaps a more pragmatic one

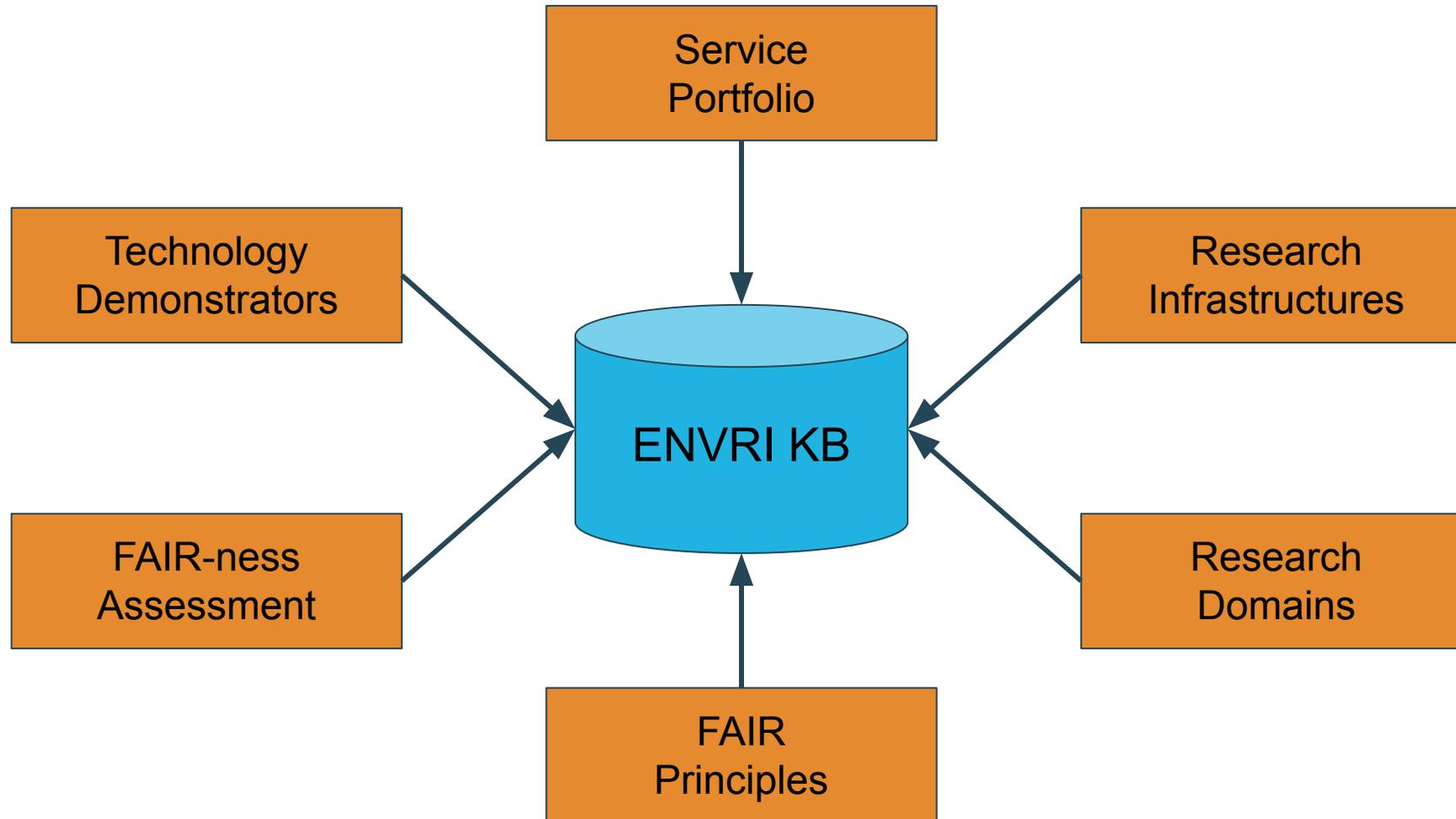


ENVRI Knowledge Base

- ENVRI KB as a tool to support RI FAIR-ification processes
- Help RI developers and managers in making RI data and services FAIR
- Understand and monitor RI FAIR-ness (where are we)
- Discover about gaps in RI FAIR-ness (what needs our attention)
- Learn about technology solutions to address gaps (what can we do)
- Along relevant dimensions, achieve technology convergence

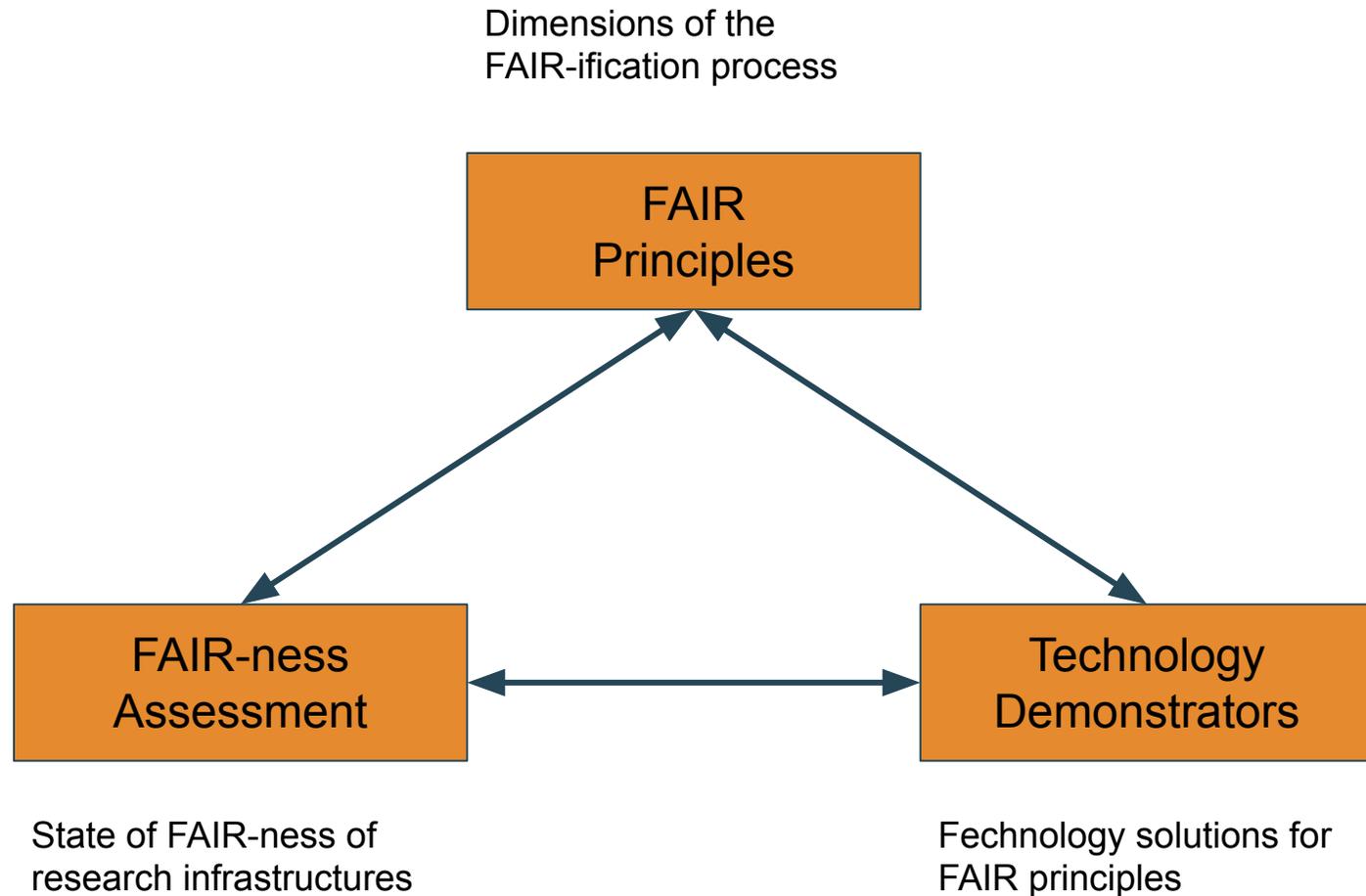


ENVRI KB: Information types





ENVRI KB: Information relations





ENVRI KB: Information integration

- All information is represented using the same technology
- All information is highly structured, machine readable
- Information is Linked Data, e.g.
 - FAIR-ness Assessments link to Research Infrastructures
 - Research Infrastructures link to Research Domains
 - FAIR-ness Assessments link to FAIR Principles
 - Technology Demonstrators link to FAIR Principles



ENVRI KB: Information access

- All information is served by a database
- Information is accessible over the Web
- Programmatically using a standard query language
- Manually using a generic user interface
- Manually using specialized user interfaces



ENVRI KB: Generic user interface

The screenshot displays a web browser window with the URL `https://envrifair1.test.fedcloud.eu/OntoWiki/index.php/resource/properties/?r=http%3A%2F%2Fenvri.eu%2Fentity%2FQ9BcgOzo4`. The page features a navigation bar with 'User', 'Extras', and 'Help' links, and a search bar labeled 'Search for Resources'. On the left, there are three panels: 'Knowledge Bases' with links to 'ENVRI Knowledge Base', 'Envri Service Portfolio with Terminology', and 'TEST'; 'Navigation: Classes' with a search box and a list of categories including 'FAIR Principle or Sub-Principle', 'research domain', 'research infrastructure', 'specification', 'FAIR assessment', 'data repository', 'Repository', 'metadata repository', 'data', and 'stations'; and a 'Login' panel with 'Local' and 'OpenID' tabs, 'Username' and 'Password' input fields, a 'Remember me' checkbox, and 'Login' and 'Register' buttons.

The main content area is titled 'Properties of Aerosols, Clouds and Trace gases Research Infrastructure'. It shows the resource URL `http://envri.eu/entity/Q9BcgOzo4` and a 'View Resource' button. Below this are tabs for 'Properties', 'History', 'Community', and 'Source'. The 'Properties' tab is active, displaying a table of properties:

<code>dc:description</code>	ACTRIS (Aerosols, Clouds and Trace gases Research Infrastructure) is a pan-European initiative consolidating actions amongst European partners producing high-quality observations of aerosols, clouds and trace gases. Different atmospheric processes are increasingly in the focus of many societal and environmental challenges, such as air quality, health, sustainability and climate change. ACTRIS aims to contribute in the resolving of such challenges by providing a platform for researchers to combine their efforts more effectively, and by providing observational data of aerosols, clouds and trace gases openly to anyone who might want to use them.
<code>rm:researchDomain</code>	atmospheric domain
<code>rm:shortName</code>	ACTRIS
<code>rdf:type</code>	research infrastructure
<code>label</code>	ACTRIS
<code>skos:prefLabel</code>	Aerosols, Clouds and Trace gases Research Infrastructure
<code>foaf:homepage</code>	http://www.actris.eu/

At the top right of the main content area, there are buttons for 'Add Property', 'Edit Properties', 'Clone', and 'Delete'. On the right side, there are two additional panels: 'Latest Comments' with an 'Enter your Comment' input field and a message 'There are no discussions yet.', and 'Instances linking here' with a link to 'infrastructure⁻¹ ACTRIS FAIR assessment'.



ENVRI KB: Generic user interface

- Aimed at expert users, e.g. RI developers
- Access to all information in the ENVRI KB
- Information can be browsed, searched as well as edited
- Flexible but generic
- No specialized information interpretation useful to ENVRI



ENVRI KB: Specialized user interface

FAIR Gap Analysis

Research Infrastructures and their repositories that do *not* meet the FAIR principles.

I2: (meta)data use vocabularies that follow FAIR principles

Demonstrator

Infrastructure	Repositories
EISCAT 3D	EISCAT Schedule; Madrigal
In-service Aircraft for a Global Observing System	IAGOS repository
Euro-Argo	Euro-Argo Data
Aerosols, Clouds and Trace gases Research Infrastructure	CLOUDNET; ACTRIS-ACCESS; EARLINET Database
European Plate Observing System	EPOS INGV; Terradue

R1.2: (meta)data are associated with detailed provenance

Demonstrator

Infrastructure	Repositories
EISCAT 3D	EISCAT Schedule
In-service Aircraft for a Global Observing System	IAGOS repository
Analysis and Experimentation on Ecosystems	ANAE- France Metadata Catalog
Svalbard Integrated Arctic Earth Observing System	Norwegian Polar Data Centre; Norwegian Meteorological Institute
Integrated Carbon Observation System	Carbon Portal
Aerosols, Clouds and Trace gases Research Infrastructure	ASC; GRES; ACTRIS-ACCESS; ACTRIS - In-Situ unit; CLOUDNET
LifeWatch	Marine Data Archive; LifeWatch Italy Portal; EUROBIS
SeaDataNet	SeaDataNet Common DATA Index (CDI); SeaDataNet Central Data Products
European Plate Observing System	Terradue; EPOS INGV; MySQL



ENVRI KB: Specialized user interface

FAIR Gap Analysis

Research Infrastructures and their re

I2: (meta)data use v

Infrastructure

[EISCAT 3D](#)

[In-service Aircraft for a Global Obs](#)

[Euro-Argo](#)

[Aerosols, Clouds and Trace gases F](#)

[European Plate Observing System](#)

R1.2: (meta)data are

Infrastructure

[EISCAT 3D](#)

[In-service Aircraft for a Global Obs](#)

[Analysis and Experimentation on E](#)

[Svalbard Integrated Arctic Earth Ob](#)

[Integrated Carbon Observation Sys](#)

[Aerosols, Clouds and Trace gases F](#)

[LifeWatch](#)

[SeaDataNet](#)

[European Plate Observing System](#)

Jupyter demonstrator (unsaved changes)

Visit repo Copy Binder link

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3

Run Code Download GitHub Binder

```

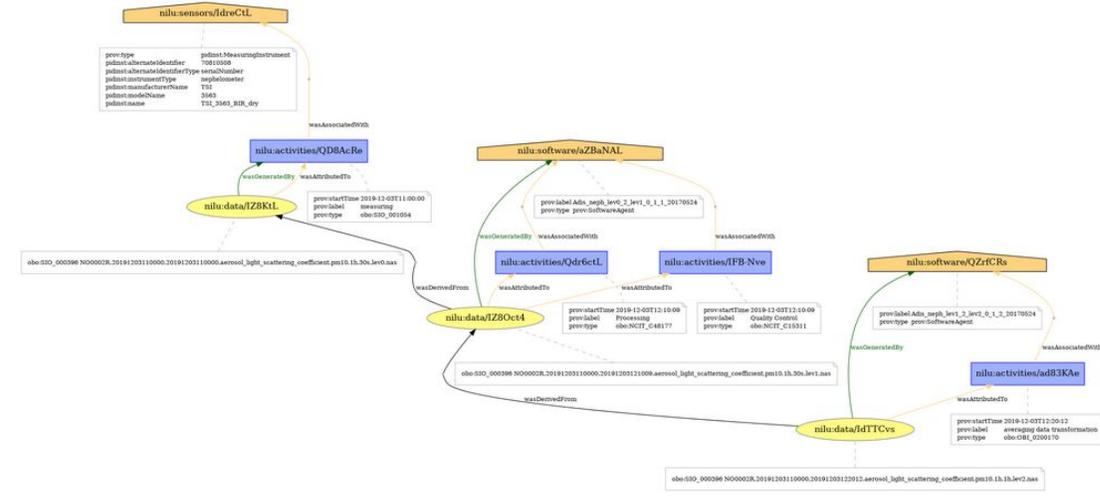
prov.wasDerivedFrom(et2, et1)
prov.wasGeneratedBy(et2, ag2)
prov.wasAttributedTo(et2, ac2)
prov.wasAttributedTo(et2, ac3)
prov.wasAssociatedWith(ac2, ag2)
prov.wasAssociatedWith(ac3, ag2)

prov.wasDerivedFrom(et3, et2)
prov.wasGeneratedBy(et3, ag3)
prov.wasAttributedTo(et3, ac4)
prov.wasAssociatedWith(ac4, ag3)

prov.serialize('data-1.ttl', format='rdf', rdf_format='ttl')

dot = prov_to_dot(prov)
display(Image(dot.create_png()))

```



Terradue; EPOS INGV; MySQL



ENVRI KB: Specialized user interfaces

- Aimed at all interested users
- Specialized access to information in the ENVRI KB
- Information is interpreted to serve ENVRI needs and requirements
- Given structured information, the possibilities here are many
- There can be many specialized user *interfaces*



Demo

<https://envri-fair.github.io/knowledge-base-ui/>



ENVRI KB: Your contribution

- Use these tools and give us feedback
- Tell us if and how they serve your needs or how they fall short
- Provide us structured data, e.g. about your RI
- Develop demonstrators for your technology solutions for FAIR principles
- Tell us about your demonstrators

Q & A

Part II

Technical details



Key ingredients

- Shared identifiers
- Common data model



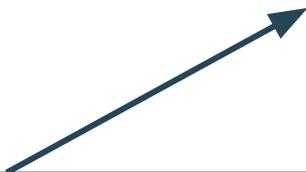
Shared identifiers

- All data are identified by URI (Uniform Resource Identifier)
 - `http://envri.eu/entity/Q9Bcg0zo4`
 - For the time being they are not resolvable!
- Enables Linked Data
 - Seamlessly crosswalk data from heterogeneous sources
 - `envri:RAwCrVJDeJ` `envri:infrastructure` `envri:Q9Bcg0zo4`

Identifies data about a FAIR assessment

An arrow points from the text box to the URI `envri:RAwCrVJDeJ` in the list above.

Identifies data about a research infrastructure

An arrow points from the text box to the URI `envri:Q9Bcg0zo4` in the list above.



Common data model

- All data are in RDF (Resource Description Framework)
- Supports integration
 - Load data from heterogeneous sources into a single storage
 - Applications with integrated view



Resource Description Framework

- Description of arbitrary resources/entities (physical, conceptual, imaginary)
- All resources are identified by URI
- Resources can be arbitrarily described by means of attributes
- It is possible to define schemas that specify these attributes
- `Statement` is the fundamental construct
- Consists of three elements: `Subject`, `Predicate`, `Object`

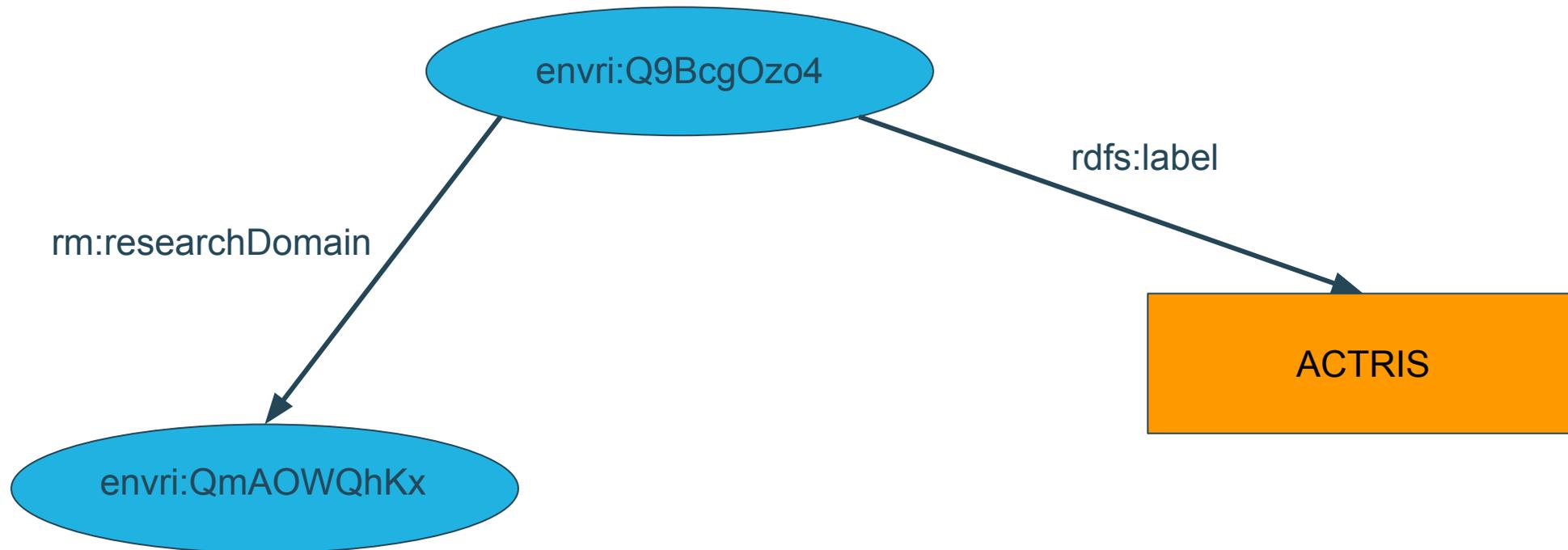


Resource Description Framework

- **Subject:** The identifier of a resource
- **Predicate:** The identifier of an attribute
- **Object:** Either the identifier of a resource or a literal value
- Graphically a statement can be represented as
 - Two nodes for the `subject` and the `object`
 - A directed edge linking the `subject` with the `object`



Resource Description Framework





Example

```
envri:Q9BcgOzo4 rdf:type rm:ResearchInfrastructure ;
```

```
foaf:homepage "http://www.actris.eu/"^^xsd:anyURI ;
```

```
skos:prefLabel "Aerosols, Clouds and Trace gases Research Infrastructure"^^xsd:string ;
```

```
rdfs:label "ACTRIS"^^xsd:string ;
```

```
rm:shortName "ACTRIS"^^xsd:string ;
```

```
rm:researchDomain envri:QmAOWQhKx ;
```

```
dc:description "ACTRIS (Aerosols, Clouds and Trace gases Research Infrastructure) is a pan-European initiative consolidating actions amongst European partners producing high-quality observations of aerosols, clouds and trace gases. Different atmospheric processes are increasingly in the focus of many societal and environmental challenges, such as air quality, health, sustainability and climate change. ACTRIS aims to contribute in the resolving of such challenges by providing a platform for researchers to combine their efforts more effectively, and by providing observational data of aerosols, clouds and trace gases openly to anyone who might want to use them."^^xsd:string .
```



Example

```
envri:Q9BcgOzo4 rdf:type rm:ResearchInfrastructure ;
```

```
foaf:homepage "http://www.actris.eu/"^^xsd:anyURI ;
```

```
skos:prefLabel "Aerosols, Clouds and Trace gases Research Infrastructure"^^xsd:string ;
```

```
rdfs:label "ACTRIS"^^xsd:string ;
```

```
rm:shortName "ACTRIS"^^xsd:string ;
```

```
rm:researchDomain envri:QmAOWQhKx ;
```

```
dc:description "ACTRIS (Aerosols, Clouds and Trace gases Research Infrastructure) is a pan-European initiative consolidating actions amongst European partners producing high-quality observations of aerosols, clouds and trace gases. Different atmospheric processes are increasingly in the focus of many societal and environmental challenges, such as air quality, health, sustainability and climate change. ACTRIS aims to contribute in the resolving of such challenges by providing a platform for researchers to combine their efforts more effectively, and by providing observational data of aerosols, clouds and trace gases openly to anyone who might want to use them."^^xsd:string .
```



Example

```
envri:Q9BcgOzo4 rdf:type rm:ResearchInfrastructure ;
```

```
foaf:homepage "http://www.actris.eu/"^^xsd:anyURI ;
```

```
skos:prefLabel "Aerosols, Clouds and Trace gases Research Infrastructure"^^xsd:string ;
```

```
rdfs:label "ACTRIS"^^xsd:string ;
```

```
rm:shortName "ACTRIS"^^xsd:string ;
```

```
rm:researchDomain envri:QmAOWQhKx ;
```

```
dc:description "ACTRIS is an international research  
initiative consolidating the efforts of  
clouds and trace gases research centres  
environmental challenges. The domain  
contribute in the resolving of such challenges by providing a platform for researchers to combine their  
efforts more effectively, and by providing observational data of aerosols, clouds and trace gases openly to  
anyone who might want to use them."^^xsd:string .
```

```
rdf:type base:ResearchDomain ;
```

```
dc:description "The domain of environmental science concerned with the atmosphere." ;
```

```
rdfs:label "atmospheric domain" .
```



Data sources

envri-fair / knowledge-base Unwatch 2 Star 0 Fork 0

<> Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights Settings

No description, website, or topics provided. Edit

Manage topics

9 commits 1 branch 0 packages 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

markusstocker added queries and documentation Latest commit 3d06c62 on 17 Feb

README.md	first commit	4 months ago
demonstrators.ttl	added	4 months ago
fair-principles.ttl	added	4 months ago
fairness-assessment.ttl	fix linking assessments and infrastructures	4 months ago
queries.rq	added queries and documentation	4 months ago
research-domains.ttl	fix linking assessments and infrastructures	4 months ago
research-infrastructures.ttl	fix linking assessments and infrastructures	4 months ago



Data storage and access

- All data can easily be loaded into a single storage system
- In this case we typically call this storage system a *triple store*
- Common index for fast data access
- Interfaces for data access, both graphically and programmatically (API)
- Programmatically via HTTP using SPARQL
- SPARQL is the de facto standard query language for RDF



SPARQL

```
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix rm: <http://www.oil-e.net/ontology/envri-rm.owl#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix dc: <http://purl.org/dc/elements/1.1/>

select ?prefLabel ?label ?domain ?homepage ?description
where {
  [] a rm:ResearchInfrastructure ;
    foaf:homepage ?homepage ;
    skos:prefLabel ?prefLabel ;
    rdfs:label ?label ;
    rm:researchDomain [ rdfs:label ?domain ] ;
    dc:description ?description .
}
```



SPARQL

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#) | [RDF views](#)

Default Data Set Name (Graph IRI)

Query Text

```

prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix rm: <http://www.oil-e.net/ontology/envri-rm.owl#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix dc: <http://purl.org/dc/elements/1.1/>

```

```

select ?label ?homepage
where {
  [] a rm:ResearchInfrastructure ;
    foaf:homepage ?homepage ;
    rdfs:label ?label .
}

```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

Execution timeout: milliseconds (values less than 1000 are ignored)

- Options:
- Strict checking of void variables
 - Log debug info at the end of output (has no effect on some queries and output formats)
 - Generate SPARQL compilation report (instead of executing the query)

(The result can only be sent back to browser, not saved on the server, see [details](#))

label	homepage
"ESONET-Vi"	"http://visobservatories.webs.com/"
"EISCAT_3D"	"https://eiscat3d.se/"
"EMBRC"	"http://www.embrc.eu/"
"EUROFLEETS2"	"http://www.eurofleets.eu/"
"IAGOS"	"http://www.iagos.org/"
"AnaEE"	"http://www.anaee.com/"
"JERICO-NEXT"	"http://www.jerico-ri.eu/"
"EUFAR"	"http://www.eufar.net/"
"Euro-Argo"	"http://www.euro-argo.eu/"
"SIOS"	"https://www.sios-svalbard.org/"
"Fix03"	"http://www.fix03.eu/"
"INTERACT"	"http://www.eu-interact.org/"
"ICOS"	"https://www.icos-ri.eu/"
"LTER-Europe"	"http://www.lter-europe.net/"
"ACTRIS"	"http://www.actris.eu/"
"CETAF"	"https://cetaf.org/"
"EUROCHAMP-2020"	"https://www.eurochamp.org/"
"ARISE"	"http://arise-project.eu/"
"EuroGOOS"	"http://eurogoos.eu/"
"IS-ENES2"	"https://verc.enes.org/ISENES2"
"EMSO"	"http://www.emso-eu.org/"
"LifeWatch"	"http://www.lifewatch.eu/"
"SeaDataNet"	"http://www.seadatanet.org/"
"EPOS-EU"	"https://www.epos-eu.org/"
"EPOS-IP"	"https://www.epos-ip.org/"
"EMPHASIS"	"https://emphasis.plant-phenotyping.eu/"
"DANUBIUS"	"http://www.danubius-ri.eu/"





Demo

<https://envrifair1.test.fedcloud.eu/OntoWiki>

Q & A