ENVRI-Hub NEXT

D7.2

First Report on Integration of Catalogue with the Analytical Framework

Status: Under EC Review

Dissemination Level: Public

Funded by the European Union

Abstract	
Keywords	Analytical Framework, API, Notebooks, VRE

This deliverable describes the connection between the Analytical Framework and the Catalogue of Services of the ENVRI-Hub Next. The Analytical framework uses the Catalogue of Services API to help scientists search and access data and metadata related to Essential Variables. Moreover, the Analytical Framework provides Virtual Research Environments to easily interact with the software provided, and a detailed description of this further possibility is provided. Future developments are described, but the deliverable focuses more on what the partners delivered in the first half of the project.

Revision History			
Version	**Date**	**Description**	**Author/Reviewer**
V 0.1	30/06/2025	ToC and first draft	Alessandro Turco (EPOS ERIC)
V 0.2	07/07/2025	External Review	Tjerk Krijger (MARIS) Ulrich Bundke (JUELICH) Claudio Dema (CNR)
V 0.3	16/07/2025	Document Review	Alessandro Turco (EPOS ERIC)
V 1.0	**20/07/2025**	**Final**	Alessandro Turco (EPOS ERIC)

Document Description			
D7.2 - First Report on Integration of Catalogue with the AF			
Work Package Number 7			
Document Type	Deliverable		
Document Status	Under EC Review	**Version**	1.0
Dissemination Level	Public		
Copyright Status	 This material by Parties of the ENVRI-Hub NEXT Consortium is licensed under a Creative Commons Attribution 4.0 International License.		
Lead partner	INGV		
Document Link	**https://documents.egi.eu/document/4039**		
DOI	**https://zenodo.org/records/16568173**		
Author(s)	• Dario De Nart (CREA) • Alessandro Turco (EPOS ERIC) • Daniele Bailo (INGV) • Rossana Paciello (INGV)		
Reviewers	• Tjerk Krijger (MARIS) • Ulrich Bundke (JUELICH)		
Moderated by:	• Matteo Agati (EGI)		
Approved by:	• Claudio Dema (CNR) - on behalf of DSB		

Terminology / Acronyms	
Term/Acronym	**Definition**
AF	Analytical Framework
CoS	Catalogue of Services
EHN	ENVRI-Hub Next
ENVRI	The ENVRI Community of Environmental Research Infrastructures
EOSC	European Open Science Cloud
ECV	Essential Climate Variables
EXV	Essential Variable
VRE	Virtual Research Environment
RI	Research Infrastructure
EH-VRE	Virtual Research Environment created for ENVRI-Hub
WF-t	Workflow tool element of the AF

Useful Reference: https://doi.org/10.5281/zenodo.14794634

Table of Contents

Table of Figures

Table of Tables

Executive Summary

The present deliverable assesses the connection between the Analytical Framework (AF) and the Catalogue of Services (CoS) of the ENVRI-Hub. The former is a collection of codes, libraries and notebooks that allow scientists to interact with the data and metadata offered by the project partners and by the ENVRI community at large; this collection of software can be downloaded and installed on local environments or directly accessed through the Virtual Research Environments provided by the project. The latter is the central repository for the mentioned data and metadata; it is the place where users should go to search for Essential Variables and related data and services.

A proper integration between the two is therefore crucial for reaching the project objectives, and the partners have worked on this since the beginning of the project. The CoS exposes an extensive and documented set of APIs which are exploited by the AF to satisfy the resource discovery and resource access use cases, which are the main scenarios identified.

The deliverable focuses more on the AF since the CoS is extensively described in D7.1. We describe the overall architecture, the libraries assembled and the notebooks collected. Moreover, the project also offers Virtual Research Environments with all the mentioned software pre-installed and ready to use by scientists. A detailed description of this powerful tool is provided.

We present the achievements of the first half of the project and highlight the improvements expected and planned for the second part. In particular, the role of Essential Variables needs to be made more visible, exploiting the work done by the related task force, which produced ad-hoc notebooks to exemplify how to interact with I-Adopt and Research Infrastructure services.

1. Introduction

The overarching goal of the ENVRI Hub Next (EHN) project is to consolidate and advance the ENVRI-Hub's robust conceptual and technical structure, enabling the ENVRI Science Cluster to provide interdisciplinary data-driven services that support climate change research, mitigation, adaptation, and risk assessment. It also aims to integrate the environmental science community into the European Open Science Cloud (EOSC), guided by the concept of Essential Climate Variables (ECVs) or more general Essential Variables (EXVs).

Access to raw data or even to pre-processed or aggregated data is often not sufficient to promote high-quality scientific results, especially when dealing with multi-disciplinary research. To overcome composition, processing and usability barriers (as stated in the project proposal), the new ENVRI hub offers the Analytical Framework (AF), including Virtual Research Environments (VREs) and Python libraries to easily interact with data/metadata services.

The goal of this document is to provide the reader with all the elements to answer the following fundamental questions:

- Is it possible to get data from the Catalogue of Services into a VRE?
- Are there any tools from the Analytical Framework to ease such a task?
- Is there any room for improvement on the Catalogue of Services side?
- Is there any room for improvement on the Analytical Framework side?

This document is aimed at system managers, software developers, and other IT professionals working at ENVRI Research Infrastructures or anyone interested in the ECV/EXVs. Therefore, the content requires a technical background that could not be condensed into the present deliverable. We prefer to address the topics directly, being sure that a modern environmental scientist/data analyst is able to find relevant resources, if needed.

This document heavily references D7.1, in which the Catalogue of Services is described in detail. We briefly recall the necessary elements, but we recommend reading **D7.1** to better understand the contents of the present document.

2. Analytical Framework Context

2.1. Overall Architecture and Scope

The Analytical Framework consists of various services aimed at improving data and service access within automated data processing environments and eventually leading to a better harmonisation of data processing pipelines.

The AF is heavily focused on VREs; therefore, a reference VRE (named EH-VRE as described in D5.1) has been made available to project participants; in addition, an application interface component (VRE-Lib) with the rest of the ENVRI-Hub through the CoS is provided. Finally, a collection of curated workflows is offered, providing documentation on how to access and use ENVRI-Hub resources to implement analytical data flows.

The AF components and their relationships are illustrated in Figure 1.

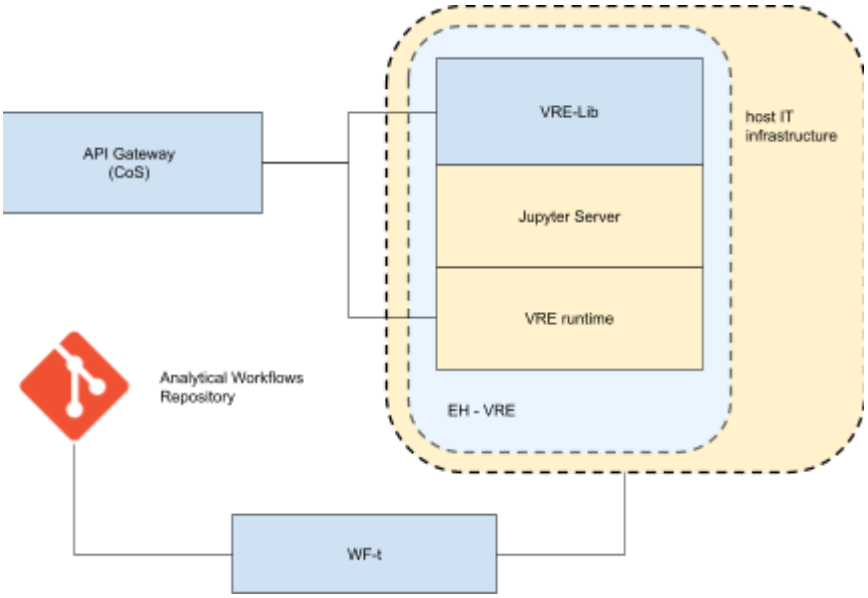

Figure 1 - *The Analytical Framework Components and Relationships*

The **VRE-Lib** is a Python package to be installed by VRE managers in their VREs and/or by individual researchers in their working environment of choice to access the services of the ENVRI-Hub. The VRE-lib is under development, and it will be published on PyPi (https://pypi.org/) as soon as a stable version will be made available by WP13 and WP14.

The **Analytical Workflow repository** is a curated collection of Jupyter Notebooks produced by the ENVRI-Hub-Next project, meant to illustrate data access and processing in the ENVRI-Hub.

The Workflow-Tools (WF-t) are a suite of software tools to be installed by Research Infrastructures (RIs) and/other institutions within their premises to allow the execution of Jupyter Notebooks as isolated functional blocks in data processing workflows.

A VRE demonstrational instance has been set up at https://vre.staging.envri.eu/, showcasing the minimum capabilities a VRE requires to be compliant with the ENVRI-Hub's Analytical Framework.

2.2. VREs and Integrations

Virtual Research Environments, herein VREs, are online systems helping researchers collaborate; typically, they are hosted and managed by a Research Infrastructure, E-Infrastructure or some other consortium and/or large organisation that distributes such a service to a number of end-users.

VREs in the research community vary in terms of underlying technologies and exposed services; however, they all support the Jupyter technology and allow running Python code.

The Analytical Framework, therefore, includes the following technologies that can be installed in any VRE in the ENVRI community:

- Python packages: software libraries that can be integrated into Python code;
- Jupyter notebooks: ready-made actionable workflows presented in a literate way;
- Jupyter Lab extensions: plugins that extend the Jupyter VRE functionalities.

In addition, the ENVRI-Hub Next project offers consortium members a dedicated VRE to develop and test workflows.

2.3. VRE-LIB

The VRE-Lib Python library is the integration component providing a Pythonic interface to the CoS to be used as a gateway for the ENVRI-Hub service portfolio.

It implements the following functionalities:

- Automated search through the CoS APIs;
- Metadata retrieval through the CoS APIs;
- Data access through the data owner's own APIs as described in the CoS.

The package is currently under development, and it will be published on PyPI as per the grant agreement by the end of the project. The final naming of the package will be decided within WP14 in accordance with the Envri-Hub-Next communication strategy.

2.4. Notebook Collection

The **Analytical Workflow repository** is, at its core, a collection of notebooks implementing data access and processing procedures, offering users insights on how individual RIs manage data access and how said data is intended to be used in processing pipelines. These notebooks indicate the current status of obtaining access to the data from the various RIs. This, in many cases, requires currently multiple different steps, which are not captured by harmonised APIs yet.

The repository is publicly available at ENVRI-Hub-NEXT / Analytical Workflow Templates · GitLab, and it includes the examples listed in Table 1:

Table 1 - RIs Notebook Examples

RIs Notebook Examples		
Notebook title	**Infrastructure**	**Illustrated Tasks**
AnaEE Weather Data Access API	AnaEE	Data access and visualization
Argo Data Access	Euro-Argo	Data access and visualization
ACTRIS_EXV-related_datasets	ACTRIS	Data access and visualization
ENVRI-Hub NEXT - Hand On the EPOS API	EPOS	Data access and visualization
IAGOS_data_access_L3	IAGOS	Data access and visualization
ICOS_data_access	ICOS	Data access
SeaDataNet - Beacon	SeaDataNet	Data access and visualization
SeaDataNet - CDI API	SeaDataNet	Data access

3. Cos-AF Interactions

In this section, we illustrate the interactions between the CoS and the components of the AF.

3.1. Use Cases

The interaction between CoS and AF is functional to the implementation of the use cases described herein.

3.1.1. Resource Discovery

Users through the scripting environment should be able to find in the CoS a dataset and/or other resources, such as the Fair Data Point repository, which implements the Catalogue of Data for ENVRI-Hub NEXT (see Deliverable 7.5 for more information).

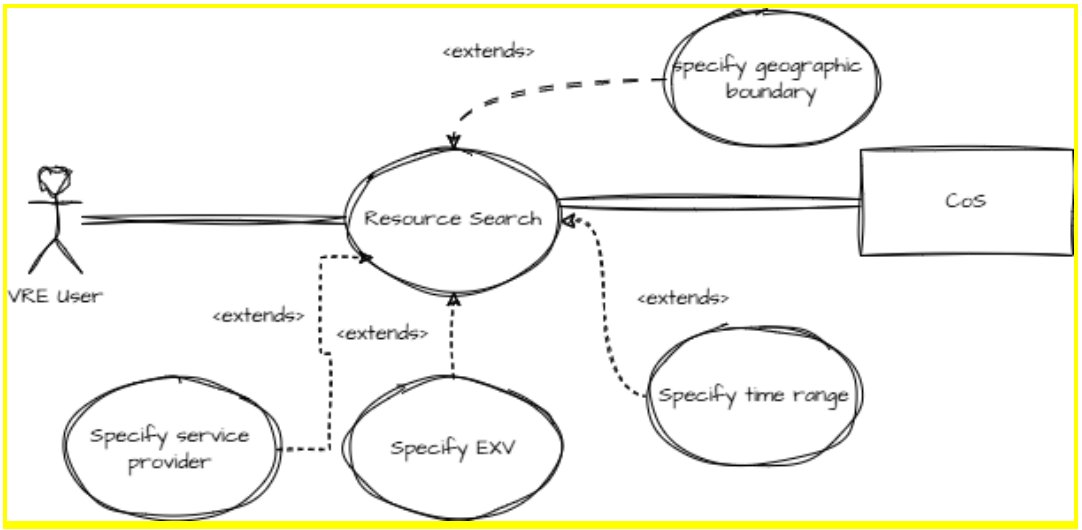

***Figure 2**: Resource Discovery Use Case*

The use case is identical to the one implemented by the WebUI, but in this case, it is realized through a text-based scripting interface, which allows the User to automate such an interaction flow as depicted in Figure 2.

3.1.2. Resource Access

Users through the scripting environment should be able to access a dataset indexed on the CoS.

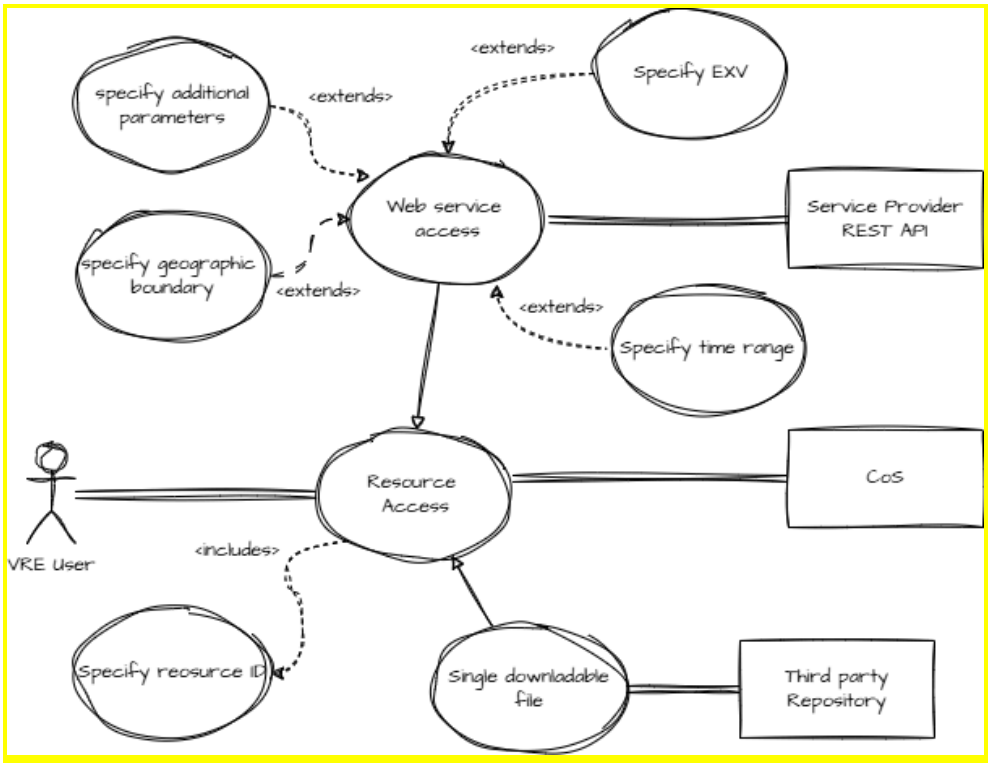

Figure 3: Data Access Use Case

In this Use Case, the CoS acts as an indexing service, allowing the user, from their scripting environment, to retrieve a resource's metadata. Depending on the type of Resource, the CoS can either provide a direct download link or return an address to another Web Service (see Figure 3).

In the latter case, according to the metadata exposed on the CoS, the user can slice the dataset using the functionalities exposed by the Service provider's REST API.

At the implementation level, the ENVRI Hub library creates a different client according to the type of service offered by the service provider.

3.2. Technical Implementation

In this section, we introduce the technical implementation work done so far.

3.2.1. ENVRI Hub Library and CoS APIs

The VRE library (VRE-Lib) is implementing the above-described use cases, and interacts with CoS in a client-server fashion, with the ENVRI Hub library querying the Resources Service described in the CoS specification (Swagger UI). Such a specification illustrates endpoints, which correspond to functionalities, and *Schemas*, which correspond to data structures that the ENVRI Hub Library internally models as classes.

3.2.2. Resource Search

The resource search use case is implemented by querying the *Search* endpoint provided by the Cos's Resource Service, which allows performing search-engine-like queries and to retrieve a specific resource given its identifier, returning *Distribution* objects.

Resources can be filtered by specifying geographical bounds, time frame bounds, or by service owner. WP7 further extended the CoS data model and Search interface to allow EXVs as a parameter in the *Search* endpoint, allowing users to retrieve resources by listing the EXV they are expected to contain.

3.2.3. Data Access

The implementation of the data access use case relies on the Distribution objects returned by the Search functionality described above or retrieved by the Resources Service, given their unique identifier. Distribution objects have a *href* property containing the link to access a *Description* of the resource, i.e. a more detailed object that ideally sits closer to the actual data than the Distribution object.

The Distribution object exposes several properties allowing to infer the data access methods, chief among which is the *type* property that can have two values:

- **WEB_SERVICE**: The resource is exposed as a Web Service, allowing parametrized interactions;
- **DOWNLOADABLE_FILE**: The resource is packaged in a monolithic file. The Available Format property contains the hyperlink to access said file. availableFormats can be found both in the Distribution and in the Detail object.

In the latter case, accessing the *href* property's content with a GET request would let the user download the file, thus allowing the user to manage it in their VRE of choice.

In the former case, resource access is mediated with a protocol, and the URL in the *href* property would produce a *Description* object including information about such a protocol. Interaction protocols can be explicitly described inside the Description object with parameter declarations or can be linked to an external URL in the *serviceDocumentation* property.

In the VRE library (VRE-Lib), according to the Distribution type attribute, different DataAccess objects are instantiated for the Distribution, implementing either file download or REST service access. In the latter case, the DataAccess object is further specialised in a series of subclasses that correspond to one of the following:

- A well-known standard web service protocol, such as SPARQL;
- A well-known bespoke protocol used across an RI;
- A bespoke protocol described in an OpenAPI 3.0 file.

From the user's perspective, according to the different types of data transfer protocol and formats provided by the Service Provider, the VRE Library instances a different Data Access Object for each resource, as shown in Figure 4.

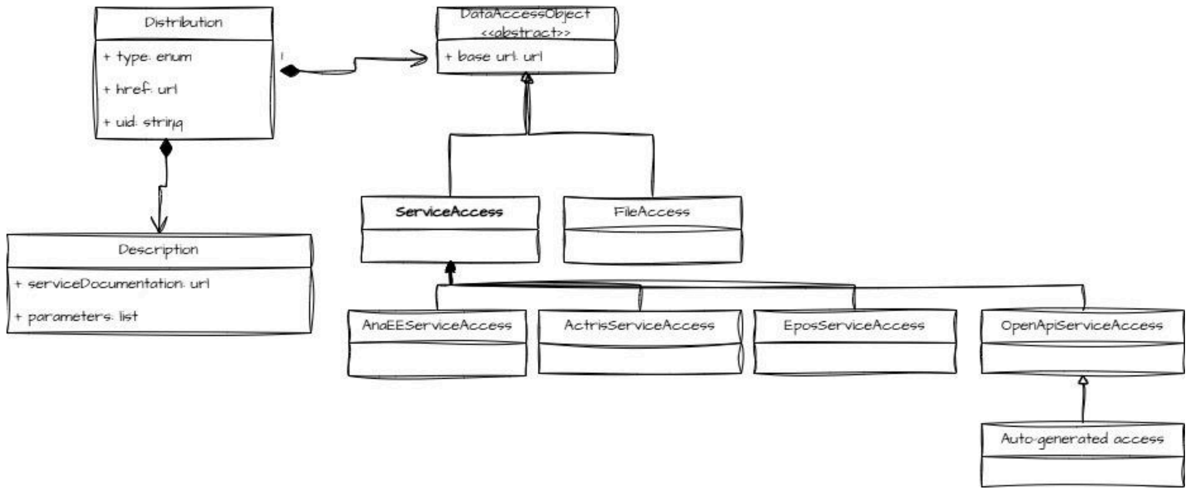

Figure 4*: Class Diagram of the VRE-LIB Entities that implement Data Access*

Right now, RIs participating in the ECV task force have provided us, using Jupyter notebooks, with working connections and data access examples to seven services, which will have their own dedicated data access object in the first release of the library. The task force will try to harmonise this as much as possible in terms of generic environmental parametrisation (geo-spatial/temporal queries).

Other, non-integrated RIs services can be accessed through their RESTful services if an OpenAPI 3.0[1] specification is made available. In such a case, the VRE Library would build a dedicated client class exposing methods and contents in a Pythonic way, as shown in Figure 5.

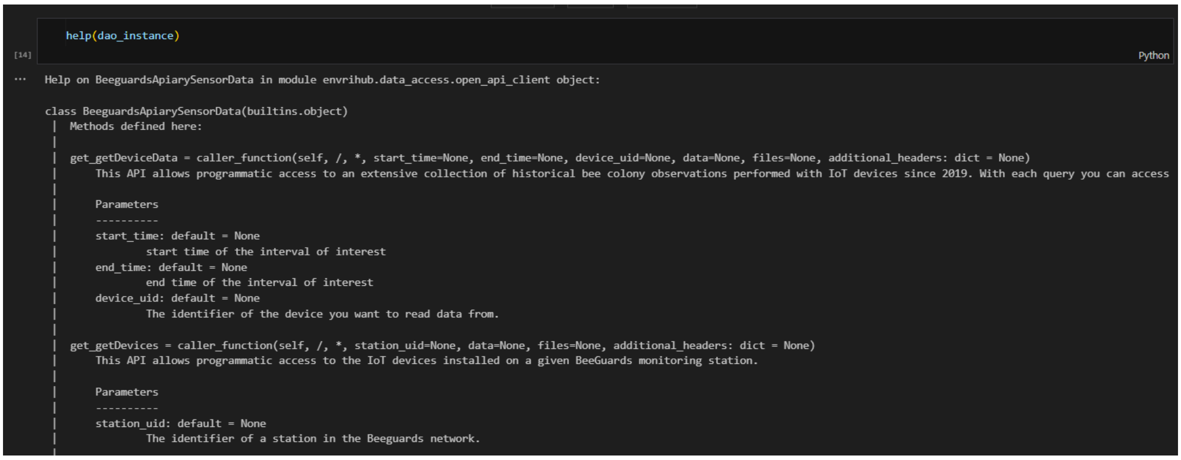

Figure 5*: Example of a Data Access Class Instance generated from OpenAPI 3 Specifications*

3.2.4. Notebooks and Data Sources

Currently, the Notebooks provided in the workflow repository do not exploit the capabilities of the CoS, but rather interact with individual RIs' services following their specific procedures. Right now, notebooks demonstrate the data access use case without interacting with the CoS services and/or adopting any piece of its data model.

[1] https://spec.openapis.org/oas/v3.0.0.html

The EXV TF, together with the WP7-8/13-14, will aim for data access harmonisation via the ENVRI-HUB library once it is published.

4. Conclusions

The Analytical Framework is the hub component that scientists are going to use to perform their studies, while the Catalogue of Services is the central repository for the data and services offered by the partners and the ENVRI community at large in the future. Therefore, the connection between the two is paramount for achieving the project goals and for making a real impact.

Although we planned further improvements for the second half of the project, the present deliverable states clearly that the integration is already solid and reliable for both the use cases of interest: resource discovery and resource access. The work done in the EXV Task Force guarantees that Essential Variables will become the primary actor by the end of the project, increasing their visibility and findability through the I-Adopt framework.

Virtual Research Environments simplify the interaction with the developed libraries. They can attract more users to the hub, and their exploitation will be assessed in the near future, since their technical robustness has been proven.

The overall status of the hub from the technical perspective is satisfactory at this stage of the project. All structural elements are in place and properly working. The work ahead now is to enrich them, facilitate access, and strengthen the connections among them so that the barriers mentioned in the project proposal can be really overcome.

5. References

References	
No	**Description/Link**
R1	**VOMS, an Authorization System for Virtual Organizations.** R. Alfieri et al. In Grid Computing. AxGrids 2003. Lecture Notes in Computer Science, vol 2970. Springe DOI: 10.1007/978-3-540-24689-3_5