



SIMDAT

Data Grids for Process and Product Development using Numerical Simulation
and Knowledge Discovery

Project no.: 511438

Grid-based Systems for solving complex problems – IST Call 2
Integrated project



Deliverable

***D 19.1.2 Report on consolidated technology prototypes
for Virtual Organisations and Ontologies***

Start date of project: 1 September 2004

Duration: 48 months

Due date of deliverable: 01.04.06

Actual submission date: 09.05.06

Lead contractor for this deliverable: European Centre for Medium-Range Weather Forecasts

Revision: 1.0

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination level		
PU	Public	X
PP	Restricted to other programme participant (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Revision history

Date	Version	Author	Modification
22.03.06	0.6	Marta Gutierrez	Version 0.7
23.03.06	0.7	Helmut Ackermann	Ontology Requirements
03.04.06	0.8	Marta Gutierrez	Add Mike Turner and Moritz Weiten Comments
04.04..06	0.81	Helmut Ackermann	Corrections Ontology and add ontology components diagram

Copyright

Copyright © European Centre for Medium-Range Weather Forecasts and other members of the SIMDAT consortium, www.simdat.org, 2006.

TABLE OF CONTENTS

INTRODUCTION	4
PART I: THE VIRTUAL ORGANIZATION (VMC)	5
1 VO Technology requirements	7
1.1 Security	7
1.1.1 VO Security Functional Requirements	7
1.1.2 VO Security Performance Requirements	8
1.2 Management and Monitoring	8
1.2.1 VO Monitoring and Management Functional Requirements	9
1.2.2 VO Monitoring and Management Performance Requirements	9
2 VO Use Cases	10
2.1 VO establishment and Membership	10
2.2 VO Management	12
2.3 Security Services	15
2.4 Monitor VMC	17
3 VO Components view	22
4 VO Technology Adoption	25
PART 2: ONTOLOGY	26
1 Ontology Technology Requirements	27
1.1 Multilingual Thesaurus	27
1.2 GT observations ontology	27
2 Ontology Use Cases	29
3 Ontology Components View	32
4 Ontology Technology adoption	34
CONCLUSIONS	35
References	35
GLOSSARY	36

INTRODUCTION

This document reviews the technical requirements for the Virtual Organisations and Ontology WP of the SIMDAT Meteo Activity.

PART I: The Virtual Meteorological Centre

The document includes some of the outputs from the VO working group meetings held in the last five months. It also follows some of the recommendations introduced by the technology champions BAE and IT Innovation.

The full description of the VO establishment and life cycle is still a work in progress; the requirements and use cases presented here are focused towards the integration of security services in the VO.

PART II: Ontology

The project plan about the ontology part presented in this document is a result of internal discussions by the meteorological partners and a workshop together with the Ontoprise GmbH. It is an overview of what should be achieved in the second project phase. The design of details is still in progress.

PART I: THE VIRTUAL ORGANIZATION (VMC)

The SIMDAT meteo-activity comprises five partners ⁽¹⁾ which have agreed to define and establish a Virtual Organization (VO) in the context of the meteorology domain.

SIMDAT meteo has proposed the concept of a Virtual Meteorological Centre for the establishment of a Virtual Organization.

The virtualisation of the five partners into a single resource to provide a unified view of meteorological data will arise some of the well-known and not that well-known challenges of the grid computing world:

- Integration of multiple data sources and formats.
- Data transfer challenges
- Data Replication
- Security
- Integration of Operational Services across the partners.

Amongst the numerous definitions of a VO, that of Camarinha-Matos ⁽²⁾, describes closer the concept of a VMC:

“A virtual organisation is a set of co-operating (legally) independent organisations, which to the outside would provide a set of services and act as if they were one organisation”

According to this definition,

The VMC is a set of organisations, people and resources with the common goal of offering integrated meteorological data and value-added services to the community.

Who might be involved in it?

- The National Meteorological Services worldwide
- WMO specific programs
- European Institutions involved with meteorological activity.
- Anyone who agrees to the data access policies of the VO to obtain meteorological data.

What are the resources?

- A dedicated operational network that transfers critical meteorological data. ⁽³⁾
- Forecasting models
- Data Archives
- Data Services: Visualization, Post processing, Delivery.

What are the relations between organisations?

- The organisations within the VMC will maintain long term relationships based on a Framework of Trust.
- There will be no hierarchy or level differentiation between the organisations; they all maintain equal rights within the VMC

The practices and software from the SIMDAT meteo activity will initially integrate into the VO the five partners involved; if successful, the VMC will interconnect with other meteorological centres in different regions world wide.

1 VO TECHNOLOGY REQUIREMENTS

The technology requirements are presented from D18.1.1 and extended to integrate security services into existing architecture.

1.1 Security

- *Trust among the entities:* All the VMC members have to establish secure connections to exchange the catalogue. In that case the authentication, authorization, data integrity and non-repudiation of the data are compulsory.
- *User access:* The VMC must be able to identify, authenticate and authorize users to access the data and services exposed through the VO. The users will be authenticated at their institutions and granted privileges to access the VO. While users will still be registered at one of the VMC organisations the VO will provide “Federation of Identity” so users are recognised at any point in the VO. Authorization services will provide flexible and fine grained access control to VO and local services within the infrastructure.
- *Data protection:* The infrastructure must ensure the integrity of the data exchanged between the partners or sent to the users.
- *Availability:* The VO must ensure that system availability is not compromised by external interfaces and that operational services maintain a high QoS.

1.1.1 VO Security Functional Requirements

VO Security Functional Requirements	Priority
<ul style="list-style-type: none"> - To authenticate • Entities of the system (Data Repository, Data Transfer Servers, ...) • Users of the system (VMC administrator, meteorologist, ...) 	High
<ul style="list-style-type: none"> - To authorize users to access data • Provide global data access policies • Provide translation services into local policies 	High
<ul style="list-style-type: none"> - To provide a global access control policy managed by each partner and integrated into their existing security policy 	High
<ul style="list-style-type: none"> - To secure communication within the entities of the system • Commands encrypted • Metadata synchronization encrypted 	High

<ul style="list-style-type: none"> - To protect the data • Ensure the integrity of the datasets circulating over the VMC. • Ensure the non-repudiation of certain types of data 	High
<ul style="list-style-type: none"> - To audit • Maintain an audit of all activities and transactions for billing and non-repudiation purposes 	High
<ul style="list-style-type: none"> - To ensure system is : • Available when needed • Reliable and performs at expected levels 	High

1.1.2 VO Security Performance Requirements

VO Security Performance Requirements	Priority
<ul style="list-style-type: none"> - To ensure that the security mechanisms like encryption doesn't affect the VMC response time. 	High
<ul style="list-style-type: none"> - To ensure that the data integrity and non-repudiation mechanisms don't affect the data transfer rates in such way that the VMC cannot fulfill its QoS requirements. 	High
<ul style="list-style-type: none"> - To ensure that the secure data transport protocols used have performances at least comparable with HTTPS or SSL-like protocols. 	High

1.2 Management and Monitoring

The VMC will be a production infrastructure available 24/7. For this reason the resilience of the system must be ensured; backup and recovery procedures have to be defined for each VMC partner.

Monitoring tools must be available to check and supervise the VMC activity. User activity, Service Level Agreements, Data exchange flow and systems must be monitored and logged.

Monitoring tools will be defined so that the administrators will be able to check the status of the system either at a local site or at the sites where the activities involved relate to other nodes.

Each partner is responsible for its VMC node and management tools will also be available at a local level:

- To create, remove users and modify users profiles
- To add, remove and modify databases and datasets in the catalogue

1.2.1 VO Monitoring and Management Functional Requirements

Monitoring and Management Requirements	Priority
<ul style="list-style-type: none"> - Ensure the resilience of the system 24/7 • Define backup procedures • Define recovery procedures 	High
<ul style="list-style-type: none"> - Monitor and log the VMC activity • At a local level on each site (user activity, data flow, SLAs) • At a global level • Provide reporting tools to analyze the VMC activity over a long time period 	High
<ul style="list-style-type: none"> - Manage at a local level the VMC • User management and registration • System management • Database management • Catalogue management 	High

1.2.2 VO Monitoring and Management Performance Requirements

VO Monitoring and Management Performance Requirements	Priority
<ul style="list-style-type: none"> - To ensure that the Logging and Management doesn't affect the core activity of the infrastructure 	High

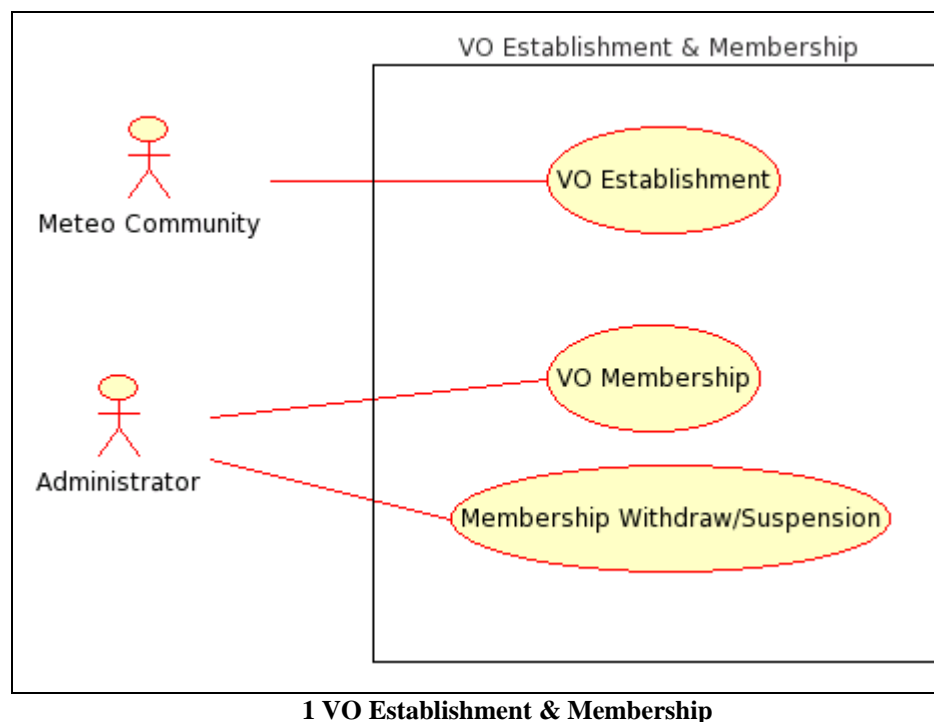
2 VO USE CASES

The following use cases describe the integration of security services and VO establishment and membership. VO monitoring and management have been included and updated from its original description D18.1.1⁽⁴⁾

In these use cases a user can be a people, a group of people (e.g. organisation) or a system. A user can have different roles depending on its interactions with the VMC:

- Administrator
- Registration officer
- Operator
- Data receiver (i.e. Scientist , NMS)
- Data provider (NMS)
- Metadata provider (NMS)

2.1 VO establishment and Membership



Note: The Meteorological Community will not be a system user as such, but it represents the community, bodies, group of people in charge of defining the VO policies.

Name	VO Establishment
UseCaseId	UC_VO2.1.1

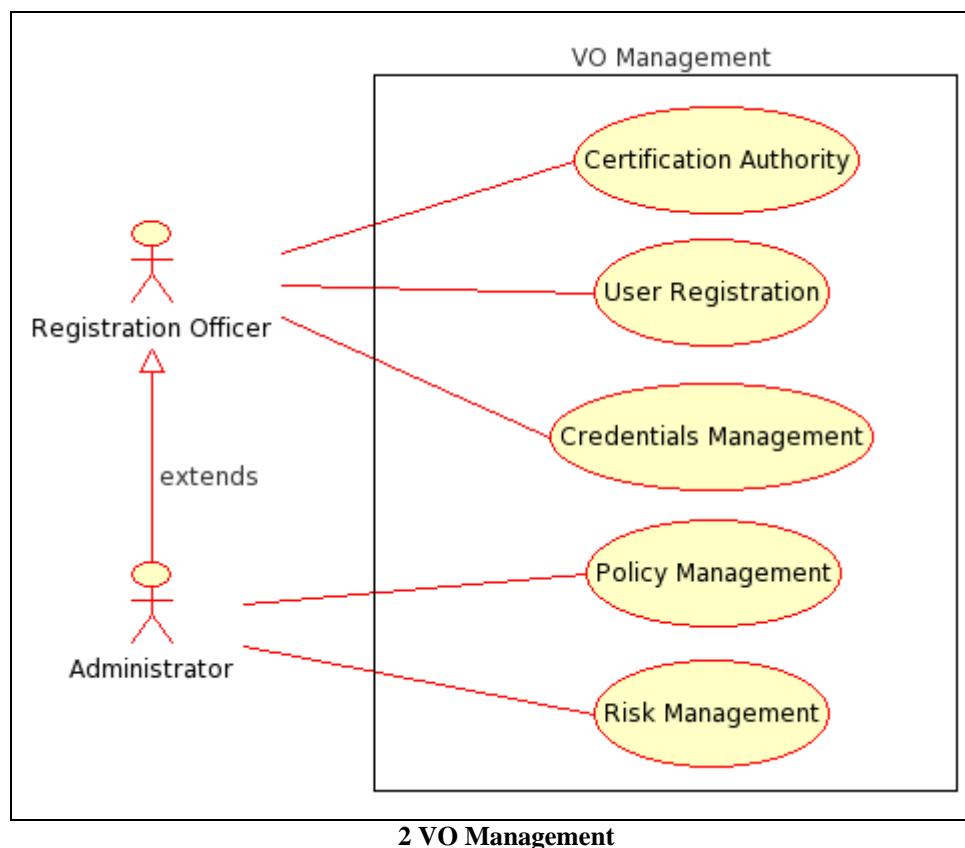
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
<p>The advantages of establishing meteorological VO have been identified. The VO policies required to operate the system as whole from each of the individual organisations are defined.</p> <p>Some of the existing data access policies are still applicable to the V-GISC e.g. WMO Resolution 40</p>	
Actors	
<ul style="list-style-type: none"> - Meteo Community 	
Pre-Conditions	
<ul style="list-style-type: none"> - The VMC concept is defined to set up a VO. 	
Post-Conditions	
<ul style="list-style-type: none"> - Global Data Access Policies, Roles and SLA's to access the VMC are stated. <p>It is also defined which organizations are eligible to join the VMC</p>	

Name	VO Membership
UseCaseId	UC_VO2.1.2
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
A new organisation joins the VMC.	
Actors	
<ul style="list-style-type: none"> - Administrator, Organisation 	
Pre-Conditions	
<ul style="list-style-type: none"> - The policies that apply to the VO are agreed. - Contracts and SLA's are signed to ensure service provider and consumers maintain requests and service levels to an agreed level. 	
Post-Conditions	
<ul style="list-style-type: none"> - Membership is established based on a Trust Framework. - Local existing policies from the new member need to be translated into the VO policies. 	

Name	Membership suspension
UseCaseId	UC_VO2.1.3
Date Created	16/03/06

Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	A member withdraws from VO membership or the VO ceases or suspends temporarily the participation of one of the members.
Actors	- Administrator
Pre-Conditions	- An organization or VMC system is part of the existing VMC
Post-Conditions	- A member notifies the VO to cease membership. - The VO notifies the rest of VMC members the suspension / withdraws of the member in question.

2.2 VO Management



Name	Certification Authority
UseCaseId	UC_VO2.2.1
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF

Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
Each organisation should run or have access to a Certification Authority that will issue PKI certificates to prove user's or host's identity.	
Actors	
<ul style="list-style-type: none"> - Registration Officer 	
Pre-Conditions	
<ul style="list-style-type: none"> - Admin has access to a Certification Authority. - Root Certificate maintains its validity for long term. (10 years) 	
Post-Conditions	
<ul style="list-style-type: none"> - X505 certificates are issued according to the policy stated by the CA - User Certificates are valid for one year. 	

Name	User Registration
UseCaseId	UC_VO2.2.2
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
A user registers with one of the organisations that belong to the VO. The VO will not maintain any centralized repository of user details.	
Actors	
<ul style="list-style-type: none"> - End user, Administrator 	
Pre-Conditions	
<ul style="list-style-type: none"> - A user agrees to the conditions stated by the VO policies to access the system. - Admin accepts registration details and a certificate/role request is processed. - An SLA is signed to provide the user with the service acceptance levels. 	
Post-Conditions	
<ul style="list-style-type: none"> - User obtains an X509 certificate and is assigned roles according to the user profile. 	

Name	Manage Credentials
UseCaseId	UC_VO2.2.3
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity

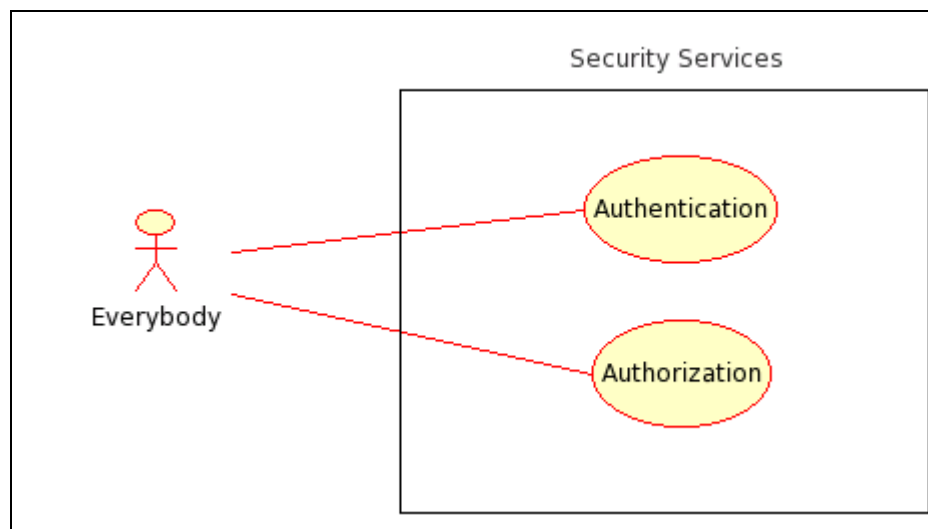
Use Case Goal
Certificates must maintain validity to be used within the VMC. Administrators shall renew certificates if still eligible to access the system or revoke them if credentials are stolen, lost or compromised.
Actors
- Admin
Pre-Conditions
- A certificate has been issued
Post-Conditions
- Certificate is renewed at expiring date if the user's agreement with the VO is still valid.
- Certificate is revoked if the keys have been compromised.

Name	Policy Management
UseCaseId	UC_VO2.2.4
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
Policies might be updated during the lifetime of the VO, the system should be able to dynamically incorporate new or updated policies	
Actors	
- Administrator	
Pre-Conditions	
- A policy has been agreed and approved by the Meteorological Community.	
Post-Conditions	
-The policy is updated and applied to the system.	

Name	Risk Management
UseCaseId	UC_VO2.2.5
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
The system should identify potential security risks and elaborate the appropriate response to mitigate the risk.	
E.g. A registration officer will revoke all certificates should the root CA certificate be compromised.	
Actors	

- Administrator
Pre-Conditions
Risks have been identified and scaled according to the consequences that it would have in the system.
Post-Conditions
- A mitigation plan is designed to respond with the appropriate measures to the risk in place.

2.3 Security Services

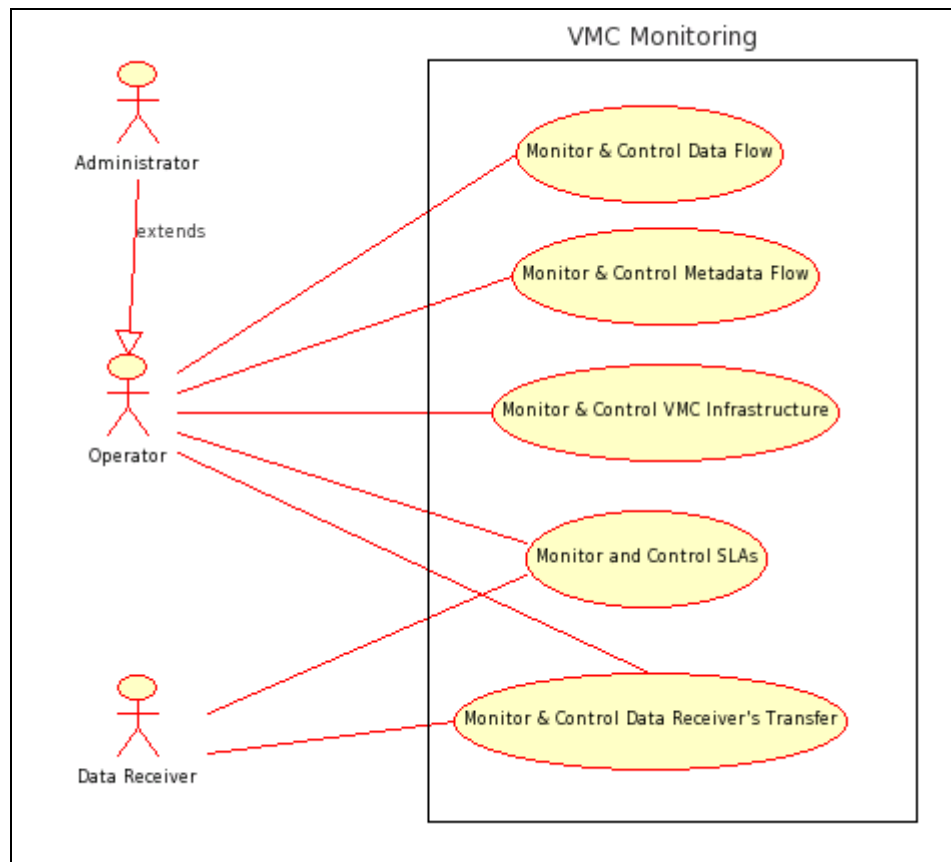


3 Security Services

Name	Authentication
UseCaseId	UC_VO.2.3.1
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
	User logs-in with the organization they belong to. Once identified by their home institution, they will be granted rights to access what they are allowed to within the VO.
Actors	
	- All
Pre-Conditions	
	- A user has previously registered with an institution, member of the VO.
Post-Conditions	
	- User obtains short lived credentials with identity and roles to maintain a session with the VO.

Name	Authorisation
UseCaseId	UC_VO2.3.2
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
<p>A user requests some data or service under restricted access. An authorization module will check the user's privileges against the data or service access policy or roles and grant/deny access.</p> <p>e.g.: admin, belongs to the Met Office, and holds the VO role: Metadata Manager. The VO policy defines that Metadata Manager can publish, update, delete and start/stop synchronization service with the rest of the nodes.</p> <p>e.g. user X registered at the metOffice holds a VO role of "scientist" and wants data kept in MARS archive The VO policy states that "a scientist can access any WMO data". Access to the MARS archive will have its own local existing policies. To provide an inbound request to access MARS (from a VO user not registered with the ECMWF), ECMWF needs to translate the incoming VO role into the local policy and grant authorization accordingly.</p>	
Actors	
- All	
Pre-Conditions	
- A user has logged in and maintains a valid representation of identity and privileges/roles	
Post-Conditions	
- The user is granted or denied access by the authorization module	

2.4 Monitor VMC



4 Monitor VMC

Name	Monitor & Control Data Flow [<i>Monitor VMC</i>]
UseCaseId	UC_VO2.4.1
Date Created	04/01/05
Last Updated by	Guillaume Aubert ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	<p>The VMC is a production system, which has to be available 24 hours a day 7 days a week. The operators or administrators need to be able to monitor and control the data acquisition and raise alarms if the expected data is not received on time. They also need to be able to monitor how the data flows between the VMC nodes and raise alarms in case of problems (see [2.4.5] Monitor & Control VMC Infrastructure for more details). They are also granted rights to restart the failing data transfers.</p> <p>Each partner's operators have local monitoring tools to check their VMC nodes status. They also have global monitoring tools to have a full view of the VMC and report problems incoming on the other nodes to the partners operators.</p> <p>Each VMC node is controlled at a local level. For example only the DWD</p>

operators can control the DWD VMC nodes. Météo France operators can only detect the problems with the monitoring tools and report them to DWD
Actors
<ul style="list-style-type: none"> - Administrator - Operator
Pre-Conditions
<ul style="list-style-type: none"> - An administrator has been authenticated and is authorised to manage the VMC - An Operator has been authenticated and is authorised to manage VMC
Post-Conditions

Name	Monitor & Control Metadata Flow [Monitor VMC]
UseCaseId	UC.2.4.2
Date Created	04/01/05
Last Updated by	Guillaume Aubert ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
	The operators or administrators need to be able to monitor and control the metadata acquisition and raise alarms if the expected data is not received on time. They also need to be able to monitor the metadata flow between the VMC nodes and raise alarms in case of problems (see [4.4.4] Monitor & Control VMC Infrastructure for more details).
Actors	
	<ul style="list-style-type: none"> - Administrator - Operator
Pre-Conditions	
	<ul style="list-style-type: none"> - An administrator has been authenticated and is authorised to manage the VMC - An Operator has been authenticated and is authorised to manage the VMC
Post-Conditions	

Name	Monitor & Control Data Receiver's Transfer [Monitor VMC]
UseCaseId	UC_VO2.4.3
Date Created	04/01/05
Last Updated by	Guillaume Aubert ECMWF
Project phase	WP19.1.2

Parent Package	Meteorology Activity
Use Case Goal	
<p>The operators or administrators or the data receivers (data receiver) himself need to be able to monitor and control the data receiver's transfers. Data receivers subscribe to receive datasets when they are available.</p> <p>Then they can monitor them and start, restart or stop the transfers. The operators have a global view of the VMC state and they can start, stop, restart the user's transfers and raise an alarm if they cannot solve the problem. The administrators have the same privileges as the operators. The operators and the administrators have precedence over data receivers and can when a major problem arises prevent the data receiver to start, stop and restart their transfers.</p>	
Actors	
<ul style="list-style-type: none"> - Administrator - Operator - Data Receiver 	
Pre-Conditions	
<ul style="list-style-type: none"> - An administrator has been authenticated and is authorised to manage the VMC - An Operator has been authenticated and is authorised to manage VMC - A Data Receiver is authenticated and authorised to manage its transfers 	
Post-Conditions	

Name	Monitor & Control SLAs [<i>Monitor VMC</i>]
UseCaseId	UC_VO2.4.4
Date Created	16/03/06
Last Updated by	Marta Gutierrez ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
<p>The operators, administrators and data receivers (data receiver) need to be able to monitor and control that services are performing according to SLA's.</p>	

Data receivers will be able to notify operators or administrators when the service has breached the SLA.

Administrators and operators should be able to check services across the VO maintain the SLA's levels and if an SLA has been breached the appropriate measures have taken place.

Actors

- Administrator
- Operator
- Data Receiver

Pre-Conditions

- An administrator and operator have been authenticated and is authorised to monitor and control SLA.
- A Data Receiver is authenticated and authorised to notify the Operator if the SLAs are breached.

Post-Conditions

Name	Monitor & Control VMC Infrastructure [<i>Monitor VMC</i>]
UseCaseId	UC_VO2.4.5
Date Created	04/01/05
Last Updated by	Guillaume Aubert ECMWF
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
<p>The operators or administrators need to be able to monitor and control the VMC nodes. They can check if the VMC is behaving properly. They have tools to monitor the state of the network between the VMC nodes. Operators are also granted to start/stop the VMC nodes causing problems and do simple recovering operations.</p> <p>Administrators can perform operations similar to the operator's ones but they have more privileges to perform deeper operations. They have access to the full software and can change, upgrade, patch the VMC nodes. They can also become a user to diagnose problems under the user's environment. They can kill jobs like irrelevant queries consuming resources (like retrieving the full catalogue). Tools are available to do the monitoring at a local (node's view) and global level (VMC view). The control will be always done at a local level (only the UKMO operators can control their nodes).</p>	
Actors	
<ul style="list-style-type: none"> - Administrator - Operator 	
Pre-Conditions	
<ul style="list-style-type: none"> - An administrator has been authenticated and is authorised to manage the VMC - An Operator has been authenticated and is authorised to manage users 	
Post-Conditions	

3 VO COMPONENTS VIEW

The following components will be integrated into the existing architecture to provide VO management and security services.

Credentials Manager: This component represents a Certificate Authority. The CA will be responsible of issuing or revocation of X509 certificates.

Credential Repository: Component to store user credentials (LDAP, KeyStore) or support delegation mechanisms (MyProxy).

Authentication Component (AuthN): This component should interact with the credential storage to obtain the user's delegated credentials and asserts its identity to any of the involved partners.

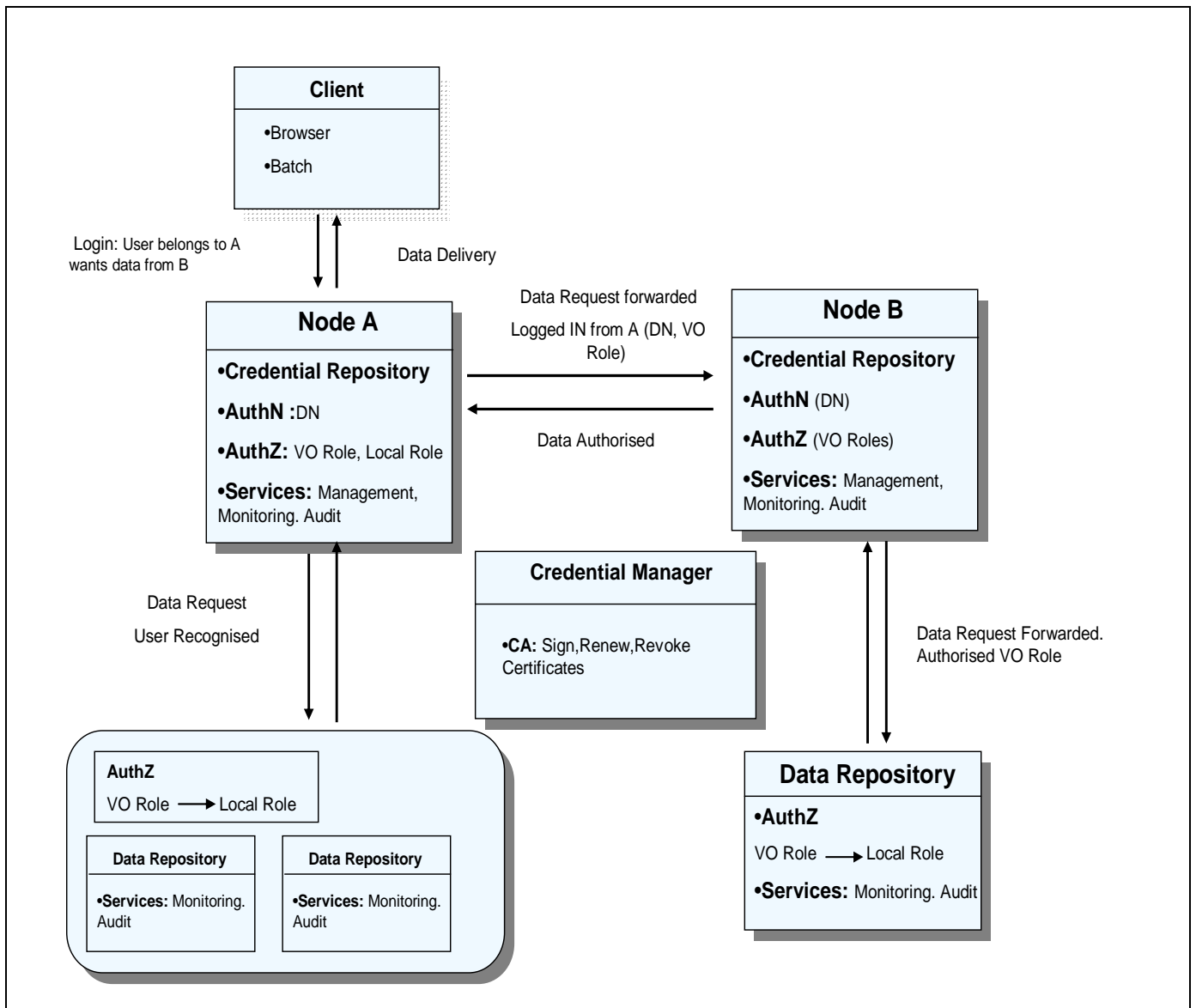
Authorization Component (AuthZ): This component should grant privileges (roles-policy based) to a user according to policies established between the VO organisations to provide access to the VO services.

Trust Management: This component will deal with any issues related to the establishment of trust between organisations.

Audit Component: This component controls statistics of user activity at any point in the VO.

Management Component This component provides the interfaces to manage the DCI. It facilitates management of users, databases and the catalogue.

Monitoring Component This component provides the monitoring services of the DCI. It allows operators to check that data and metadata acquisition are on schedule, to supervise the state of the different DCI nodes, to monitor user data transfers and interrupt, to delay or promote data transfers as appropriate.



5 Components View

The diagram 5 displays a view of the components within the existing architecture. Security and Management services will be integrated with the Node and the Data Repository.

The following scenario is a walk through the system components where a user that is registered with organisation A requires data from organisation B.

- 1) User logs in through a batch or browser client.
- 2) If delegation is required a short lifetime proxy certificate would be obtained and kept in a credential repository

- 3) The user's session will contain information with the user's identity and privileges granted by the organization A.
- 4) The user will be provided with some authority recognised throughout the VO at the Node level. (This would be a VO Role assigned to the user at registration time by acceptance of the terms and conditions of use)
- 5) A data request is identified at organization B, and Node A will forward the request and the user's session details to the Node B.
- 6) When the request arrives to Node B, the services:
 - Will verify that the user comes from a Trustworthy source, generally membership of the VO should provide with the basis for the trust relationship.
 - Will look at the privileges that Node A has granted, and check against the VO policy if the request can go forward to the data repository. (Perform VO authorization)
 - Once in the data repository, the authorisation privileges need to be mapped or translated the local policies and extract the data for delivery. (Perform local authorization)

From step 1 communication would be encrypted all the way through.

In the cases where a VO member provides access to more than one data repository the authorisation service will be abstracted and possibly offered as a standalone service.

Special Agreements:

When an organisation requires special access and conditions to dedicated members, a local policy will then be published to the local Node ONLY, so users could agree to the local terms.

For example, Met Office wants to offer access to critical data - only to its users- and not make it available to the rest of the VO. Met Office users will need to agree to the local policy, explicitly stated to this organisation.

In the same line, there might be services only available to specific organisations and not to the whole VO.

The replication of certain type of data, e.g. observations will be a service only available for the NMS, as ECMWF and EUMETSAT do not distribute this sort of data. A dedicated policy involving the members concerned, should then be published to provide the replication service.

4 VO TECHNOLOGY ADOPTION

The second phase of SIMDAT will start with an evaluation of the existing security frameworks. By the end of phase two the VO WP will deliver the security components and an initial design of the monitoring services.

The technology champions are providing recommendations, best practices ⁽⁵⁾ and roadmaps of the relevant standards and adaptors to some security frameworks that will fulfil some of the VO requirements.

As the existing components in SIMDAT Meteo migrate towards the integration of WSRF interfaces, the security services will be implemented considering the service oriented architectures standards and frameworks.

The following projects of relevance for the VMC have been identified and therefore will be watched closely or adopted if requirements are fulfilled.

PBAC ⁽⁶⁾ (Process Based Access Control) This module is integrated in GRIA and provides authorization to stateful web services interfaces based on a business process workflow and policy enforcement points.

TrustCoM ⁽⁷⁾: FP6 project that is building a framework for trust, security and contract management for VO supporting Grid Infrastructures and SOA.

GT4 ⁽⁸⁾: Provides some security components such as:

- GSI: to support single sign-on and delegation mechanisms
- simpleCA: Certification Authority
- MyProxy Server: Credential Repository
- SGAS: SweGrid Accounting System, with components for policy management, accounting, user logging.

GridShib ⁽⁹⁾: Is an NMI project providing a policy controlled grid framework for the Shibboleth ⁽¹⁰⁾ project.

NDG AA ⁽¹¹⁾: The NERC Data Grid (NDG) has designed and implemented an Attribute Authority (AA) model. The AA issues Attribute Certificates with user roles information. AC are presented to the different organisations and reissued by the local AA according to a trust relationship and role mapping into local roles.

VOMS and **CAS** will not be considered as they are based on centralised infrastructures.

Web Services Security Standards of relevance:

- **WS-Security**: Message Integrity and Confidentiality. (Implementation Apache WSS4J)

- **WS-Trust:** Issue and exchange of security tokens in the presence of a Trust Framework
- **WS-SecureConversation:** Establish security context.
- **WS-Federation:** Different security realms federated by allowing and brokering trust of identities, attributes, authentication
- **WS-Policy:** Syntax and general model to define policies for service providers and requestors.

PART 2: ONTOLOGY

The prototype developed for SIMDAT phase I, did not use any ontology technology explicitly. Some of the reasons for it:

First semantic discussions and consensus amongst the meteorology domain experts is always a long term and active process.

Second, the choice of a community standard as the basis of a potential meteorological ontology will in the long term benefit the interoperability and usability but meanwhile, the meto activity has implemented the prototype with the very first versions of the outputs coming of a standardization process.

To describe datasets within the VMC, a standard called WMO Core Profile is being defined by the meteorological community. This profile is an extension of the ISO 19115 metadata standard which is a UML based standard describing geo-referenced data. The WMO Core Profile is at the heart of the VMC prototype and is used to encode the metadata elements describing all its datasets. Members of the SIMDAT meteo activity are actively participating to the work done by the group defining the WMO Core Profile and are influencing the development of the future standard for meteorological metadata. A new version should be released in summer 2006. Modeling of the WMO Core Profile using an Ontology language (OWL/RDF) could help in the definition of the metadata standard. Reasoning engines based on this Ontology could then be used to build the semantic layer of the VMC (discovery/browsing).

Discussion with potential users of the VMC had revealed that there is a need to have a search for synonyms in particular there are special meteorological terms like precipitation which have to be related to every day words like rain and snow, in order that the user of the VMC will not miss data sets when searching the system. Further there should be a multilingual support because there will be end users that are only able to search the system in their native language.

1 ONTOLOGY TECHNOLOGY REQUIREMENTS

1.1 Multilingual Thesaurus

Meteorological data will be demanded by highly different groups of users, like scientists, engineers, agriculturists and fire-fighters. In general these groups are accustomed to different terminologies, they use partly different terms for the same concept. In order to provide to all these data consumers a simple access to the meteorological data available from the VMC the simple keyword search should be replaced by a search function for concepts, by including sets of synonyms. Further the data should be classified into different taxonomies, so a user can browse the catalogue with the classification scheme she is accustomed to.

For that reason a thesaurus should be developed and provided to the VMC. Because a living terminology has to be open for exchanges and modifications a service has to be installed to extend and modify the thesaurus of the VMC by authorized people continuously. The thesaurus has to include the controlled vocabulary of the WMO Core Profile.

Moreover the thesaurus should be multilingual, so a user may search meteorological terms in her native language, in particular in English, French and German. Also the classifications should be available with labels in different languages.

1.2 GT observations ontology

For the dissemination of meteorological data by the Global-Telecommunication-System (GTS) metadata - type, location, data format and creation time - are transformed into a very squeezed code consisting of few letters and numbers in the header of a GTS record. The coding rules are defined in GTS-manuals published by the WMO. Behind these rules there is a classification schema for meteorological messages, using the concepts of types, formats, locations, times and durations. These classification schemas are well-known to senior meteorological experts.

The extension of the ISO 19115 standard that is called the WMO Core Profile is currently under construction and development, there are still a lot of open questions in particular for different message types that are handled by the GTS. There is the hope that making the rules of the GTS-manuals explicit by representing it in a ontology will support the definition process of the WMO Core Profile for observational data transported in the GTS.

A further reason for a "GTS-message-ontology" is to provide to the senior meteorologist the classification and terms they know and to integrate it into the metadata catalogue.

Name	Thesaurus Maintenance Service
Activity/Prototype	Meteo / Demonstrator
Creation Date	2006-03-23 by Helmut Ackermann
Technology	Ontology
Impl. Date	2006-06-30
Priority	High

Description
<p>A web service for extending and modifying a multilingual meteorological thesaurus has to be implemented, installed and provided. The service should provide a Webportal for authorized editors. The portal should offer functions for searching, edit new resources, modify resources, add labels in different languages and add annotations.</p> <p>There should be a download function for the current version of the thesaurus over a http-channel in a RDF/OWL representation.</p> <p>It is suggested to use classes and properties of the W3C SKOS draft for structuring the thesaurus.</p>

Name	Thesaurus extended knowledge discovery
Activity/Prototype	Meteo / Demonstrator
Creation Date	2006-03-23 by Helmut Ackermann
Technology	Ontology
Impl. Date	2006-09-30
Priority	High
Description	<p>Information of the thesaurus has to be included into the catalogue node. So the keyword search will be extended by a search function for related synonyms and the corresponding labels in different languages. Further the catalogue should incorporate a multi-hierarchy, which will be populated by the classifications defined in the thesaurus.</p>

Name	Metadata Editor with controlled vocabulary
Activity/Prototype	Meteo / Demonstrator
Creation Date	2006-03-23 by Helmut Ackermann
Technology	Ontology
Impl. Date	2006-11-30
Priority	Medium
Description	<p>Integrate the thesaurus into the metadata editor, which is an interactive tool to create and maintain the WMO Core Profile metadata of data repositories at the site of the data providers. The integration of the thesaurus should support the use of the controlled vocabulary by presenting proper selection lists.</p>

Name	GTS message ontology
Activity/Prototype	Meteo / Demonstrator
Creation Date	2006-03-23 by Helmut Ackermann
Technology	Ontology
Impl. Date	2006-06-30
Priority	Low
Description	<p>Create an ontology that makes the semantic of the GTS manuals explicit in representing the ontology in a RDF/OWL file and provide tools to browse and search in the ontology</p>

2 ONTOLOGY USE CASES

The following use cases describe the integration of a multilingual thesaurus and a GTS message ontology. Section 3.1.4 “Ontology” has been replaced from its original description D18.1.1 ⁽⁴⁾

In these use cases a user can be a human being, a group of people or a system. A user can have different roles depending on its interactions with the VMC:

- Meteorological data consumer (i.e. scientist, engineer, agriculturists, etc.)
- Metadata provider (i.e. scientist , meteorological expert)
- Editor (i.e. language translator, terminology expert)
- WMO Core Profile designer (i.e. scientist , meteorological expert)

Name	Multilingual Search by Synonyms
UseCaseId	UC_onto_01
Date Created	23/03/06
Last Updated by	Helmut Ackermann DWD
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	Each meteorological data consumer should be able to search the VMC product catalogue by asking for meteorological terms in his native language together with other search constraints like locations and temporal intervals. The VMC should then be able to present a selectable result list which all meteorological products satisfying the constraints and which are related to the terms given. In the beginning the supported languages are restricted to English, French and German.
Actors	- Data Consumer
Pre-Conditions	
Post-Conditions	

Name	Multilingual Browsing
UseCaseId	UC_onto_02
Date Created	23/03/06
Last Updated by	Helmut Ackermann DWD
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	Each meteorological data consumer should be able to restrict his search for products in the VMC data catalogue by browsing a thesaurus that is

structured in way to which he is accustomed to in his application area. The labels for the hierarchy node has to be chosen by the language the data consumer has chosen in his session profile
Actors
- Data Consumer
Pre-Conditions
Post-Conditions

Name	Thesaurus Maintenance
UseCaseId	UC_onto_03
Date Created	23/03/06
Last Updated by	Helmut Ackermann DWD
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
	The meteorological thesaurus have to be extended when new type of data repositories become available or new user groups with a different terminology and classification should have access. The same will be true if new native languages have to be added. Further an editor needs to have access when labels, resources or properties have to be corrected.
Actors	
	- Editor, Metadata or Data Provider
Pre-Conditions	
	- consistent thesaurus
Post-Conditions	
	- consistent thesaurus

Name	Metadata Editor Support
UseCaseId	UC_onto_04
Date Created	23/03/06
Last Updated by	Helmut Ackermann DWD
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
	A metadata provider has to define, extend, modify and correct the metadata descriptions he is responsible for. These descriptions have to conform to the WMO Core Profile for that he will use a special metadata editor that ensures that at the end valid metadata records will be created. To support the use of the controlled vocabulary the thesaurus will be integrated and the selection list of the metadata edit will be populated by proper terms from the thesaurus.
Actors	
	- Metadata Provider
Pre-Conditions	

Post-Conditions

Name	WMO Core Profile Designer Support
UseCaseId	UC_onto_05
Date Created	23/03/06
Last Updated by	Helmut Ackermann DWD
Project phase	WP19.1.2
Parent Package	Meteorology Activity
Use Case Goal	
<p>WMO Core Profile designer can browse the GTS ontology to get a better understanding of the terminology and categories used in the current distribution for daily meteorological messages. The GTS ontology should be generated by a reasoner engine from the semantic experts based on the rules defined within the WMO GTS manuals.</p>	
Actors	
<ul style="list-style-type: none"> - WMO Core Profile designer 	
Pre-Conditions	
Post-Conditions	

3 ONTOLOGY COMPONENTS VIEW

The following components have to be provided to extend the VMC with an extended semantic search and browse functionality on the base of a multilingual meteorological thesaurus.

Thesaurus Service: This will be a Webservice that provides access to the current version of the thesaurus. In a second step the authentication and authorization service component of the VMC have to be integrated. A catalogue node or a portal from the VMC may have access to download the current thesaurus as a file in order to update their catalogue query engines.

Thesaurus Maintenance Portal: A Web application will provide a user interface for data providers and editors to update the thesaurus. This tool should use components from ontology editor OntoStudio from Ontoprise GmbH.

Metadata Thesaurus Enhancer: This component includes a database with the essential information from the thesaurus and it's relations to the metadata catalogue it extends the metadata search and browse engine in order to perform an efficient search and browse process with synonyms in different languages. At the moment it is not yet decided, if this component will be integrated into the catalogue node or the VMC portal.

Metadata Editor: This editor supports metadata providers to create and maintain valid metadata. The editor will use the information available in the thesaurus to create valid metadata which conforms to the established controlled vocabulary.

GTS ontology Viewer: A web portal or a client application that should allow browsing the GTS ontology, probably just an OntoStudio variant.

The following component is a helper tool in the phase of development

Glossary Importer: This tool should generate, from an established hypertext manual, a first meteorological thesaurus.

Figure 6 shows the dependency of the different ontology components. The Thesaurus Service is the central component, it manages the ontology database. The Metadata Thesaurus Enhancer and the Metadata Editor import the ontology data which they need from the Thesaurus Service.

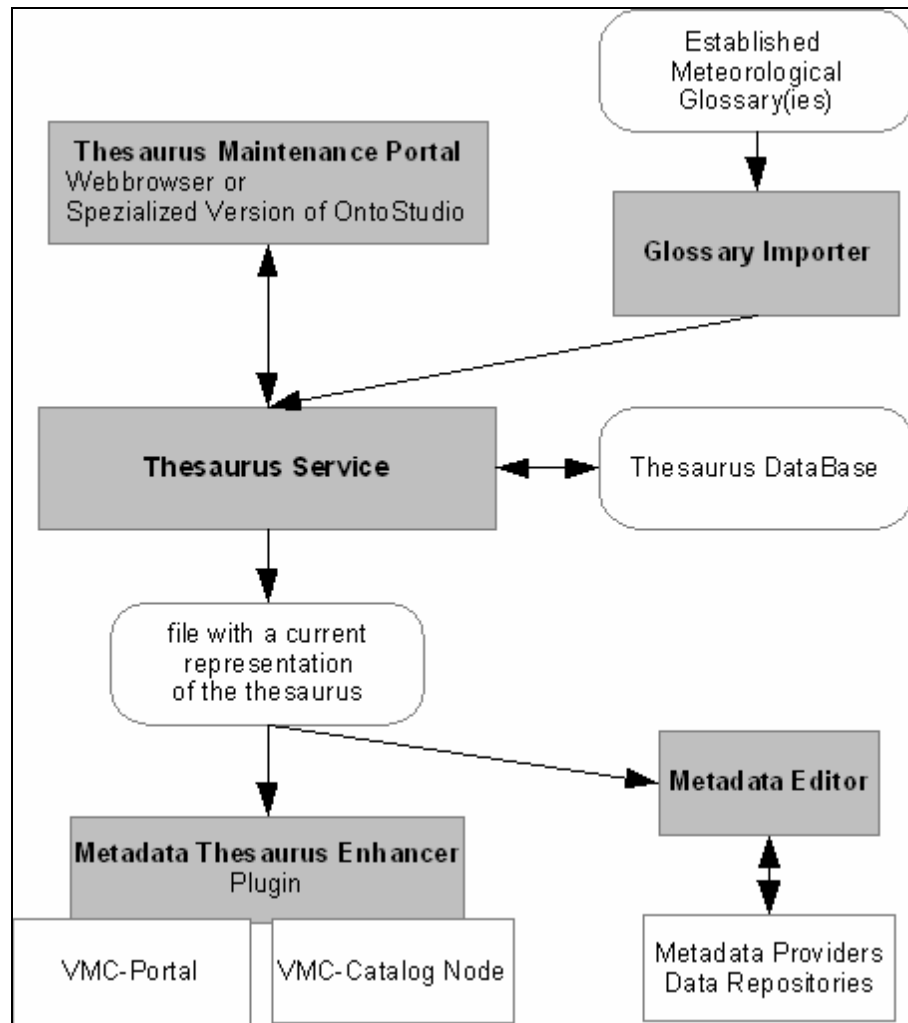


Figure 6 Ontology Components View

4 ONTOLOGY TECHNOLOGY ADOPTION

A first workshop with a semantic expert of Ontoprise GmbH took place on 2006-03-01 in Karlsruhe. There, the technologies which Ontoprise can provide to build and integrate a thesaurus for an improved knowledge discovery were discussed like other applications areas of semantic technologies in the range of the Meteo activity.

We have already started with an evaluation of ontology languages and their different representations. Especially, the ontologies for representing multilingual thesauri will be considered. We had already a look at WordNet⁽¹²⁾ and its multilingual extension EuroNet⁽¹³⁾. Particularly promising for our application will be the W3C “Simple Knowledge Organisation System” (SKOS)⁽¹⁴⁾ which is in the state of a draft.

Further we will evaluate established glossaries of meteorological terms in English, French and German, in particular we will look at the glossary from the American Meteorological Society that is available in hypertext, because we intend to generate from that a first core thesaurus that should be corrected and later continuously extended with the help of the Thesaurus Maintenance Service.

Beside that we will also evaluate ontologies that are developed and used by other earth science portals, like the set of SWEET⁽¹⁵⁾ ontologies, the Climate and Forecast ontology from the LEADS⁽¹⁶⁾ project and the three years old incomplete representation of the ISO 19115 in a RDF file. Also we will investigate open source tools for viewing, editing and searching RDF/OWL as well as the JENA⁽¹⁷⁾ framework for handling RDF files with Java. In particular we will study the Editor OntoStudio from Ontoprise GmbH and how we can use it as interface to maintain our thesaurus.

The Ontobroker from Ontoprise GmbH will be evaluated as a tool to extract the proper information for integrating the thesaurus into the searching and browsing functionalities of the VMC.

An application has to be written to create a first test thesaurus from the AMS glossary⁽¹⁸⁾. A first simple test thesaurus has to be written as soon as possible to become acquainted with the structure. Then a workshop with the experts of Ontoprise will be arranged to discuss how to build the thesaurus maintenance service, to check the consistency of the thesaurus and to design the integration of the thesaurus enhancement into the VMC portal or catalogue node so an efficient search will be possible.

For the GTS message ontology, we will introduce an expert of Ontoprise GmbH into the meaning and the rules that are defined in the WMO GTS manual⁽¹⁹⁾. We hope that then she can extract formal rules from that manual which she can feed into Ontobroker to generate the GTS ontology automatically which should afterwards exported in a RDF description and made available on a Webservice for browsing it with Ontostudio. This work will in addition be interesting to demonstrate semantic technology like

reasoner in relation to meteorological applications and help evaluate semantic query technology which may become of interest after phase II.

CONCLUSIONS

The analysis and implementation of ontology and security services in the second phase of the SIMDAT meteo activity will provide the existing prototype with enhanced services and a closer representation of the meteo scenarios.

- Implementation of security services will enable the VMC to expose a wider range of meteorological datasets that require a strict access control.
- The security services will address the exploitation of the system with real users providing navigation through all the VO in a seamless, easy and secured way.
- Ontology services will provide the end users with an enhanced discovery experience based on richer relationships established between the meteorological descriptions.
- Multilingual Thesaurus addressing internationalization issues, standard definition of meteorological concepts and synonyms relationships will be a key factor for the description of meteorological datasets.

REFERENCES

- ⁽¹⁾ VMC Partners: DWD, Météo-France, UKMO, ECMWF, EUMETSAT
- ⁽²⁾ Camarinha-Matos L.M., Afsarmanesh H, Ollus M., "Virtual Organisations Systems and Practices", Springer Science + Business Media 2005
- ⁽³⁾ RMDCN: <http://www.ecmwf.int/services/computing/rmdcn/>
- ⁽⁴⁾ Consolidated Meteorology Requirements (D18.1.1)
- ⁽⁵⁾ Virtual Organisation Best Practice/ Template (D.4.1.2)
- ⁽⁶⁾ <http://www.gria.org/index.html>
- ⁽⁷⁾ www.eu-trustcom.com
- ⁽⁸⁾ <http://www.globus.org/toolkit/docs/4.0/security>
- ⁽⁹⁾ <http://gridshib.globus.org/>
- ⁽¹⁰⁾ <http://shibboleth.internet2.edu/>
- ⁽¹¹⁾ http://ndg.badc.rl.ac.uk/public_docs/AHM_2004_BNL.pdf

(12) <http://wordnet.princeton.edu/>

(13) <http://www.illc.uva.nl>

(14) <http://www.w3.org/2004/02/skos>

(15) <http://sweet.jpl.nasa.gov/EnablingFinal.doc>

(16) <http://ams.confex.com/ams/pdfpapers/102272.pdf>

(17) <http://jena.sourceforge.net/>

(18) <http://amsglossary.allenpress.com/glossary>

(19)

http://www.wmo.ch/web/www/ois/Operational_Information/WMO386/ManOnGTS.html

GLOSSARY

Term	Definition
AuthN	Authentication Service
AuthZ	Authorization Service
DN	Distinguish Name
VMC	Virtual Meteorological Centre
RMDCN	Regional Meteorological Data Communication Network
VMC partners	DWD, Météo-France, UKMO, ECMWF, EUMETSAT
DCI	Data Communication Infrastructure
DCL	Database Communication Layer
WSRF	Web Service Resource Framework
SOA	Service Oriented Architecture
GSI	Grid Security Infrastructure
NMI	NSF Middleware Initiative
WMO	World Meteorological Organization
GTS	Global Telecommunication System