



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

***D20.1.4 SIMDAT Meteorological scenario:
Evaluation/Validation report
D 20.1.2 Software Design of SIMDAT Meteorological
Scenario***

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
06.03.06	0.1	Guillaume Aubert	Initial version
17.03.06	0.5	Guillaume Aubert	Add Matteo's Comments
22.03.06	0.7	Guillaume Aubert	Add additional Comments
23.03.06	1.0	Guillaume Aubert	Add Helmut's Ontology Lessons Learned
23.03.06	1.1	Matteo Dell'Acqua	Approved
03.04.06	1.2	Guillaume Aubert	Add Mike Turner and TMB's comments
07.04.06	1.3	Guillaume	Add Moritz Weiten comments

Copyright

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

Table of contents

1	Introduction	4
2	PM 12 Prototype.....	4
3	Lessons Learned and Achievements	5
3.1	Architecture	5
3.2	Grid Technologies	6
3.3	Distributed Data Access	7
3.4	Virtual Organisation.....	10
3.5	Ontology/Metadata.....	10
4	SIMDAT Phase II Meteorology Activity Scenario.....	11
4.1	Synchronization Engine Enhancement.....	11
4.2	Network Management Component	12
4.3	WSRF Grid migration	12
4.4	Metadata Manager Enhancement	13
4.5	Virtual Organization.....	13
4.6	Ontology.....	14
4.7	Analysis Services	15
5	Requirements.....	16
5.1	Grid Requirements	16
5.2	DDA Requirements	17
5.3	VO Requirements	19
5.4	Ontology Requirements.....	20
5.5	Analysis Services Requirements	21
6	Conclusion.....	22
7	References	22

1 Introduction

The SIMDAT project is now 18 months-old and it is time to reflect on the pros and cons within the Meteorology activity and evaluate what has been built during SIMDAT phase I. The aim of this document is to list the achievements and lessons learned from the prototype in each of the technologies defined by the project, validate the main parts of the prototype and consequently define the scenario for SIMDAT phase II and the associated requirements.

This document covers 2 deliverables:

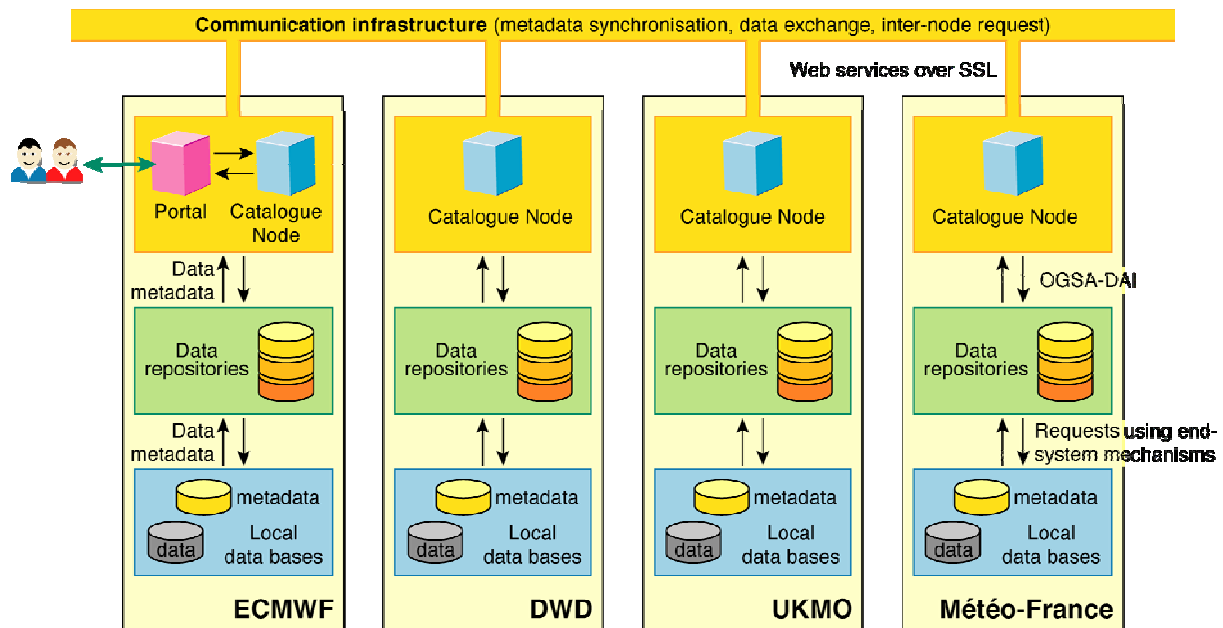
- *D20.1.4 SIMDAT Meteorological scenario: Evaluation/Validation* report where the PM12 prototype is described and the lessons learned presented.
- *D 20.1.2 Software Design of SIMDAT Meteorological Scenario* detailing the scenario for SIMDAT phase II and presenting the associated requirements.

2 PM 12 Prototype

The Virtual Meteorological Centre (VMC) prototype built by the meteorology activity is a virtual database offering discovery and data access services to the users in order to access scientific datasets.

Three main components, the Data Repository, the Catalogue Node and the Portal were defined in order to build the virtual meteorological centre. The architecture is shown below in figure 1. These components are installed on each of the partner sites and are interconnected through a secure communication infrastructure.

- The Data repository is the interface with the local databases and offers the necessary information to describe search and locate the data. It also offers a standard interface to retrieve the data from the associated local databases.
- The Catalogue Node maintains the catalogue of all the data available within the system and ensures the synchronization of the metadata. It harvests metadata from the Data Repository, requests data and serves its two clients: the portal or other Catalogue Nodes.
- The portal offers the interface to search and browse the catalogue of the Virtual Meteorological Centre.



1. VMC layered Architecture

3 Lessons Learned and Achievements

Below is given an overview of the achievement and lessons learned of the SIMDAT Phase I.

3.1 Architecture

- The architecture designed during the SIMDAT phase I (described above) is fully decentralised and based on a distributed database where each node has a copy of the synchronized catalogue. Therefore it is possible to build a reliable distributed meteorological centre where a user can discover, retrieve and collect datasets from any Catalogue Nodes. If a Catalogue Node is temporarily unavailable, its users can be redirected to another one. The main drawback for this architecture is that it is more difficult to handle than a fully centralized system with a unique catalogue and requires having a Virtual Organization with well-defined business rules to manage and operate the infrastructure.
- The architecture follows the Service Oriented Architecture and is ready to integrate analysis services where the distributed database is the first and core service.
- The architecture is extensible, as it is very easy to add new Catalogue Nodes and new data sources (Data Repository)
- The architecture is flexible as it offers different kind of deployment. For the demonstrator phase I, the Catalogue Nodes were deployed on the network DMZ and the Data Repositories on the LAN but it would be very easy to have a Catalogue Node accessing remote Data repositories installed on another site.
- The architecture is scalable, since the current implementation can integrate Data Repositories exposing many more datasets to the users.
- One of the architectural choices was to rely on a synchronization engine to build the virtual database. This solution adds complexity and must be handled carefully to avoid metadata inconsistency.

The architecture provides a solution for the meteorology (meteo) Activity problem: The Weather Information System (WIS) can be based on it. The current architecture will continue to be used in SIMDAT Phase II for the meteorology activity and most of the tasks planned are improvements or additional features of the architecture.

Architecture Lessons Learned
The Architecture provides a Solution for the meteo activity problem <ul style="list-style-type: none"> - It is decentralized and has no central weak point - It is flexible, scalable and extensible - It is a SOA ready to integrate analysis services using the distributed database
The decentralized quality of the architecture adds complexity and the VO defined in phase II will have to well-define business rules to operate the system
The Catalogue synchronization adds complexity and must be handled carefully

3.2 Grid Technologies

This section summarises the feedback and thoughts gathered from the meteorological activity regarding Grid Technologies.

- 2005 was a transition year for the Grid technologies as the main Grid frameworks (Globus Toolkit, etc) were migrating from OGSi (Open Grid Services Infrastructure) to WSRF (Web Service Resource Framework). As a consequence, none of the new frameworks were mature enough to be used in production and the meteo activity was not willing to invest in technologies production-ready (OGSI compliant) but programmed to disappear in a near future. When possible flexible and versatile technologies like OGSA-DAI have been used in the demonstrator. Since the release of frameworks like the Globus Toolkit 4.0.1 or Apache WSRF 1.1 (end of 2005), WSRF can now be considered as a beneficial technology with some interesting prospects for the meteo activity. The Globus toolkit 4 has now a large community, is widely used (OGSA-DAI, China Grid, E-Science community in USA, etc) and offers lots of services in term of security, data delivery, on top of the core WSRF services.
- The current demonstrator has three different types of interfaces: RMI interfaces have been built to offer discovery and data access from the Catalogue Node, Web-Services interfaces have been defined for the Inter-Node request forwarding and synchronization mechanisms and OGSA-DAI interfaces have been used for the Data-Repositories. One of the main tasks of SIMDAT phase II is to unify these interfaces by implementing them using WSRF. These new interfaces will provide better chances to be interoperable within the meteorological community.
- OGSA-DAI used during phase I has been perceived as a very promising framework, well-built and simple to use. Some issues like the non-support of the concurrency have now been solved. This is why OGSA-DAI will continue to be used as a way to access the data repositories during SIMDAT phase II.

Grid Technologies Lessons Learned
2005 was a transition year but in 2006 WSRF seems to be ready with emerging robust implementations like GTK4
WSRF can now be considered as a beneficial technology with some interesting prospects for the Meteo Activity (especially GTK 4.0.1)
The Catalogue Node interfaces needs to be unified and implemented using WSRF in order to provide interoperability with the other partners of our community
OGSA-DAI is a promising technology and can be used to access meteorological Data Repositories. The Data Repositories need to be migrated from OGSA-DAI OGSi to OGSA-DAI WSRF

3.3 Distributed Data Access

The main purpose of the Meteo Activity is to build a virtual distributed database from which a user will be able to discover access and collect scientific datasets. The database is built on top of a Catalogue Service storing metadata describing the scientific datasets and offering discovery services like full-text, time-based, location-based searches and directory browsing facility.

- The distributed database is generic and could be used to publish, discover and access any kind of datasets

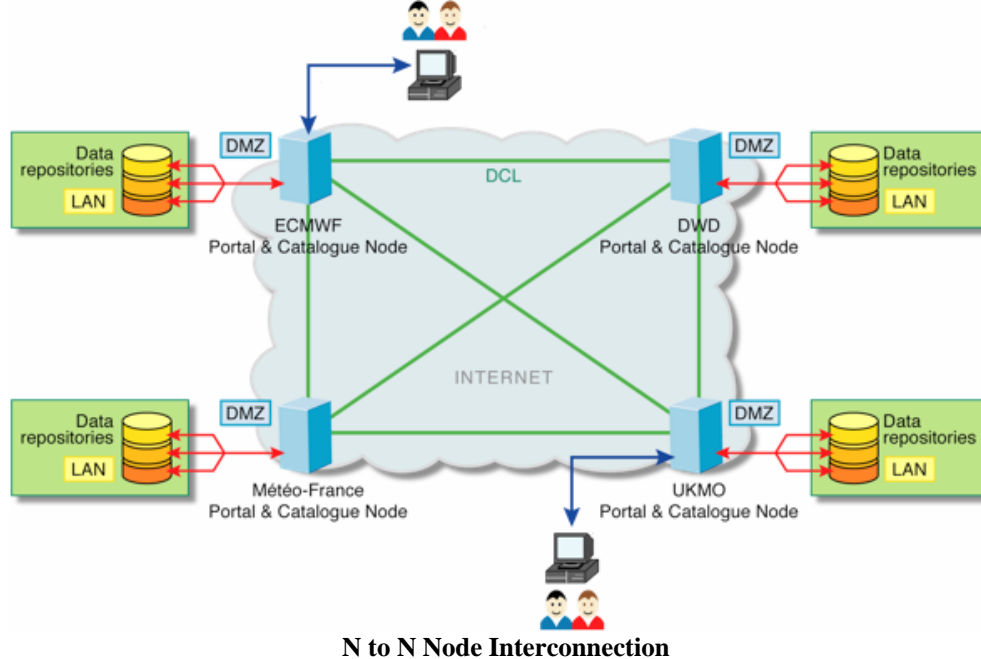
The Catalogue service (also called metadata manager) is a generic data registry implemented within a RDBMS MYSQL 4.1.

- The implementation needs to be standardized to be easily exportable in another RDBMS (postgresql for example) and this could be achieved by using an object to relational mapping tool.
- The discovery part of the Catalogue service needs also to be enhanced: the full-text search should not rely on MYSQL specificities and the browsing discovery needs to be extended. Indeed, it would be interesting to present to the users different ways to browse the catalogue by navigating through different datasets classifications depending on the user's profile (if the user is a climatologist, a forecaster, etc). For the demonstrator one classification has been implemented: the directory classification.
- The data registry implemented by the meteorology activity of SIMDAT has been built using proprietary technologies and XML registry standards like ebXML should be investigated to facilitate the catalogue exchanges within the meteorology community

A synchronization service is used to have the same catalogue on each node and offers a unique view of the available datasets to the user. The synchronization service uses an XML protocol called OAI-PMH derived from the Open Archive Initiative (OAI <http://www.openarchives.org/>).

- OAI-PMH is suitable to synchronize catalogue with static contents but lacks information and is not designed to synchronize catalogues with a fairly dynamic content as used in the VMC.
- The synchronization protocol needs to evolve to avoid metadata inconsistencies and enhance performances. The protocol has to be transformed to be a journalized system using version numbers and not time stamps and extra information is needed to handle a very dynamic turn over of metadata (real-time metadata is completely changed within 3 days in the Virtual Meteorological Centre).

- The synchronization engine implementation forces a node to be directly inter-connected with the other nodes in order to get a complete catalogue. In a near future, the network will be constituted by at least 50 nodes and the current implementation is not suitable. The synchronization engine needs to be implemented in such a way that a node can get from a unique node a complete image of the catalogue. The network of Catalogue nodes used should be a mesh network which is self-organising and where a node can be dynamically inserted without modifying the configuration of each nodes.



- The synchronization engine needs a push (notification) interface to distribute metadata updates with a high priority. Indeed information such as severe weather conditions warnings will have to be spread very quickly over the network of nodes.

Metadata harvesting and data access mechanisms have also been defined in the Data Repository to provide a universal way for the database service to access the different meteorological databases.

- The Data Repository is a universal way to access any kind of meteorological data sources.
- The implementation needs to be improved and a zero-development Data Repository needs to be created to be easily used by partners who don't have java development knowledge.

DDA Lessons Learned
The distributed database is generic solution that could be used to publish, discover and access any kind of datasets
The discovery part of Catalogue Service needs to be enhanced
The Catalogue Service has to migrate to standard technologies to facilitate the catalogue exchange within the meteo community and being interoperable
The Synchronization Engine must be enhanced: <ul style="list-style-type: none"> - Protocol enhancements to improve performances and avoid catalogue inconsistencies - Implementation should be modified to build a mesh network of nodes where a new node can get a complete copy of the catalogue from at least one node
The Data Repository is a universal interface to access local meteorological data sources

3.4 Virtual Organisation

This part collects the lessons learned regarding the security aspects of the Virtual Organisation (VO) as this was the only part of the VO approach during phase I.

The Security for the prototype was sufficient but minimal and lacked consistency. It was mainly implemented at the transport level using SSL:

- X509 certificates coming from the ECMWF CA have been delivered to the users to access the portal in order to establish an HTTPS connection,
- The Portal and the Node have a X509 Certificate and communicate in RMI over SSL,
- The Node-to-Node connection is SOAP or REST over SSL. The connection is established using the nodes certificates.

The Data transfers are protected by a ticketing mechanism where a ticket is delivered by the Catalogue Node and needs to be showed to be authorized to perform the download. The ticket is invalidated after a short period of time.

There was no authorization model to provide a fine-grained access to the datasets.

- One of the main tasks of phase II is to design and build a global security model. A complete Authentication service with a distributed registration authority registering users and delivering certificates has to be built. An authorisation service will be also designed and integrated in current architecture to provide a fine-grained access control to the datasets depending on the user's profile (or Role).

VO Lessons Learned
The security for the prototype was minimal and lack consistency
The prototype in phase I did not provide authorisation mechanisms beyond the data repositories.

3.5 Ontology/Metadata

The prototype developed for SIMDAT phase I did not use any ontology technology explicitly. Discussions between the semantic and meteorology experts had revealed that the creation of an ontology leading to the definition of a universal meteorological request language would be a very long and difficult process. Each meteorology centre has its own way to describe meteorology phenomenon and building mapping software with Ontology technologies would force this software to support complex mathematical operation modifying the data like grid interpolation, unit's conversion. For example the soil temperature can be measured one or two meters over the floor depending of the centres and would force the retrieval tool to transform on the fly the datasets in order to convert the temperature in the chosen type.

Regarding metadata, a standard called WMO Core Profile is being defined by the meteorological community. Modelling 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.

To have more information please refer to [3] *D19.1.2 Report on consolidated technology prototypes for virtual organisations and ontologies*.

Ontology Lessons Learned
The development of an Ontology leading to the definition of a universal meteorological language has not been necessary.
The project supports the development of a meteorological metadata standard within WMO
The knowledge discovery should be enhanced by a multilingual thesaurus.

4 SIMDAT Phase II Meteorology Activity Scenario

Using the lesson learned of the phase I as a starting point the following scenario has been defined.

The SIMDAT Phase II scenario will be based on the phase I architecture which will be consolidated and enhanced in order to solve its present weaknesses:

- Catalogue nodes will be part of a self-organizing meshed network and the synchronization engine is going to be improved to handle content with a short life-time,
- The security services of the VMC VO will be implemented and integrated with the other services,
- A semantic layer improving the discovery service will be built.

The following tasks and milestones have been defined

4.1 Synchronization Engine Enhancement

The VMC (Virtual Meteorological Centre) built by the Meteorology activity is a distributed information system based on metadata catalogues nodes and data repositories holding the actual meteorological datasets (observations, outputs models, etc). The metadata is updated when new datasets are entered in to the system. In order to increase the performance and have a responsive discovery, metadata is synchronized between the nodes and stored in a catalogue on each node. Some of the Nodes are responsible for collecting real-time observations coming from the WMO GTS (Global Telecommunication System) and have to synchronize associated real-time metadata. This metadata has a short life time (up to 3 days) and consequently, a big portion of the metadata catalogue will have to be wiped out and replaced every three days. The synchronization protocol used during the phase I based on OAI-PMH is suitable to synchronize catalogues with static content but lacks information and is not designed to synchronize catalogues with a fairly dynamic content as used in the VMC.

The Synchronization engine needs to evolve to be decoupled from the component managing the elements to synchronize (the metadata manager). Indeed the synchronization engine should be generic to build the synchronized catalogue of meteorological metadata but also to replicate the GTS real-time data among some nodes in order to create a resilient system.

For these reasons, the synchronization engine has to be improved and transformed into a journalized system. Techniques and principles used in synchronization components like journalized file systems, source version control systems like CVS, Subversion will be investigated. SyncML (Synchronization Markup Language) developed by the Open Mobile

Alliance in order to synchronize devices such as mobile phones and pdas with computers will be also studied.

To have more information on the techniques and protocol used, please refer to the VMC synchronization internal document (*[1] VMC Synchronization Design*).

ECMWF is responsible for this module.

A first version is planned for end of June 06 and a consolidated final version will be release for September 06

4.2 Network Management Component

The current synchronization engine forces to have a fully connected topology in order to get a complete catalogue. The actual solution is unsuitable to build the VMC network that should contain 50 to 100 nodes. In some cases the direct network connection between the nodes will also be impossible for security reasons. For example, some nodes could be accessible to others only through a specific node acting as a bridge between two different networks.

The solution for this problem is to build a Network Management component organizing the network topology. The Network Management component will organize the nodes as a mesh network. Mesh networking is an emerging network organization used by the wireless community and in peer2peer software. It is a way to route data between nodes and it allows for continuous connections and reconfiguration around blocked paths by "hopping" from node to node until a connection can be established. Mesh networks are self-healing: the network can still operate even when a node breaks down or a connection goes bad. As a result, a very reliable network is formed.

To have more information on the mesh network component built, please refer to the Dynamic node insertion and message routing document (*[2] Mesh Network Management*).

INTEL is responsible for designing and implementing this module taking into account the requirements of the meteorology activity

A first version is planned June 06 and a consolidated final version will be release for month September 06.

The synchronization engine will rely on the network management software to routes its synchronization messages.

4.3 WSRF Grid migration

As detailed in the lessons learned, the demonstrator phase I has three different interfaces causing difficulties to be interoperable in the future within the meteorology community. Considering the actual state of the Grid technologies WSRF is now seen as a first step to build interoperability.

Globus Toolkit 4.0.1 (GTK), which is the reference implementation for WSRF seems to be mature and offers a consistent set of services that could be used as a base to solve some of the problems of the meteorology activity. For example GTK 4 has different security components, which could be used to build the VMC Virtual Organization and some of the data transfers modules such as GridFtp, RFT (reliable file transfer) could also be used to efficiently transport data.

One of the main tasks for phase II will be to migrate the interfaces of the Catalogue Node to WSRF using GTK4 core part. Other parts of GTK4 such as the security and data transfer modules will be evaluated to see how their integration in the architecture could benefit the

meteorology activity. Each partners of the meteorology activity will migrate its Data Repositories to the last version based on OGSA-DAI WSRF (relying on GTK4).

ECMWF is responsible for migrating the nodes interfaces and lead the GTK4 evaluation. All the meteorology partners (DWD, EUMETSAT, Meteo-France and UKMetOffice) are responsible for migrating their Data-Repository to the new WSRF synchronization and data-requesting interface.

A final version is planned for Month 25 (Oct 06).

4.4 Metadata Manager Enhancement

The Catalogue service is implemented in the Catalogue node within the Metadata Manager. The Metadata Manager needs to be simplified and standardized to be easily exportable in another RDBMS. An object to relational mapping tool such as iBatis (<http://ibatis.apache.org/>) or Hibernate (<http://www.hibernate.org/>) will be used.

The discovery part of the Catalogue offering geographical, temporal and full text search has to be also modified to be totally independent of MySQL. The Browsing offers a way to discover the datasets through a directory structure (classifications). An interesting feature would be to present to the users different classifications depending on the user's profile (forecaster, oceanographer, climatologist). Technologies such as ebXML or the semantic registries could be used to build this enhanced browsing service and will be investigated.

The catalogue service is currently built using proprietary technologies but to facilitate the catalogue exchanges and integration in other scientific catalogue services, standard registry technologies such as ebXML should be investigated. Indeed some of the leading project of the meteorology community working on similar subjects are moving toward these technologies.

ECMWF is responsible for this activity.

A new version of the Catalogue service simplifying the database schema and implementation will be delivered in June 06.

UK Met Office will conduct a study on the ebXML features and how to use it to build the Metadata Manager. A document will be produced.

The introduction of multiple datasets classifications will be investigated.

4.5 Virtual Organization

The security part of the VMC virtual organization will be designed and developed during the phase II of SIMDAT.

The VO will not maintain any centralised information for the security aspects so, user registration will be handled by each of the partners. The VO will provide instead, the mechanism to authenticate a user at its home institution and provide 'proof of identity' to the rest of the partners according to the trust relationships established between the VO members.

Authorisation services will check against access policies (global or local) concerning either the data or services to grant or deny access across the VO. The VO will need to identify the policies required in order to be part of the VMC and establish the levels of trust that the VO will account for.

Service Level Agreements (SLA) will ensure that the system operates to the agreed level and that appropriate actions take place if the SLAs are breached. Under these conditions the

partners (as individual members of the VO) and the VO (as a unified entity of services and resources) should all guarantee that the Quality of Service does not get compromised over the existing operational services.

Security Grid technologies like MyProxy, Proxy certificates, Gribshib (see <http://www.globus.org/>) will be evaluated in order to build the VO Authentication and Authorization service.

ECMWF is responsible for this activity.

A first version of the VO security components will be delivered in Sept 06 and a consolidated final version will be release for March 07.

To have more information please refer to *[3] D19.1.2 Report on consolidated technology prototypes for virtual organisations and ontologies*.

4.6 Ontology

Ontology will be investigated to add a semantic layer to the distributed portal. This will enhance the discovery service of the portal and will allow the VMC to reach a wider audience. For example, in meteorology the term "precipitation" is used for rain, snow or hail. A user searching for "rain" data should find "precipitation" related datasets.

Ontology will be used to build a controlled domain-vocabulary (thesaurus) and to categorize the keywords using synonyms or hierarchies of specialization. Existing ontology such as SWEET or CF will be analysed to see if they are suitable to create a meteorological thesaurus. The thesaurus will be created and extended to have multi-lingual support. The VMC discovery service will be enhanced to make use of the thesaurus and will be implemented in the portal or in the node search engine. The "thesaurus" service could also be used in the metadata generation and maintenance tools in order to control the vocabulary that can be used in the metadata.

Ontology could also be used in another area: the current observation network called GTS has been in placed for more than 30 years and is transporting various and numerous observations. A classification for these observations has been defined during the years but it is very tedious and difficult to understand it as the documentation is network-oriented and more concerned about routing mechanism and messages structures than the observations itself. Starting from this documentation, ontology could be defined and used as a logical model to clearly define the relations between the observations. This logical classification could then be used in the future to have a better understand of the different available observations and generate their associated metadata which will be used in the VMC.

The Deutscher Wetterdienst (DWD) is responsible for this activity.

A first version of the discovery service including the use of the thesaurus will be delivered in Sept 06 and a consolidated final version will be release for Feb 07.

A first version of the Ontology modelling the differents GTS observations will be delivered in Jan 07.

To have more information please refer to *[3] D19.1.2 Report on consolidated technology prototypes for virtual organisations and ontologies*.

4.7 Analysis Services

Most of the meteorological data laying in the databases and archives are “raw”, i.e. they are as they were received or produced. There exist numerous processing tools to perform data transformation such as sub-area extraction, interpolation of global numerical weather prediction system output, re-projections of satellite images or extraction of individual items (e.g. one temperature measurement) from composite observation messages. More elaborate tools are able to combine various data to produce derived products, to visualise or plot those products.

These tools could be provided as Grid services (WSRF services). A survey of the available tools at each partner’s sites will be carried out and a first catalogue of available services will be established. This catalogue will be implemented in a service registry associating the analysis services to different dataset types. A user will have then the possibility to create larger services using workflow tools just by chaining these atomic analysis services.

For the phase II the meteorology activity will produce a design paper explaining how the services will be integrated within the current architecture and associated with the virtual database.

In order to build relevant use-cases the subscription service will be seen as a first service of the Grid.

5 Requirements

Below are described the requirements resulting from the lessons learned and scenario phase II

5.1 Grid Requirements

Summary

- Provide an efficient infrastructure across the 5 partners as basis for higher level services and capable of supporting the wide variety and large volumes of data
- Provide interfaces to achieve interoperability with standards and protocols used by other WIS actors (GISC, DCPC). ***Adopt open standard technologies and implements WSRF interfaces***
- Provide an extendable and scalable infrastructure ***based on a self-organising mesh network*** able to accept new partners
- ***Provide QoS and resources managements*** : Traffic prioritization, request cost evaluation, queuing mechanism
- ***Provide robust and efficient transport*** for Node to Node and Node to Client data exchange
- ***Grid Technologies must offer mechanism to integrate virtual database with analysis services***

Name	Build an Efficient Grid Infrastructure
Activity/Prototype	Meteo / Demonstrator Phase II
Creation Date	16.03.06 by Guillaume Aubert
Technology	Grid
Impl. Date	End of August 06
Priority	high
Description:	Provide an efficient infrastructure across the 5 partners as basis for higher level services and capable of supporting the wide variety and large volumes of data

Name	Adopt Standard Interface
Activity/Prototype	Meteo / Demonstrator Phase II
Creation Date	16.03.06 by Guillaume Aubert
Technology	Grid
Impl. Date	End of October 06
Priority	high
Description:	Provide interfaces to achieve interoperability with standards and protocols used by other WIS actors (GISC, DCPC). <i>Adopt open standard technologies and implements WSRF interfaces</i>

Name	Build a Mesh Network
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	Grid
Impl. Date	End of August 06
Priority	high
Description:	Provide an extendable and scalable infrastructure <i>based on a self-organising mesh network</i> able to accept new partners

Name	Provide Quality of Service
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	Grid
Impl. Date	
Priority	high
Description:	<i>Provide QoS and resources managements</i> : Traffic prioritization, request cost evaluation, queuing mechanism

Name	Robust and efficient Transport
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	Grid
Impl. Date	Phase III
Priority	medium
Description:	<i>Provide robust and efficient transport</i> for Node to Node and Node to Client data exchange <i>Provide mechanism to transport very large data volumes</i>

Name	Accept Analysis Services
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	Grid
Impl. Date	Design for end of Phase II and implementation in Phase III
Priority	high
Description:	<i>Grid Technologies must offer mechanism to integrate virtual database with analysis services</i>

5.2 DDA Requirements

Summary

- *Provide functionalities to search for and find datasets of interest:*
 - o Human to formulate queries directly to catalogue (interactive access)
 - o Machine-to-machine queries (batch access)
- Provide the unified view of all the shared datasets through a *synchronized catalogue*.
Implement a reliable synchronization service that can rely on a mesh network topology

- Provide mechanisms to **replicate and cache** the data (real-time data) across the VMC nodes
- Provide mechanism for a node to **quickly announce** to the other nodes the availability of some special datasets (storm tsunami warnings, etc) like *notification, push mechanisms, QoS, prioritization*
- Provide a uniform Interface with existing meteorological databases (flat file repositories, RDBMS database, off-line archives, XML DB). *Extend the Data-Repository component*
- Provide **efficient and secure** data transfers
- *Provide the ability to easily add new data sources*
- *Provide administration tools for the Virtual Database*
 - o Administration of the existing local database must be independent of the administration of the Virtual Database

Name	Provide discovery/requesting functionalities
Activity/Prototype	Meteo / Demonstrator Phase II
Creation Date	16.03.06 by Guillaume Aubert
Technology	DDA
Impl. Date	
Priority	high
Description:	<i>Provide functionalities to search for and find datasets of interest:</i> <ul style="list-style-type: none"> - Human to formulate queries directly to catalogue (interactive access) - Machine-to-machine queries (batch access)

Name	Implement catalogue synchronization
Activity/Prototype	Meteo / Demonstrator Phase II
Creation Date	16.03.06 by Guillaume Aubert
Technology	DDA
Impl. Date	2006-06-30
Priority	high
Description:	Provide the unified view of all the shared datasets through a <i>synchronized catalogue</i> . <i>Implement a reliable synchronization service that can rely on a mesh network topology</i>

Name	Implement Data Replication/Cache
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	DDA
Impl. Date	First version end of Phase II, final version Phase III
Priority	high
Description:	Provide mechanisms to replicate and cache the data (real-time data) across the VMC nodes

Name	Provide a Universal Interface for Meteorological Databases
Activity/Prototype	Meteo / Demonstrator

Creation Date	16.03.06 by Guillaume Aubert
Technology	DDA
Plan	Done in Phase I / Enhancement during Phase II
Priority	high
Description:	Provide a uniform Interface with existing meteorological databases (flat file repositories, RDBMS database, off-line archives, XML DB). Extend the Data-Repository component Provide the ability to easily add new data sources

Name	Provide Virtual Database Administration tools
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	DDA
Impl. Date	Phase III
Priority	high
Description:	Provide administration tools for the Virtual Database - Administration of the existing local database must be independent of the administration of the Virtual Database

5.3 VO Requirements

Summary

- Build a VO organized as a **federation of identity** where a user registered from A is recognised at B as a VO member
- Implement a flexible authentication service identifying users and batch client and associating them to VO Roles. The registration of users will be still handled by individual organizations.
- Implement an authorization service providing a **flexible fine-grained** access control mechanism to the datasets. The authorization service will implement a global access control policy managed by each partners and integrated into their existing security policies

Name	Build a Federation of Identity
Activity/Prototype	Meteo / Demonstrator Phase II
Creation Date	16.03.06 by Guillaume Aubert
Technology	VO
Impl. Date	End Of September 06
Priority	high
Description:	Build a VO organized as a federation of identity where a user registered from A is recognised at B as a VO member

Name	Implement a flexible Authentication Service
Activity/Prototype	Meteo / Demonstrator Phase II
Creation Date	16.03.06 by Guillaume Aubert
Technology	VO
Impl. Date	End Of September 06
Priority	high
Description:	- Implement a flexible authentication service identifying users and batch client and associating them to VO Roles. The registration of users will be still handled by individuals organizations.

Name	Implement a flexible Authorization Service
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	VO
Impl. Date	End Of September 06
Priority	high
Description:	Implement an authorization service providing a flexible fine-grained access control mechanism to the datasets. The authorization service will implement a global access control policy managed by each partners and integrated into their existing security policies

5.4 Ontology Requirements

Summary

- **Establish a controlled vocabulary** (Thesaurus) for keywords and categories in the meteorological domain
- Build a **discovery interface** that enables non-meteorological users to query for datasets using their own vocabulary
- Investigate existing Ontology in the context of meteorology like SWEET, CF
- Use Ontology to create a logical model of the available observations travelling on the WMO GTS

Name	Establish a Controlled Vocabulary (Thesaurus)
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	Ontology
Impl. Date	End of September 06
Priority	high
Description:	Establish a controlled vocabulary for keywords and categories in the meteorological domain

Name	Enhance the discovery service using the Thesaurus
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	Ontology
Impl. Date	End of September 06
Priority	high
Description:	Build a discovery interface that enables non-meteorological users to query for datasets using their own vocabulary

Name	Use Ontology to create a logical model of the GTS observations
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	Ontology
Impl. Date	End of Phase II
Priority	low
Description:	Use Ontology to create a logical model of the available observations travelling on the WMO GTS

5.5 Analysis Services Requirements

- Establish a **list of the current meteorology services/tools** eligible to become services on the Grid
- Build a **catalogue of analysis services** classified following the datasets type they can accept as input and the datasets type they produce
- Investigate the integration of analysis services with the VMC architecture

Name	Establish a list of the gridable meteorology services/tools
Activity/Prototype	Meteo / Demonstrator Phase II
Creation Date	16.03.06 by Guillaume Aubert
Technology	Analysis Services
Impl. Date	End of Phase II
Priority	high
Description:	Establish a list of the current meteorology services/tools eligible to become services on the Grid

Name	Build a Catalogue of analysis services
Activity/Prototype	Meteo / Demonstrator Phase II
Creation Date	16.03.06 by Guillaume Aubert
Technology	Analysis Services
Impl. Date	End of Phase II
Priority	high
Description:	Build a catalogue of analysis services classified following the datasets type they can accept as input and the datasets type they produce

Name	Investigate the integration of analysis services within the VMC architecture
Activity/Prototype	Meteo / Demonstrator
Creation Date	16.03.06 by Guillaume Aubert
Technology	Analysis Services
Impl. Date	Design for End of Phase II
Priority	high
Description:	Investigate the integration of analysis services with the VMC architecture

6 Conclusion

- The meteorology activity has delivered with the prototype of the SIMDAT phase I a first version of a framework that provides the pillars for discovery, collection and distribution of meteorological data. The architecture is proving to be a key helping factor in the development and establishment of some of the concepts covered by the WIS (Weather Information System) systems as a future long term infrastructure introduced by the WMO.
- The project has created a lot of interest in the meteorology community and the VMC is now one of the leading WIS project. Collaborations with the Chinese, Japanese and Korean meteorology agencies and the Australian Bureau of Meteorology have now started in order to demonstrate the architecture at a *worldwide level*.
- *The second phase will prove a challenging phase to demonstrate scalability of the architecture in order to include in the VMC, datasets for more than ten different locations.*
- Synchronization, security and semantic discovery will be addressed during phase II and others like associating analysis services with the VMC virtual database will start to be under consideration for integration in phase III

7 References

- [1] VMC Synchronization Design available from the SIMDAT BSCW server
 [2] Mesh Network Management available from the SIMDAT BSCW server
 [3] D19.1.2 Report on consolidated technology prototypes for virtual organisations and ontologies available from the SIMDAT BSCW server