VAMDC

Virtual Atomic and Molecular Data Centre

D5.3

–

Infrastructure Support Report 2

Version 0.2

Grant agreement no: 239108

Combination of Collaborative Projects & Coordination and Support Actions

Project Information

Project acronym: VAMDC

Project full title: Virtual Atomic and Molecular Data Centre

Grant agreement no.: 239108

Funding scheme: Combination of Collaborative Projects & Coordination and Support Actions

Project start date: 01/07/2009

Project duration: 42 months

Call topic: INFRA-2008-1.2.2 Scientific Data Infrastructure

Project web sites: http://www.vamdc.eu

http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/WebHome

Consortium:

Beneficiary Number *	Beneficiary name	Beneficiary short name	Country	Date enter project**	Date exit project**
1(coordinator)	Centre National de la Recherche Scientifique	CNRS	France	Month 1	Month 42
2	The Chancellor, Masters and Scholars of the University of Cambridge	CMSUC	UK	Month 1	Month 42
3	University College London	UCL	UK	Month 1	Month 42
4	Open University	OU	UK	Month 1	Month 42
5	Universitaet Wien	UNIVIE	Austria	Month 1	Month 42
6	Uppsala Universitet	UU	Sweden	Month 1	Month 42
7	Universitaet zu Koeln	KOLN	Germany	Month 1	Month 42
8	Istituto Nazionale di Astrofisica	INAF	Italy	Month 1	Month 42
9	Queen's University Belfast	QUB	UK	Month 1	Month 42
10	Astronomska opservatorija	AOB	Serbia	Month 1	Month 42
11	Institute for Spectroscopy RAS	ISRAN	Russian Federation	Month 1	Month 42
12	Russian Federal Nuclear Centre All-Russian Institute of Technical Physics	RFNC-VNIITF	Russian Federation	Month 1	Month 42
13	Institute of Atmospheric Optics	IAO	Russian Federation	Month 1	Month 42
14	Corporacion Parque Tecnologico de Merida	CTPM	Venezuela	Month 1	Month 42
15	Institute of Astronomy of the Russian Academy of Sciences	INASAN	Russian Federation	Month 1	Month 42

This project is funded under *"Combination of Collaborative Projects and Coordination and Support Actions"* Funding Scheme of The Seventh Framework Program of the European Union

Document

Deliverable number:	D5.3
Deliverable title:	Infrastructure Support Report 2
Due date of deliverable:	June 2011
Actual submission date:	August 2011
Authors:	P. Le Sidaner, K Benson, F. Kosmala, M.L. Dubernet and WP5 team
Work Package no.:	WP5-SA1
Work Package title:	Infrastructure Deployment
Work Package leader:	CNRS
Lead beneficiary:	CNRS
Dissemination level:	PU
Nature:	Report
No of pages (incl. cover):	

Abstract	The objective of D5.3 is to describe VAMDC Infrastructure Deployment Report for Cycle 2. This report corresponds to Activities in WP5: SA1 "Infrastructure Deployment". This report will be included in the VAMDC Periodic Report for Cycle 2.

Versioning and Contribution history

Version	Date	Reason for modification	Modified by
V0.1	**June 2011**	**Compilation of nodes contributions for WP5**	**F. Kosmala**
V0.1	**July 2011**	**WP5 Report for P2**	**P. Le Sidaner**
V0.1	**July 2011**	**Making of D5.3 including 2 report**	**F. Kosmala**
V0.1	**August 2011**	**Check**	**M.L. Dubernet**
V0.2	**August 2011**	**Improvement of P2 content**	**P Le sidaner & K. Benson**
V0.2	**August 2011**	**Final Check**	**M.L. Dubernet**

Final Version (v0.2) released by		Circulated to	
Name	Date	Recipient	Date
M.L. Dubernet	24th August 2011	Mrs Asero	24th August 2011

WP5 ACTIVITIES DESCRIPTION

Work package number	5		**Start date or starting event:**				3				
Work package title	SA 2: Support to the Infrastructure										
Activity Type	OTHER										
Participant id	1	2	3	8	11	12	13	15			
Person-months per beneficiary: **(Total = EU + Node Contributions)**	54	36	24	3	18	5	6	11			

Table of Content

1. WP5 Objectives

SA2 provides support for the delivery of the VAMDC e-infrastructure to users and producers (SA1) . SA2 will be responsible for the maintenance and monitoring of the core infrastructure;

Implementing Grid technology within the VAMDC: providing direct support to the users of the scientific data infrastructure as they enter the VAMDC portal and for the preservation and storage of digital data.

WP5 leader is CNRS (1)

2. WP5 Milestones and Deliverables

Milestones

M5.1	Deployment of Monitoring	WP5	CNRS	Months 10, 22, 34, 42	
M5.2	Deployment of Help Desk	WP5	CNRS	Months 10, 22, 34, 42	
M5.3	Deployment of Grid Operation	WP5	CNRS	Months 10, 22, 34, 42	
M5.4	Deployment of Preservation and QA	WP5	CNRS	Months 10, 22, 34, 42	

Deliverables

D5.1 Service Support Plan (PM 3)
D5.2 Infrastructure Support Report to be included in report to the EU– Year 1 (PM 10)
D5.3 Infrastructure Support Report to be included in report to the EU – Year 2 (PM 22)
D5.4 Infrastructure Support Report to be included in report to the EU – Year 3 (PM 34)
D5.5 Final Report of Service Support to be included in final report to the commission (PM41)

Annual Service Support Plan revisions included in Revised Annual VAMDC Project Plans – Year 1,2,3

3. WP5 Tasks Description

WP5 Leader	P. Le Sidaner / K. Benson CNRS:UMS / UCL :MSSL	
Task Number	Leader	Other Partners
1	A.Shih (CNRS:UMS)	All
2	J. Bureau (CNRS: LPMAA)	All
3	K. Benson (UCL)	All
4	P. Le Sidaner (CNRS: UMS)	All
5	M. Doronin (CNRS: LPMAA)	All

Description of work

Task1: Maintenance and monitoring of the core infrastructure (CNRS(1), all SA2 partners)

The core infrastructure will include partners who maintain existing databases and services. All the actors will be in charge of providing access to the databases/services deployed in SA1. The services include accessing the databases via different protocols, access to dictionaries and publishing registries. Task1 involves setting up the quality assurance of the infrastructure activities, service heartbeats and development and use of unit test packages. The monitoring activities will be implemented at VO-Paris Data Centre using the NAGIOS software. We will need to develop plugins specific to the various protocols which will need validation. Monitoring Software implemented at VOPARIS Data Centre will be distributed to regional centres.

Task 2: Grid Operations (CNRS(1))

The infrastructure includes the possibility to use the GRID technology in order to run numerical codes that produce AM data or that use AM data on hardware provided outside VAMDC. This is separate from and complementary to the execution of codes on hardware provided at VAMDC nodes (the latter facility is part of SA1). Task 2 will make selected codes useable on the grid. The work involves:

- making the codes executable on grid nodes, either by making the codes themselves portable or by packaging them in virtual machines;
- providing grid portals where the codes can be invoked and from which the results can be retrieved;
- negotiating access for VAMDC users with grid providers, especially with EGEE

Task 3: Support to "users" of the infrastructure (UCL(3) with partners (2), (12), (15))

"Users" of the infrastructure, meaning all people interacting with the infrastructure, will need to have access to information concerning the composition of the infrastructure, the services which are available, the procedures about how to enter the infrastructure, the procedures about how to implement the standards, how to use or adapt the various tools. The support to the "users" will be operated in Task 3 through the provision of on-line support materials, a help desk and a service providers/users forum where people could share best operation practice. We intend to produce a self-studying e-tutorial that can be incorporated in university courses on molecular and atomic physics, astronomy, energy systems, environment (etc). Also we intend to operate an e-tool for general public to take a virtual guided tour of VAMDC: statistics, content, geography of clients and producers, databases locations.
Dissemination and Tutorials organized in WP3 will show and teach how to implement and use the infrastructure, will a vertise all those tools. Note that the actual generation of the training materials and support events will be organised by WP3 (NA2).

Task 4: Preservation of digital data and resources (CNRS(1))

The Preservation of digital data and resources is one of the key aspect of sustainability. It is the purpose of SA3 to set up a system of preservation through archiving and mirroring. Some nodes will act as repositories: the nodes already supporting such preservation (nodes linked to VALD, CHIANTI, etc..), VOPARIS Data Centre which will act for most of CNRS resources and could be extended to other partners. The first proposed technology will be to create a virtual machine for a certain number of projects who will implement their resources and we will implement synchronisation. This first step of a mirroring site is the simplest approach and will be implemented during the whole project. During Phase 1 we will work at the EPT level in order to follow preservation activities in other areas. We will adjust our preservation policy accordingly in Phase 2.

Task 5: Quality Assurance of data and resources (CMSUC(2), with partners 3, 11, 12, 15)

Another crucial point is the reliability of the data transferred via the various protocols. The database providers are responsible for the entries in their own database. The usual and slow way of accessing data via classical web interface or via ftp obliges the user to understand the structure of the database, to read instructions in order to get the meaning, definition of columns and lines. An interoperable e-infrastructure will remove some of this verification process of the user. Therefore it is indispensable to check that all resources (core and new ones) use the protocols, standards in the best and reliable way. In Task 5 small groups of VAMDC people understanding the protocols/standards and the physics of the retrieved data will test the output of databases in order to check the good use of protocols, whenever there is a new release handling new cases.

4. WP5 Tasks Plan for Period 2

Period: 01/07/2010 – 30/06/2011
WorkPackage: WP5/SA2 Support to the Infrastructure
WorkPackage Leader and co-Leader: CNRS (P. Le Sidaner),
Participants in the WorkPackage: CNRS, CMSUC, UCL, INAF, ISRAN, RFNC-VNIITF, IAO, INASAN.

Part 1

Objectives and details for each task in Year 2.

Year 2 for the WP5 is dedicated to strengthen the infrastructure and to setup bases for a reliable system.

Task 1 Maintenance and monitoring of the core infrastructure :

1.1 Using the monitoring sytem already in place:
install new services when they become available, using registry information
contact each service to allow them to receive problems alert and have access to the web portal. The policy should be site dependent (number of contacts receiving alert, number of contacts to acknowledge problem alert or to schedule downtime ...)
provide information and help on nagios uses and open account to all relevant VAMDC people
Follow the standards development of data access protocols in order to monitor them properly.
Make development of dedicated plugins for these protocols.

1.2 Promote statistics on reliability of services

Task 2 Grid Operations
2.1 Evolution of Portal :
promote a portal to launch code, make a how to and propose on-line documentation to submit code.

2.2 Promote access to Observatoire de Paris Node:

This service will need different steps and have to be validated:
Access to EGEE certificate, this will be done locally by the local entity. explanation and How-To need to be provided
Access to Astronomy & Astrophysics Vorg, the approval will be made by C. Loomis.
Access to Observatoire de Paris UI machine to submit job. This will be made also in two steps : first getting an account to access SSH gateway of the Observatory then opening an account on the UI machine.

2.3 Possibly, create a VAMDC Vorg – This last step might delayed up till Period 3. All depends upon progresses in Period 2

Task 3 Support to "users" of the infrastructure

3.1 Technical support to users and to service providers will be provided by email. The issue-tracking system RT will be installed at Paris Observatory to manage this email traffic. RT allows incoming messages to be sorted and copied to the appropriate pool of experts for resolution.

3.2 During period 2, groups of experts will be formed to respond to particular types of issues, by consultation with other work packages. Probably areas of expertise are question to developers, technical problem in service deployment, technical end-user problems.

3.3 RT keeps a permanent record of issues raised. From this, a list of frequently-asked questions will be complied during periods 2 and 3.

3.4 Within the project, code will be stored in various repositories: some is already in the public GitHub repository and a Subversion repository is available at the Paris observatory. The Redmine code-tracking product will be installed at the Paris observatory to give a common view of these repositories from a single point on the WWW.

Task 4 Preservation of digital data and resources

- Ensure replication of the 3 other test databases
  - Coordination with Scientific and Technical contact to see modality for replication. Firstly with the 4 candidates where Table Access Protocol are available.
  - Construct virtual machines adapted to the service requirement
  - Open access to make an efficient mirror : construct replication
  - Define with each data provider and scientific responsible for the policy related to updating the mirror.
- Extend it to other new candidates
- Follow the protocols "Data Access Layer" on VAMDC to ensure miror for databases compliant to this protocol.
- In coordination with the person responsible for each database, register the mirror service in Local Astrogrid Registry. Define policy for technical responsibility on mirror service
- Use of tape archive for saving Virtual machine in order to allow a backup recovery in case of major injury.

Task 5 Quality Assurance of data and resources

5.1 Set up technical validation of web-services to check:
- provision of mandatory features;
- provision of optional features;

- correct syntax for request parameters (e.g. the names for query languages and output formats);
- correct syntax of results (e.g. XML outputs must be well-formed and schema-valid);
- correct handling of errors (e.g. error documents in TAP must be VOTables).

These checks detect problems in implementation or deployment and will be run occasionally by WP5 staff.

5.2 Set up monitoring of web-service availability. A service is "available" if it is both running and connected to its resources (e.g. databases), so availability checks imply simple queries to the service. The tests should be made regularly, by an automated system that discovers services from the registry.

5.3 Extract, from the early service-deployments, a list of good practices that reduce the chance of errors in later deployments

5. WP5 Tasks Reports for Period 2

Period: 01/07/2010 – 30/06/2011
WorkPackage: WP5 deployment
WorkPackage Leader and co-Leader: G. Rixon, A. Shih

Participants in the WorkPackage: CNRS, CMSUC, UCL, OU, UNIVIE, UU, KOLN, INAF, RFNC-VNIITF, IAO, IVIC

Part 1

A summary of progress towards objectives and details for each tasks

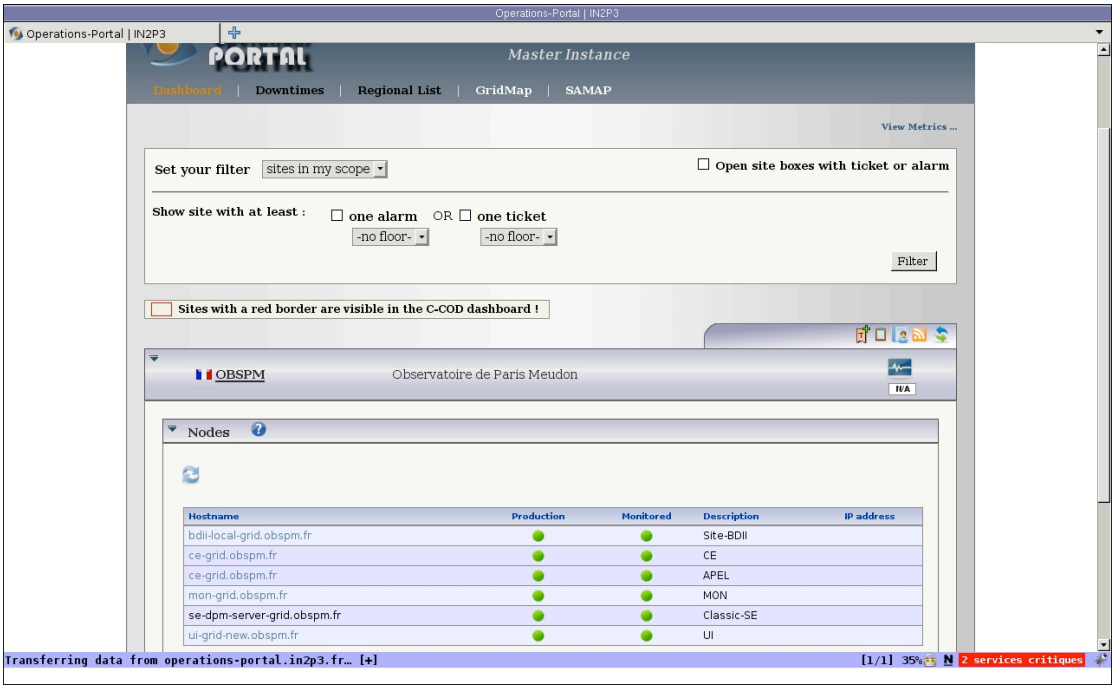
Task 1 On Maintenance and monitoring of the core infrastructure :
All the services are now monitored in Nagios. Services are checked every 5mn on each faillure an email alert is send to the service provider and the responsible for VAMDC infrastructure in charge of Nagios.
Nagios system is on https (authenticated). This mean all service provider have user account and password access to the interface. It allow us to acknowledge any trouble, to specify down time period for maintenance …
Each time a new service appear, this procedure is done.

A document in draft status is written about monitoring uses and access.
Plugin for Nagios have been modify to fulfill the specification of VOSI availability specification.

Statistics are promoted on liability of services by the Nagios Manager.
For all services statistics are available using the web interface with tables like

State	Type / Reason	Time	% Total Time	% Known Time
UP	Unscheduled	262d 14h 19m 34s	71.944%	99.863%
	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**262d 14h 19m 34s**	**71.944%**	**99.863%**
DOWN	Unscheduled	0d 8h 40m 26s	0.099%	0.137%
	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**0d 8h 40m 26s**	**0.099%**	**0.137%**
UNREACHABLE	Unscheduled	0d 0h 0m 0s	0.000%	0.000%
	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**0d 0h 0m 0s**	**0.000%**	**0.000%**
Undetermined	Nagios Not Running	0d 0h 0m 0s	0.000%	
	Insufficient Data	102d 1h 0m 0s	27.957%	
	Total	102d 1h 0m 0s	27.957%	
All	Total	365d 0h 0m 0s	100.000%	100.000%

Task 2 Grid Operations
The EGEE node promote is now fully operational and accept the VAMDC users.



Two tutorials where organized with access to VAMDC users :
One specially dedicated to access to grid using Nut Shell, and another using tools like portals for managing job or data.
Documents are available to help user in step by step access to the grid from the beginning (get his certificate, ask for an account on the GUI and to the Observatoire de Paris, construct a job and launch it)
First initiative on preparation, formation and test on the use of the GRID for data production using The Opacity Project / The Iron Project tools.q

Task 3 Support to "users" of the infrastructure
Description of VAMDC Infrastructure is available (on-line and PDF) –
See http://www.vamdc.eu/vamdc-infrastructure

HelpDesk have been install using Request Tracker https://voparis-vamdc-support.obspm.fr with single point of entry support@vamdc.eu.
This tool is one of the most use for help desk and ensure to keep traces of event and be sure that any request have been answer on a reasonable time.

For code repository : Redmine have been install at Observatoire de Paris on http://voparis-vamdc-project.obspm.fr:3000/ but finally GIT was chosen by the development team.

Task 4 Preservation of digital data and resources
First databases have been replicated and backup using the virtualisation storage based on Active Circle. Yet only a disk copy is done. At the stage of the project we don't store them in tape.
Mirroring has been set up also for the registry. As Registry is the heart of the VO information system. It's also the single point of failure. Now replication is possible.
For the mirroring process :
Documentation is about to be released so that data provider can start database mirroring. The mirroring process is a join venture between technical manager of the

initial service and the IT team of VO-Paris.
There are two options to install such a procedure. To dedicate a virtual machine manage by the provider or to do all the procedure from Observatoire de Paris. The second option requires to give us SSH access to the service and data server to synchronize periodicaly.

Task 5 Quality Assurance of data and resources
A TAP-VAMDC Service Validator has been build testing compliance of the service and validation of the answer according to the data model schema. The validator is provided with a user's Documentation for testing databases and service or for building a correct service.

VALD database has also developed a quality control system for their own data.

Significant results (Activities and Deliverables)

a) A complete documentation start to provide full access to user to use and improve VAMDC.
To complete user information a Help Desk system has been build to ensure a good support to any kind of questions.

b) Grid tutorials where successful and will help user if they plan to use the grid in their scientific project.

c) Infrastructure becomes more robust by a complete monitoring, evolution of tools to provide data and soon a full replication system.

Deliverables to EU

D5.1 Service Support Plan –DONE-
See http://www.vamdc.eu/public-deliverables/16-deliverables-wp5

D5.2 Infrastructure Support Report to be included in report to the EU– Year 1- DONE – see http://www.vamdc.eu/public-deliverables/16-deliverables-wp5

D5.3 Infrastructure Support Report to be included in report to the EU – Year 2- DONE –see http://www.vamdc.eu/public-deliverables/16-deliverables-wp5

Annual Service Support Plan revisions included in Revised Annual VAMDC Project Plans – Year 1,2
See D1.2 and D1.5 http://www.vamdc.eu/public-deliverables/12-deliverables-wp1

Internal Deliverables

1) **Nagios Monitoring :** . https://voparis-vamdc-monitoring.obspm.fr (document currently drafted to explain system)
2) **GRID** = explanations about how to use the GRID at http://www.vamdc.eu/vamdc-infrastructure/114 - In addition the grid tutorial provides some more information to users at http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/GridWorkshopTutorial
3) **Support to Users**
   - Help Desk and "Description of Infrastructure" (see report for links)

4) VAMDC-TAP Service Validator Tool : http://www.vamdc.eu/software

Deviations from the contract (Annex I) and reasons for them (if applicable)

none

Failures to achieve critical objectives and/or not being on schedule and reasons for them (if applicable)

a) Service replication has only started and will be operational when publication tool will be in a very stable state. Typically in period 3. b) In the Support to user, the Group of expert to answer user request has not been define yet. We have focus on period two to write documentation dedicated to providers and users. c) Currently the services are opened as prototypes bug report and features enhancement are discussed via the developers mailing lists. There is no material in RT to construct FAQ in period 2. This will be done in stable release. d) Code repository : Due to developer habit, VAMDC project use gitub http://github.com, all the Django code for deploying VAMDC software deployment is at https://github.com/VAMDC/NodeSoftware. The Redmine installed has not been used. No EU resources was used for Redmine implementation. No impact on project

Proposed corrective actions (if applicable)

a) Will be done in Period3 b) Will be done in Period 3 c) Will be done in Period 3 d) Nothing to correct