# VAMDC

# Virtual Atomic and Molecular Data Centre

**D5.2**

–

**Service Support Report 1**

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 |

## Document

| | |
|---|---|
| Deliverable number: | D5.2 |
| Deliverable title: | Service Support Report 1 |
| Due date of deliverable: | June 2010 |
| Actual submission date: | 8 September 2010 |
| Authors: | P. Le Sidaner, M.L. Dubernet |
| Work Package no.: | WP5-SA2 |
| Work Package title: | Support to the Infrastructure |
| Work Package leader: | CNRS |
| Lead beneficiary: | CNRS |
| Dissemination level: | PU |
| Nature: | Report |
| No of pages (incl. cover): | |

| Abstract | The objective of D5.2 is to describe VAMDC Service Support Report for Period 1. This report corresponds to Activities in WP5: SA2 "Support to the Infrastructure". This report is included in the VAMDC Periodic report for Period 1. |
|---|---|

**Versioning and Contribution history**

| Version | Date | Reason for modification | Modified by |
|---------|------|------------------------|-------------|
| V0.1 | 20/07/2010 | Section 5 | P. Le Sidaner |
| V0.2 | 20/07/2010 | Draft D5.2 + some links added to Section 5 | M.L. Dubernet |

| Final Version (v0.2) released by | | Circulated to | |
|---------|------|------------------------|-------------|
| Name | Date | Recipient | Date |
| M.L. Dubernet | 8th September 2010 | Mrs Asero | 8th September 2010 |
| | | | |

# 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:<br>(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 | Testing by Users Panels of prototype software |
| M5.4 | Deployment of Preservation and QA | WP5 | CNRS | Months 10, 22, 34, 42 | Text of Call Available on Public Website |

### 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 | L.A. Molina (CNRS:LUTH, LPMAA) | All |
| 3 | Not yet decided as it starts on M12 | |
| 4 | P. Le Sidaner (CNRS: UMS) | All |
| 5 | Not yet decided as it starts on M12 | |

**Description of work** (possibly broken down into tasks)

*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 advertise 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 Description for Period 1

**Task 1:** Maintenance and monitoring of the core infrastructure

- 1.1 Install of Nagios (M6)
- 1.2 The list of machines and services to monitor as well as contacts will be defined by the first 9 months and will be updated over time. Using existing plug-ins or defining new.
- 1.3 New plug-ins will depend on the protocol defined inside VAMDC
- 1.4 Installation and maintenance of the monitoring process for machines and services

**Task 2:** Grid Operations

- 2.1 Making an inventory of codes where grid launch should be useful : this task has to be started by the first 7 months (codes producing A&M data or application codes using A&M data)  - Provide List of codes
- 2.2 Gridify codes if necessary  and test codes in EGEE grid  - Use 2 codes and Record Pbs, edit documentation - Starting Month 7
- 2.3 Find the most useful portal for launching and monitoring these codes on the Grid– Provide Documentation-  Starting Months 7

**Task 3:** Support to "users" of the infrastructure – Starts only on Month 12 (Cycle 2)

**Task 4:** Preservation of digital data and resources
- 4.1 After first inventory of services, some more point will be defined by contacting contact persons for  services at each node res:
  o What technologies to be used (OS, BDD, language required, total volume ...)
  o Who will be the  technical contact to give access to for replication of service
  o Who will be the  scientific contact who will  validate mirror
- 4.2 This jobs will be start just after inventory (and only for few sites):
  o Virtual machine installation (M11)
  o Negotiation on each service (M11)
  o Installation of services and replication process for chosen sites (M11)

**Task 5:** Quality Assurance of data and resources - Starts only on Month 12 (Cycle 2)

## 5.  WP5 Tasks Report for Period 1

**Period**: 01/07/2009 – 30/06/2010
**WorkPackage:** WP5-SA2- Service Support
**WorkPackage Leader and co-Leader**: P. Le Sidaner, K. Benson
Participants in the WorkPackage:  CNRS, CMSUC, UCL, INAF, ISRAN, RFNC-VNIITF, IAO, CTPM, INASAN

| Part 1 |
| --- |

A summary of progress towards objectives and details for each tasks

| |
| --- |
| Task 1 Maintenance and monitoring of the core infrastructure<br>After a census of services made in collaboration with WP4 (census used in task 1 and 4  of WP5):<br>    • System of monitoring of core infrastructure have been installed (NAGIOS) and configured.<br>    • Specific plug-ins have been developed to monitor VAMDC Table Access Protocol services.<br>The realisations are in accordance with time-table of previsions.<br><br>Task 2 Grid Operations<br>After a census of codes (: http://voparis- |

twiki.obspm.fr/twiki/bin/view/VAMDC/GridCodesCensus, 2 candidates where chosen to be launched on the grid.

- Gridification process was successful for the 2 codes.
- Portal Tools to manage codes were tested
- Grid Infrastructure of EGEE node at Observatoire de Paris is now up and running with almost 200 core, available to VAMDC participant according to EGEE policy.

Task 4 Preservation of digital data and resources
- start of service replication to ensure high availability and service quality.
- build an archive to store data for longer time preservation
- installation of Astrogrid infrastructure to compose a self consistent replication system with a dedicated registry.

Task 5 Quality Assurance of data and resources
- Quality control of the user request (together with WP7) on XSAMS format of VALD database.
- Scientific Quality Assurance Activities on all databases involved in the process of output of databases using XSAMS.

Significant results (Activities and Deliverables)

a) monitoring activities is implemented at VO-Paris Data Centre using the NAGIOS software.
Nagios server is up and running with restricted access at
https://voparis-vamdc-monitoring.obspm.fr/
16 host are already monitored and 17 services related to theses hosts.

On every service or host problem, monitoring system send regular mail alert to the proper persons. The system allow possibility of acknowledgment of service provider

| Host ⬆⬇ | Service ⬆⬇ | Status ⬆⬇ | Last Check ⬆ | Duration ⬆⬇ | Attempt ⬆⬇ | Status Information |
|---|---|---|---|---|---|---|
| CDSD | CDSD is the calculated line list of 7 isotopoloques of carbon dioxide. | CRITICAL | 06-23-2010 07:15:36 | 28d 14h 15m 2s | 3/3 | CRITICAL - Socket timeou |
| OA_Cagliari_PAH | Theoretical spectral database of polycyclic aromatic hydrocarbons | OK | 06-23-2010 07:16:25 | 3d 7h 38m 33s | 1/3 | HTTP OK: HTTP/1.1 200 |
| SMPO | Spectroscopy and Molecular Properties of Ozone | OK | 06-23-2010 07:21:37 | 4d 1h 43m 21s | 1/3 | HTTP OK: HTTP/1.1 200 |
| axis_xms | AXIS optical spectra XMM | OK | 06-23-2010 07:20:25 | 7d 19h 14m 33s | 1/3 | HTTP OK: HTTP/1.1 200 |
| carbon_cluster | Theoretical spectral database of pure carbon clusters | OK | 06-23-2010 07:24:25 | 3d 7h 40m 33s | 1/3 | HTTP OK: HTTP/1.1 200 |
| cdms | CDMS-tap-availability | OK | 06-23-2010 07:16:36 | 3d 17h 58m 22s | 1/3 | HTTP OK: HTTP/1.1 200 |
|  | CDMS: Cologne database for molecular spectroscopy | OK | 06-23-2010 07:18:38 | 8d 23h 26m 21s | 1/3 | HTTP OK: HTTP/1.1 200 |
| chianti | CHIANTI-tap-availability | CRITICAL | 06-23-2010 07:22:06 | 5d 16h 36m 52s | 3/3 | HTTP CRITICAL: HTTP/1 found - 1273 bytes in 0.029 |
| dijon_methane_lines | Service calculated dijon-methane-lines | OK | 06-23-2010 07:21:37 | 8d 23h 23m 21s | 1/3 | HTTP OK: HTTP/1.1 200 |
| kida | kida | OK | 06-23-2010 07:21:37 | 8d 23h 23m 21s | 1/3 | HTTP OK: HTTP/1.1 200 |
| stark_b | Stark-B | OK | 06-23-2010 07:18:25 | 7d 17h 46m 33s | 1/3 | HTTP OK: HTTP/1.0 200 |
| tipbase | TIPbase | OK | 06-23-2010 07:18:25 | 0d 14h 36m 33s | 1/3 | HTTP OK: HTTP/1.1 200 |
| topbase | TOPbase | OK | 06-23-2010 07:23:37 | 0d 14h 41m 21s | 1/3 | HTTP OK: HTTP/1.1 200 |
| umist | UMIST Database for Astrochemistry | OK | 06-23-2010 07:15:37 | 1d 15h 49m 21s | 1/3 | HTTP OK: HTTP/1.1 200 |
| vald_moscow | VALD acces in Moscow | OK | 06-23-2010 07:17:37 | 1d 13h 7m 21s | 1/3 | HTTP OK: HTTP/1.1 200 |
| vald_uppsala | VALD acces in Uppsala | OK | 06-23-2010 07:18:38 | 0d 0h 36m 20s | 1/3 | HTTP OK: HTTP/1.1 200 |
| vald_wien | VALD acces in Vienna | OK | 06-23-2010 07:18:25 | 8d 23h 26m 33s | 1/3 | HTTP OK: HTTP/1.1 200 |
| vamdc.eu | Basecol-tap-availability | CRITICAL | 06-23-2010 07:21:37 | 71d 23h 40m 35s | 3/3 | HTTP CRITICAL: HTTP/1 response time |
|  | Site VAMDC | OK | 06-23-2010 07:21:37 | 8d 23h 23m 21s | 1/3 | HTTP OK: HTTP/1.1 200 |
|  | tomcat-astrogrid | OK | 06-23-2010 07:18:25 | 27d 18h 1m 34s | 1/3 | HTTP OK: HTTP/1.1 200 |

to stop alert or downtime period for maintenance.

…Statistical report are associated to nagios monitoring and can be used as quality assurance for infrastructure like host or service availability

| State | Type / Reason | Time | % Total Time | % Known Time |
|---|---|---|---|---|
| UP | Unscheduled | 61d 5h 24m 2s | 37.489% | 99.853% |
|  | Scheduled | 0d 0h 0m 0s | 0.000% | 0.000% |
|  | Total | 61d 5h 24m 2s | 37.489% | 99.853% |
| DOWN | Unscheduled | 0d 2h 9m 36s | 0.055% | 0.147% |
|  | Scheduled | 0d 0h 0m 0s | 0.000% | 0.000% |
|  | Total | 0d 2h 9m 36s | 0.055% | 0.147% |
| 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 0h 0m 0s | 62.456% |  |
|  | Total | 102d 0h 0m 0s | 62.456% |  |
| All | Total | 163d 7h 33m 38s | 100.000% | 100.000% |

b) Task 2 : codes candidate to be launch on the grid.
 **PDR** come to be the first candidate to test an access portal to EGEE. At this time the code is in the testing phase. This test has been documented, on the VAMDC Wiki http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/PdrOnGrid

**gXSTAR** (Juan González & Claudio Mendoza). One of the major interests of the VAMDC Venezuelan node is to make it possible for users to run complex applications on the available grid infrastructure. As part of the initial tests, we have worked on porting XSTAR to the grid, integrating web interfaces developed for wXSTAR to simplify input file building and configuring routines required to work with the EELA2 grid. The result of such integration has been gXSTAR, a web tool designed to prepare the JDL file and libraries needed to run on the grid. Using the gXSTAR portal, users can create a grid-working directory designed to be fully compatible with the grid, requiring only an addition FTP transfer from the user's computer to the grid User-Interface (UI) node. In order to employ it after input file creation, the user must have a valid EELA2 account on the UI with a proper user certificate.

c) Background research and initial census of end user tools that can be useful to the project. From this census a choice and classification includes two web portals candidates. The documentation is available on the VAMDC wiki : http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/GridTools.

Task 4
a) First inventory of the services have been made in collaboration with WP4. The result is available on the twiki. It contains the description of services and local contacts, size and type of Databases.

b) Fisrt candidate is Bascol where contact are easy and access also.
The replication machine has been build in FreeBSD
The replication procedure have been done for the MySql databases, using master/slave capability.
Some test are still in progress to see if PHP evolution of the application is possible into 5.3 version in order to ensure security and evolution of the site.
Moreover, VALD project has already service replication that can be consider as a prototype candidate. INASAN describe this support of Moscow mirror-site of VALD database. Archiving the data and control of the synchronization process http://www.astro.uu.se/valdwiki/ValdBackup.

1  An instance of Astrogrid is running at Paris Observatory
   http://voparis-vamdc-astrogrid.obspm.fr:8080/
Public access will be given when the environment will be operational.
It will allow to define all the replicate services locally in the Astrogrid registry. The voparis-vamdc site will have the services replicates and the registry to locate and query them. This self consistent system will prevent from failure of service.

- Long time preservation
Copies of the VAMDC virtual machine will allow long time preservation.
VO-Paris as a part of IT team have build a long time preservation system based on storage virtualisation using distributed and shared storage system.
This system will allow to store data in different cells with a pre defined quality of service (for example 2 disk instance and one tape instance). The storage software will constantly verify the number of instance and duplicate on new support if there is any fail on a cell.

The storage system is virtualised which allow to so solve storage units bounded lifetime,

The tape storage use open format tar which gave no dependency on the virtualisation system.

All these environment give possibility to save the all virtual machines in a long time storage system. Paris Observatory is the oldest Europeen observatory, which give a formal guaranty of long time maintenance.

Task 5 : Quality Assurance

A unified description of VALD atomic transitions for the subsequent transfer to XSAMS format http://www.astro.uu.se/valdwiki/VALD3linelists. Conversion of part of VALD data to unified format and preparation VALD as a prototype for demonstration of VAMDC working schema by the end of the first year (joint work of partners 11 and 15) http://www.astro.uu.se/valdwiki/AtomicLevel.

Quality control of the user request (together with WP7) http://voparis-twiki.obspm.fr/twiki/pub/VAMDC/PmCycleOne/WP5_t5.pdf, http://voparis-twiki.obspm.fr/twiki/pub/VAMDC/PmCycleOne/pakhomov_work.pdf, http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/CodingResources, http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/VALDJAXB.

Scientific Quality Assurance activities have been carried out at other nodes because people need to check the output of their databases using the various protocols

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

NOT APPLICABLE

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

a) Some delay come on task 4, only one service was replicate on cycle 1. Because the Work Package leader & Task Leader (Pierre Le Sidaner) had an accident on the 8th of February.  He was out of work for 3,5 month.
Service replication for the first 4 candidates is delayed and shifted to cycle 2.

b) Task 5 has started before cycle 2 from the scientific point of view of Quality Assurance. It must be understood that other Scientific Quality Assurance activities have been carried out at other nodes because people need to check the output of their databases using the various protocols. This is a normal procedure that will continue during the whole project.

Proposed corrective actions (if applicable)

a) The job not done on cycle 1 task 4 will be added to the cycle 2 program and will be achieved on cycle 2 period.

D5.1 Service Support Report 1

*(approximate length of Part 1: 2 pages)*