VAMDC

Virtual Atomic and Molecular Data Centre

D4.3

–

Infrastructure Deployment Report 2

Version 0.3

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:	D4.3
Deliverable title:	Infrastructure Deployment Report 2
Due date of deliverable:	June 2011
Actual submission date:	August 2011
Authors:	G. Rixon, F. Kosmala, M.L. Dubernet and WP4 team
Work Package no.:	WP4-SA1
Work Package title:	Infrastructure Deployment
Work Package leader:	CSMUC
Lead beneficiary:	CSMUC
Dissemination level:	PU
Nature:	Report
No of pages (incl. cover):	

Abstract	The objective of D4.3 is to describe VAMDC Infrastructure Deployment Report for Cycle 2. This report corresponds to Activities in WP4: 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 to WP4	F. Kosmala
V0.1	July 2011	WP4 report for P2	G. Rixon
V0.1	July 2011	Making of D4.3 with inclusion of P2 report	F. Kosmala
V0.2	August 2011	correction P2 plan section	M.L.Dubernet
V0.3	August 2011	Update of P2 report section	G. Rixon
V0.3	August 2011	Final Check	M.L. Dubernet

Final Version (v0.1) released by		Circulated to	
Name	Date	Recipient	Date
M.L. Dubernet	24th August 2011	Mrs Asero	24th August 2011

WP4 ACTIVITIES DESCRIPTION

Work package number	4		Start date or starting event:			3				
Work package title	SA 1: Infrastructure Deployment									
Activity Type	OTHER									
Participant id	1	2	3	4	5	6	7	7	12	13	14
Person-months per beneficiary: (Total = EU + Node Contributions)	66	36	36	12	6	6	8	11	5	9	24

Table of Content

1. WP4 Objectives

To provide Data Access via a homogeneous environment where the distributed user community can retrieve AM resources through a standard interfaces. This involves implementing standard outputs for the AM databases, finding the resources by interrogating registries, using querying and pipeline tools.

WP4 leader is CMSUC (2)

2. WP4 Milestones and Deliverables

Milestones

M4.1	Deployment of Data Access	WP4	CSMUC	Months 10, 22, 34, 42	
M4.2	Deployment of Infrastructure	WP4	CSMUC	Months 10, 22, 34, 42	
M4.3	Evaluation of Available Software	WP4	CSMUC	Months 10, 22, 34, 42	Testing by Users Panels of prototype software
M4.4	Open Call for New Resources	WP4	CSMUC	Months 24	Text of Call Available on Public Website

Deliverables

D4.1 Infrastructure Deployment Plan (PM 3)
D4.2 Infrastructure Deployment Report to be included in report to the EU – Year 1 (PM 10)
D4.3 Infrastructure Deployment Report to be included in report to the EU – Year 2 (PM 22)
D4.4 Infrastructure Deployment Report to be included in report to the EU – Year 3 (PM 34)
D4.5 Final Report of Service Deployment to be included in final report to the commission (PM41)
Annual Infrastructure Deployment Plan revisions included in Revised Annual VAMDC Project Plans – Year 1,2,3

3. WP4 Tasks Description

WP4 Leader	G. Rixon (CMSUC)	
Task Number	Leader	Other Partners
1	G. Rixon (CMSUC)	All partners
2	L. Molina (CNRS)	CMSUC (2), IVIC (14), UU (6) + others TBD
3	K. Benson (UCL)	CNRS (1), UU (6), RFNC-VNIITF (12)
4	ML Dubernet (CNRS)/G. Rixon (CMSUC)	UU(6), UCL (3), KOELN (7)
5	M. Doronin (CNRS)	UU(6), UCL (3), KOELN (7)
6	M.L. Dubernet (CNRS)	All partners

Description of work

The VAMDC infrastructure will be designed as an homogeneous environment where any AM producer or "community" users will be able respectively to publish their AM data or to retrieve and manipulate those data. The AM producers range from atomic physics to molecular physics handling complex molecules, solids and surfaces. The communities encompass astrophysics users from very different areas: stellar, galaxies, interstellar medium (those application areas are handled by the IVOA and Euro-VO projects), planetology and small bodies of the solar system (EuroPlanet Project), solar-earth system (EGSO and SPASE projects), atmospheric users (studies of earth atmosphere), environmental and combustion chemistry, fusion physics and industrial applications. The IVOA community is the most advanced project as far as building an interoperable infrastructure for astronomy and we will use some of their achievements, i.e. standards, tools, services when those are relevant to the project.

Task 1: Standard access to AM data (lead by CMSUC(2), all SA1 partners)

We will provide standard service interfaces to AM databases. JRA1 will define these interfaces and this task is to implement them on the existing databases held by each VAMDC node. Participants at all nodes will be involved.

Task 2: Standard access to numerical codes (lead by CNRS(1), with partners (2), (6))

Where a VAMDC node has a useful numerical code for AM analysis or modelling, we will make it available as a service. These services will provide a uniform way of launching the codes and recovering their outputs. This task is complementary to the grid adaptation of code in SA2: the codes run on resources contributed by the node owners and need not be made portable to an external grid.

Task 3: Implementing registries (lead by UCL(3) with partners (1), (6), (12))

The registry facilities defined by JRA1, and implemented with the software produced by JRA2, must be populated with information. This task gathers the meta data for the services at each node and adds it to the registries.

Task 4: Augmenting VODesktop (lead by UCL(3))

The EuroVO's VODesktop is a generic interface for the virtual observatory. It allows access to all VObs data, plus launching numerical codes and sharing of data between desktop visualization tools. We will adapt A-M desktop applications to work with VODesktop and the underlying VObs applications environment.

Task 5: Publishing desktop software (lead by CNRS(1), with partner (5))

We will collect and make available to end users chosen A-M applications for the desktop.

Task 6: Expansion of the infrastructure (co-lead by CNRS(1) and CMSUC(2) with (5), (6))

Once the core infrastructure is deployed, new resources will be included in the infrastructure via an open call to producers of AM resources. Those new resources will need to be deployed and tested within the infrastructure. Task 6 will be devoted to the technical inclusion and testing of these new AM resources. The choice of these resources will be made in NA1 by the VPB by the EPT.

4. WP4 Tasks Plans for Period 2

Period: 01/07/2010 – 30/06/2011
WorkPackage: WP4 SA1 Deployment
WorkPackage Leader and co-Leader: G. Rixon and A. Shih
Participants in the WorkPackage: CNRS, CMSUC, UCL, OU, UNIVIE, UU, KOLN, INAF, RFNC-VNIITF, IAO, IVIC

Part 1

Objectives and details for each task in Year 2.

The emphasis in period 2 is on the web services providing access to data. We shall deploy data-access services on all data-sets for which we have an applicable data-model (XSAMS) and web-service protocol (TAP, TAP-XSAMS). Data-sets of other, more-specialized types will receive their data-access services in period 3. We shall continue to develop UIs and the registry of resources to support use of the data-access services. We do not intend to provide general access to numerical codes during period 2, but shall instead develop the framework by which such access may be provided during period 3.

Task 1: access to data

TAP-XSAMS services will be deployed for all VAMDC data-sets in the deployment census for which the XSAMS model is suitable. This includes data-sets in existing databases (Task 1.6), for which the service implementation has to be fitted to the database schema and data sets currently stored in flat files (Task 1.5), for which the data will be ingested into relational databases using a standard schema (Task 1.4). A TAP-XSAMS implementation (in Python, based on the Django framework) was created in period 1. This implementation will be refined during period 2 and distributed to VAMDC sites as a complete package with all software dependencies (Task 1.7).

A prototype data-access service for the solid-state-spectroscopy data-model (SSDM) will be developed and deployed. This will support testing of SSDM and will inform the design, by WP6, of a standard data-access protocol for SSDM. The prototype is not expected to form part of the level-2 service release.

Task 2: access to codes

An existing, grid-enabled code (to be specified during period 2, as part of WP8 Task 3.1) will be adapted to takes its input data directly from VAMDC data-services (Task 2.2). This will serve as a demonstrator of how VAMDC might be exploited in large computations. It will not be part of the VAMDC core in the long term.

Services will be deployed to execute a code that is bound to a VAMDC data-set at the site where those data are stored (Task 2.1). This is a proof-of-concept installation for the code-execution technology, so the choice of code is not critical; Xstar and codes associated with CHIANTI are likely choices. Two such services will be deployed for a given code, one using AstroGrid Common Execution Architecture and one using SOAPlab2, so that we can compare the two systems.

Wider deployment of code-access services is deferred to period 3.

Task 3: registry

New services will be added to the registry as they are deployed (Task 3.4).

The registry contents will be cleaned up (Task 3.6) and reorganized around data-

---

collections rather than individual services (Task 3.5). This will makes it easier to show scientifically-relevant views of VAMDC in the UIs.

Task 4: user interfaces

The web portal produced in period 1 will be refined. Its navigation will be simplified, and the scientific focus improved (Task 4.5). A user interface for the TAP-XSAMS services will be added (Task 4.4). Features of particular scientific interest will be added, such as the ability to raise equivalent data from many VAMDC datasets and to merge or compare them. Whereas the period-1 portal was an engineering tool, the portal at the end of period 2 can be regarded as a prototype of the eventual science-portal.

An interface to the TAP-XSAMS services will be added to VODesktop (Task 4.3) and to Taverna (Task 4.6).

Task 5: distribution of desktop software

A list of useful and desirable applications will be drawn up during period 2 (Task 5.2). Some items from this list may be distributed during period 2, but we expect the major distributions to happen in period 3, after the data-access services mature.

VAMDC's generic UIs (extended versions of VODesktop and Taverna) will be made available to VAMDC users.

Task 6: expansion
No activity planned in period 2.

5. WP4 Tasks Reports for Period 2

<p align="center"><i><u>VAMDC</u></i>
Periodic Report P2 (per Workpackage)</p>

Period: 01/07/2010 – 30/06/2011
WorkPackage: WP4 – deployment
WorkPackage Leader: G. Rixon
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

A new procedure has been adopted for release of infrastructure. For each formal update to the VAMDC infrastructure-standards there is a corresponding update to the web services implementing those standards (portal, registry, data-access services). Thus, we expect four releases of infrastructure per year during the VAMDC project. Subsequent versions of the services appear on different URLs; the two most-recent versions are kept available.

The first such release occurred at the end of period 2, for the 2011.05 version of standards. The next is expected in July for the 2011.07 standards and this latter update will form the level-2 release that is presented for the year-2 review.

Task 1: access to data

The VAMDC-TAP protocol (formerly TAP-XSAMS) is now the standard means of access to data within VAMDC. No further deployments of IVOA TAP and SLAP services are planned, and authors of client software for VAMD will be encouraged to use VAMDC-TAP.

Databases which implement and register a VAMDC-TAP protocol automatically become available through the VAMDC web-portal.

VAMDC now supports two reference implementations of a VAMDC-TAP web-service: a Django/Python implementation derived from the VALD prototype and a Java implementation derived from the BASECOL prototype. VAMDC nodes are encouraged to use or adapt one of these rather than writing their own implementation. A tutorial on use of the reference implementations for data providers was held at Meudon during January 2011.

The following databases have or will have VAMDC-TAP services. The availability column reflects release to beta testers: those marked period 2 were available during the reporting period; those marked period 3, 2011.07 update became available shortly after the end of period 2; the rest are to be completed later in period 3.

Data set	Site	Availability
VALD	Uppsala	Period 2, 2011.05 update
Chianti	Cambridge/UCL	Period 2, 2011.05 update
CDMS	KOELN	Period 2, 2011.05 update

BASECOL	CNRS-LPMAA-LUTH	Period 2, 2011.05 update
STSP/GhoSST	CNRS-LPG/IPAG	Period 3 (see note below)
UDfA	QUB	Period 2, 2011.07 update
KIDA	CNRS-LAB	Period 3, 2011.07 update
PAH	INAF-OAC	Period 3, 2011.07 update
LASP	INAF	Period 3 (see note below)
Spectr-W3	VNIITF	Period 2, 2011.05 update
CDSD	IAO	Period 2, 2011.05 update
eMol	OU	Period 3
HITRAN	UCL	Period 2, 2011.05 update
NIST atomic database		Period 3
Methane lines	CNRS-ICB	Period 3, 2011.07 update
S&MPO	CNRS-GSMA/IAO	Period 2, 2011.05 update
Ethylene	CNRS-GSMA	Period 2, 2011.05 update
Stark-B	CNRS-LERMA/AOB	Period 3
Lund atomic data	Uppsala	Period 2, 2011.05 update
uADB (formerly XStarDB)	CTPM/Cambridge	Period 3
TipTopBase	CTPM/CNRS-LERMA	Period 2, 2011.05 update

GhoSST and LASP contain solid-spectroscopy data which cannot be fully accessed through VAMDC-TAP. A specialized data model and web-service protocol (as similar to VAMDC-TAP as possible) for these data are being developed at LPG/IPAG for release in period 3.

Task 2: access to numerical codes

IVIC and UCL are investigating ways to offer numeric codes relevant to VAMDC as web services. These services could be called from users' own scripts and applications and would take their inputs from the VAMDC data-services.

Prototype web-services for the Xstar simulation-package have been created using the SOAPlab[1] system and the AstroGrid Common Execution Architecture[2] with the intention of comparing the usefulness of the two systems. The experiment is currently waiting for the data services to be completed to the 2011.05 standards and the comparison will now be made in period 3.

If the model of numeric-software-as-service is found scientifically useful and technically suitable, services of this kind will be made available to users in period 3.

Task 3: registry

The registry has been extended with a new XML-schema that allows more metadata in

---

[1]   http://www.myexperiment.org/workflows/1206.html

[2]   Runs with Taverna, using VAMDC extensions. See the registry entry http://casx019-zone1.ast.cam.ac.uk/registry/main/viewResourceEntry.jsp?IVORN=ivo%3A%2F%2Fvamdc%2Fxstarcmdappwspectrumfile

the registration of VAMDC-TAP services. This schema will become part of the 2011.07 standards.

The registry is a single point of failure for VAMDC. A mirror scheme is being developed to reduce this risk and will be included in the level-2 release.

The registry can be inspected through its web UI.[3]

Task 4: user interfaces

The "level-1" portal used in period 1 and 2 focused on the needs of developers within VAMDC. This portal is due to be replaced by the "level-2" portal which reflects the needs of end users. An early prototype of the level-2 portal is already available within VAMDC. The first release of this portal to users will be in the level-2 release in July.

VAMDC now provides and supports extensions ("plug-ins") to the Taverna workflow-system that give access to VAMDC-TAP services.

VAMDC no longer supports or recommends the use of VODesktop. Taverna and the web portals replace this software.

Task 5 Desktop software

A validation tool for VAMDC data-services is available.[4]

Task 6 expansion of infrastructure

The Lund archive of atomic data has been added to VAMDC, with support from the Uppsala node.

Significant results (Activities and Deliverables)

Deliverables to EU
D4.1 Implementation Plan – DONE –   *See* **http://www.vamdc.eu/public-deliverables/16-deliverables-wp4**   *D4.2 Implementation Report to be included in report to the EU – Year 1 – Done –*   *See* **http://www.vamdc.eu/public-deliverables/16-deliverables-wp4**   *D4.3 Implementation Report to be included in report to the EU – Year 2 – Done –*   *See* **http://www.vamdc.eu/public-deliverables/16-deliverables-wp4**   *Annual Implementaion 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   a) Deployment workshop in Meudon, January 2011.   b) System release (2011.05 standards) in June 2011.

---

3   http://casx019-zone1.ast.cam.ac.uk/registry/ - this views the evolving registry used for VAMDC development rather than the fixed registry used for releases; the latter was not available at the time of reporting.
4   http://vamdc.eu/software

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

No deviations.

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

A) Some databases not yet available on-line (not accessible with standard services from the portal): Ghosst and LASP because of the complexity of the solid spectroscopy data model; KIDA because of delayed registration; PAH, Stark-B, Methane because the features they needed did not come available before May 2011 (official release of standards); uADB because of lack of time by people in Cambridge and complexity of interaction with CTPM; eMOl because we believe that the content was not created prior to start of VAMDC.

B) Numeric-software-as-service experiment not yet completed due to lateness of data services.

Proposed corrective actions (if applicable)

A) For availability of databases:
  a) Ghosst, KIDA, PAH, Stark-B, Methane are very close to being accessible → before mid-P3
  b) LASP employs somebody from June 2011, so the access should be OK
  c) EMol might request some external help from other nodes either foran implementation at the node or at a hosting site (see point d)

d) Where appropriate, arrange for sites to host data-access services on behalf of other sites (curation of the databases remains with the sites providing the data) : for example uADB is planned at Cambridge and EMol could be hosted at Cambridge or UCL.

B) Carry forward the software-as-service experiment into the start of period 3.

(approximate length of Part 1: 2 pages)