VAMDC

Virtual Atomic and Molecular Data Centre

D6.5

–

Final Interoperability Report

Version 1.0

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	CPTM	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:	D6.5
Deliverable title:	Final Interoperability Report
Due date of deliverable:	December 2012
Actual submission date:	January 2013
Authors:	M. Doronin, M.L. Dubernet and the WP6 team
Work Package no.:	WP6-JRA1
Work Package title:	Interoperability
Work Package leader:	CNRS/KOLN
Lead beneficiary:	CNRS/KOLN
Dissemination level:	PU
Nature:	Report
No of pages (incl. cover):	

Abstract	The objective of D6.5 is to describe VAMDC Interoperability activities during the whole project. This report corresponds to Activities in WP6: JRA1 "Interoperability".

Versioning and Contribution history

Version	Date	Reason for modification	Modified by
V0.1	Dec 2012	WP6 Report	M.L. Dubernet
1/0	January 2013	D6.5 doc	M.L. Dubernet

Final Version (v1.0) released by		Circulated to	
Name	Date	Recipient	Date
M.L. Dubernet	12th February 2012	Mr Bodo	12th February 2012

WP6 ACTIVITIES DESCRIPTION

Work package number	6			Start date or starting event:			3			
Work package title	JRA1: Interoperability									
Activity Type	RTD									
Participant id	1	6	7	8	9	10	12	14			
Person-months per beneficiary: (Total = EU + Node Contributions)	36	18	10	5	12	6	3	6			

Table of Content

1. WP6 Objectives as in Annexe I

Define all standards necessary to build an interoperable infrastructure

WP6 Leader is KOLN with co-leader CNRS

2. WP6 Milestones and Deliverables as in Annexe I

Milestones

M6.1	Technical Meetings	WP6	UU	Months 5,10, 16, 22, 28, 34, 40, 42	Minutes. Presentations on internal Website
M6.2	Evaluation of standards releases	WP6	UU	Months 10, 22, 34	

Deliverables

D6.1 Interoperability Plan (PM 3)
D6.2 Interoperability Report to be included in report to the EU – Year 1 (PM 10)
D6.3 Interoperability Report to be included in report to the EU – Year 2 (PM 22)
D6.4 Interoperability Report to be included in report to the EU – Year 3 (PM 34)
D6.5 Final Report of Interoperability to be included in final report to the commission (PM41)
Annual Interoperability Plan revisions included in Revised Annual VAMDC Project Plans – Year 1,2,3

3. WP6 Tasks Description as in Annexe I

WP6 Leader (co)	S. Schlemmer (KOLN)/ M.L. Dubernet (CNRS: LPMAA/LUTH)	
Task Number	Leader	Other Partners
1	M.L. Dubernet/J. Bureau(CNRS:LPMAA/LUTH)	All partners
2	T. Millar (QUB)	All partners
3	N. Piskunov (UU)	CNRS:LPMAA/LUTH
4	M. Doronin (CNRS:LPMAA)	UU
5	S. Schlemmer (KOLN)	All partners

Description of work (possibly broken down into tasks)

Task 1: Data Models and XML Schema (lead by CNRS(1), all JRA1 partners)
The current data models and XML schema are a description of atomic and molecular linelists for use in an astrophysical context and a description of atomic and molecular elementary processes. The documents have been designed by a small number of people, those documents are still in draft mode, do not cover all application fields and have not been discussed extensively among users and producers. These preliminary versions will be completed and extended in order to cover a wider range of species, a wider range of processes and will include the effect of the environment. For now we identify the following extensions:

- inclusion of solid, surface spectroscopy for interstellar medium and planetology
- inclusion of larger molecules such as PAH
- description of atomic and molecular line shapes arising from different sources
In connection to the user & producer board of NA2, more extensions will be considered if necessary. Through the Standards and Processes Committee all standards will be connected to International efforts of standardisation.

Task 2: Dictionaries (lead by QUB(9), all JRA1 partners)
 In order to uniquely identify resources we will need to define and build dictionaries both general and specific to applications. At present we identify the following dictionaries:
- single identification of databases and services
- list of conventions (link to IUAPAC and other convention bodies)
- list of names of species (in relation with other fields such as chemistry)
- list of processes and coding of processes
- list of quantum numbers
 Other lists will be identified during the course of the project in relation with activities in WP2 (NA2)

Task 3: Access Protocols and Query/Retrieval Languages (lead by UU(6), all JRA1 partners)
We will define protocols retrieving different types of resources: numerical data, libraries, documentation, references. Those protocols will cover asynchronous queries and the retrieval of huge sets of data. In a second step we will design a general query language allowing to access and retrieve any atomic and molecular data.

Task 4: Registries (lead by CNRS(1) with partner 6)

D6.5 Final Interoperability Report

Registries provide a mechanism with which applications can discover and select resources-- e.g. data and services--that are relevant for a particular scientific problem. We will start from the registries defined in the IVOA and see how to adapt and/or extend the documents to our own purpose. In particular we wish to implement ways of finding resources at various levels of granularity.

Task 5: Other Documents (lead by KOLN(7), all JRA1 partners)
Other basic definitions/standardisations might be necessary in order to find/identify resources and documents produced by the IVOA will be assessed in order to keep/adapt to our own needs.

4. WP6 Final Tasks Reports

Period: 01/07/2009 – 31/12/2012
WorkPackage: WP6 - Interoperability
WorkPackage Leader and co-Leader: M.L. Dubernet (CNRS: LPMAA/LUTH) / S. Schlemmer (KOLN)
Participants in the WorkPackage : UCL, UU, CMSUC, KOELN, CNRS are the most active participants in WP6. Other partners provide some punctual inputs.

Part 1

A summary of progress towards objectives and details for each tasks

WP6 dealt with the definition of all standards necessary for the interoperable exchange of atomic and molecular data. It includes the definition of data models for atomic and molecular data, as well as for solid spectroscopy, the definition of "keywords", the definition of query protocols and data access, the definition of registries, of units, of versioning processes, of a uniform protocol for web application to process XSAMS files. All these components are described below.

All Standards Documents can be found on http://www.vamdc.eu/standards
This page provides the detailed standards in HTML for the latest release and keep at the bottom the history of the successive releases (Download section).
Before being released on the official website, documents are handled in a repository.

 Task 1: Data Models and XML schema Documents
   1. <u>VAMDC-XSAMS:</u>

Started with the International Atomic Energy Agency's XML Schema for Atomic, Molecular and Solid Data (XSAMS) version 0.1.1, VAMDC consortium has found that modifications/additions were necessary in order to meet the needs of implementation and queries. This effort has been carried out within the WP6 workpackage including contributions/requirements from all VAMDC partners. This effort has resulted in the so-called VAMDC-XSAMS schema that is used within VAMDC. The XSAMS schema provides a framework for a structured presentation of AMPSI data in an XML file. It is based on the physical representation of interaction between various objects through description of the physical states and interaction

D6.5 Final Interoperability Report

characteristics. Therefore, XSAMS provides rules for presentation of

- states of atoms, molecules, solids (surfaces) and some elementary particles,
- characteristics of interaction between physical objects,
- sources of the data
- environment

The VAMDC release documents for the VAMDC-XSAMS schema include: a tarball including the VAMDC-XSAMS schema files (main file is xsams.xsd) and some example fil, the current reference guide and its PDF fil, the documentation created directly from the schema, the changelog document that provides the log between the successive versions of VAMDC-XSAMS, the documentation for the case-by-case description of molecular states.

VAMDC-XSAMS has been continuously updated throughout Period 1,2,3,4 with official release in r12.07. Now the VAMDC-XSAMS schema is at version v1.0 at http://www.vamdc.org/documents/standards/#data-model

2. Schema for Solid Spectroscopy (CNRS/LPG)

A detailed and well-structured Solid Spectroscopy Data Model (SSDM) has been fully developed in order to describe accurately the solid samples (down to the species), the experiments, the spectra, the bands and states and their products, plus several extensions to describe natural samples. It has been ambitiously designed to be able to describe the laboratory data of any solid spectroscopy data provider, from almost any type of research or applied fields.

The 'final' stable version of SSDM (v0.5.1 + several extensions), its dictionary and documentation (http://ghosst.osug.fr/wiki/index.php/SSDM_data_model, ~350p) have been delivered on 23th October 2012.

A large set of European data providers, representing 15 laboratories, were involved in the SSDM development through the "SSDM Expert Working Group" which met 3 times in Grenoble during the last 3 years to give their advice on the SSDM-GhoSST development.

Task 2: Dictionaries Documents

In VAMDC, different pieces of software need to communicate to each other. Apart from protocols and schema, a common vocabulary is needed. By this we mean a list of "global keywords" that should consist of reasonably short, human-readable keywords which uniquely define a certain type of information or data. In the following we describe how the keywords were created and how they are used in different parts of VAMDC software. The common gain in the various aspects is that the vocabulary allows to split the tasks that are common to all data sets from the database-specific information and routines. Thereby it becomes possible to implement software that can be re-used by multiple datasets, reducing the deployment on a new data set to implementing the parts that are truly specific for it.

Keywords origin

In order to compile a list of well-defined names for all kinds of information that VAMDC datasets can contain, we started from the XSAMS schema for atomic and molecular data, that is used as a main data model within the project.

Flattened and stripped, xsams-derived keywords took form like AtomStateLandeFactor, SourceAuthorName, MolecularSpeciesIonCharge.

The keywords representing desired branches of XSAMS like Species,Processes, RadiativeTransitions,Collisions were added, those would find use in future VSS2 query language.

Keywords use

The VAMDC keywords form three overlapping subsets:

- *Restrictables*, used in registries and in VSS query language, any client software and VAMDC user portal must use them to be able to request the data from VAMDC.
- *Returnables* that are currently used in registries and internally in the Django TAP-VAMDC service implementation, they define placeholders in XSAMS tree for user data output.
- *Requestables* that are due to be added to the VSS2 version of the query language. They would describe the branches of the XSAMS schema client wants to see in the output document produced by the service.

InchiKey :
In order to uniquely identify common species across participant VAMDC databases, the Standard IUPAC International Chemical Identifier, and in particular a hash (based on SHA-1) of this identifier (the Standard InChIKey) must be generated for each species (i.e. atom or molecule) within each participant VAMDC node.
http://www.vamdc.eu/documents/standards/inchi/index.html)
So InchiKey has become in r11.12 a compulsory element of standards in order to allow interoperability. But it appeared by end of P3 that this is not enough and that some cases are still non distinguishable. Therefore in June 2012 it has been decided to define a "VAMDCRegistryIdentifier" that will uniquely identify species in VAMDC. This has been formalized in P4 in version 1.0 of VAMDC-XSAMS.

Task 3: Access Protocols and Query/Retrieval Language Documents

Many data-sets in VAMDC include information that can be rendered in the VAMDC-XSAMS data model. Data in that common model could be transformed to and from a table model which uses the same columns for all data-sets. If all the data-sets had this table model as part of the schemata of their databases, then a SQL query to that model would work on all data-sets, and the results could be written in a common format.

VAMDC-TAP is a protocol for data-access services that provide the common table

model matching VAMDC-XSAMS and which can return the results of queries in VAMDC-XSAMS. VAMDC-TAP services accept queries in a restricted form of SQL (VSS2: VAMDC SQL Sub-set #2) and return results in VAMDC-XSAMS or in certain tabular formats. Implementations of VAMDC-TAP map queries from the common table-model to the actual schemata of their databases.

VAMDC-TAP provides "virtual data". I.e., it associates data selection criteria, defined by a query text, with an archived data-set, defined by the address to which the query is sent, the two combined in one URL. Each such URL represents the results of the query as if they had been pre-computed and stored on a web server. The data URLs are semi-permanent; they can be copied between application, bookmarked, emailed to colleagues, etc.

VAMDC-TAP is based on IVOA 's Table Access Protocol (TAP). TAP already provides virtual data and allows us to plug in our query language VSS2 and our data model VAMDC-XSAMS.

VAMDC-TAP is defined as a web-service protocol. That means that VAMDC-TAP services are driven by GET and POST requests to HTTP (or HTTPS) URIs. Low-level details of the protocol are defined by the HTTP RFCs. Further, the service can be implemented in any language and on any database engine without breaking interoperability.

Documentation for Data Access can be found at:
http://www.vamdc.org/documents/standards/dataAccessProtocol/index.html

Task 4: Registry Documents

The International Virtual Observatory Alliance (IVOA) registry allows astronomers to search, obtain details of, and leverage any of the resources located anywhere in the IVO space, namely in any Virtual Observatory. The IVOA defines the protocols and standards whereby different registry services are able to interoperate and thereby realise this goal. IVOA registry defines interfaces on how to query and share resources. Software is written to conform to standard interfaces in order to assist scientific utilities to access particular resource. A resource in this context is represented in XML form and is stored in the registry. A resource may describe anything about an observatory, particular instrument, another registry, and services such as catalogue or table type services, cone searches. Extensions can be made if necessary and this functionality is made available for VAMDC.

he End User does not have the capability to access registry via this Web interface. Only Scientist and other Technical users of VAMDC can use the registry to add or update resources in the VAMDC registry. End Users use other client programs such as the Astrogrid VODesktop to query on the resources located inside the registry.

Location, home page and menus

Production registry: http://registry.vamdc.eu/registry-12.07/

Registry of legacy system: http://registry.vamdc.org/registry-11.12/

Development registry: http://casx019-zone1.ast.cam.ac.uk/registry/

Registry documentation can be found at
http://www.vamdc.org/documents/standards/registry/index.html

 Task 5: Other Documents
 - A Document describing the process of Release and of version numbers
http://www.vamdc.eu/documents/standards/management/index.html

- A document describing "XSAMS processors services"
http://www.vamdc.eu/documents/standards/dataConsumerProtocol/index.html

Applications to process data in XSAMS format may be made available as web sites,
which makes them accessible for interactive use, or as web services, making them
accessible to scripts and other software. This standard prescribes a form for these web
applications that includes both the interactive web-site and scriptable web-service.

Web applications conforming to this standard can be registered in the VAMDC
registry. Registration makes the applications available to generic UIs such as the
VAMDC portal.

Conforming web applications can read data either from a URL (e.g. the portal passes
a data-extract URL leading to a VAMDC database) or from an uploaded file (e.g. a
user loads data from a file on his computer).

 - A document describing Units
http://www.vamdc.eu/documents/standards/units/index.html

Significant results (Activities and Deliverables)

D6.1 Interoperability Plan – DONE –
See **http://www.vamdc-project.vamdc.eu/public-deliverables/18-deliverables-wp6**

D6.2 Interoperability Report to be included in report to the EU – Year 1 – Done –
 See **http://www.vamdc-project.vamdc.eu/public-deliverables/18-deliverables-wp6**

D6.3 Interoperability Report to be included in report to the EU – Year 2 – Done –

D6.5 Final Interoperability Report

See **http://www.vamdc-project.vamdc.eu/public-deliverables/18-deliverables-wp6**

D6.4 Interoperability Report to be included in report to the EU – Year 3 – Done –
See **http://www.vamdc-project.vamdc.eu/public-deliverables/18-deliverables-wp6**

Annual Interoperability Plan revisions included in Revised Annual VAMDC Project Plans – Year 1,2,3
 See D1.2, D1.5, D1.7 **http://www.vamdc-project.vamdc.eu/public-deliverables/12-deliverables-wp1**

Internal Deliverables

Standards are released every 6 months with versioning numbers on the www.vamdc.eu website which is the ONLY official source for the standards in VAMDC.

Final Release [12.07] of standards and software (in html and PDF) has been published.

VAMDC documentation and software versioning policy

Guidelines and Procedures for VAMDC Document Standards, Schemas and Software Management [v 11.07] [r 12.07]

Data access protocol, query language and dictionaries

Data access protocol [v 12.07] [r 12.07]
Change log for data-access protocol
Query language [v 12.07] [r 12.07]
Change log for query language
Dictionaries [v 12.07] [r 12.07]

Data model

VAMDC-XSAMS reference guide [v 1.0] [r 12.07]
VAMDC-XSAMS changelog [v 0.2] [r 12.07]
Case-By-Case reference guide [v 11.05] [r 12.07]
VAMDC-XSAMS schema files are available for download below.
VAMDC-XSAMS schema documentation can be viewed in the browser or downloaded from below.
Case-By-Case schema documentation can be viewed in the browser or downloaded from below.

Registry

Registry user's guide [v 12.07] [r 12.07]
Change log for registry guide

Units

Unit Conversion Table [v 11.12] [r 12.07]

InChI Generation

InChI/InChIKey [v 11.12] [r 12.07]

XSAMS Processor service

Data consumer protocol [v 12.07] [r 12.07]

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

D6.5 Final Interoperability Report

Not Applicable

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

Not Applicable

Proposed corrective actions (if applicable)

Not Applicable

D6.5 Final Interoperability Report