# VAMDC

# Virtual Atomic and Molecular Data Centre

**D6.2**

**–**

**Interoperability Report 1**

Version 0.6

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 |

D6.2 Interoperability Report 1

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: | D6.2 |
| Deliverable title: | Interoperability Report 1 |
| Due date of deliverable: | June 2010 |
| Actual submission date: | 8th September 2010 |
| Authors: | M.L. Dubernet |
| Work Package no.: | WP6-JRA1 |
| Work Package title: | Interoperability |
| Work Package leader: | KOLN/CNRS |
| Lead beneficiary: | KOLN/CNRS |
| Dissemination level: | PU |
| Nature: | Report |
| No of pages (incl. cover): | |

| Abstract | The objective of D6.2 is to describe VAMDC Interoperability Report for Period 1. This report corresponds to Activities in WP6: JRA1 "Interoperability". 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 | WP6 Draft | M.L. Dubernet, L. Nenadovic |
| V0.2, V0.3, V0.3, V0.5 | 31/08/2010 | Updates with inputs from various sources | L. Nenadovic |
| V0.6 | 02/09/2010 | Update on Solid Spectroscopy | B. Schmitt |

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

# 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 )***

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 Tasks Plan for Period 1

Full task activities are detailed at the VAMDC wiki on the WP6/JRA1 pages – see http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/WP6

**Task 1:** Data Models and XML schema Documents

- 1.1 – Investigation of XSAMS – Leader CNRS/LPMAA with all participants -
  - Presentation of XSAMS at Kick-Off (M4)
  - Investigation of Molecular Parts by Partners (M4-M9)
  - Produce document where modifications to XSAMS is proposed by M9 –
- 1.2 – Schema for Solid Spectroscopy - Leader CNRS/LPG with internal and external partners (EuroPlanet RI FP7 project)
  - Organisation of Solid Spectroscopy Workshop to discuss Schema (M7)
  - Draft documents M9; M11
- 1.3 – Schema for PAH – Leader INAF + CNRS/CESR
  - Draft documents M9; M11
- 1.4 – Schema for Line Shapes related to pressure broadening – Leader UCL with participants CNRS/GSMA, ICB, IAO, CFA
  - Draft documents M9; M11
- 1.5 – Schema for Atomic Line Shapes – Leader AOB + CNRS/LERMA
  - Draft documents M11

**Task 2:** Dictionaries Documents

Drafts of lists will be provided in M9 and updated in M11

- 2.1 List of Species Names: make decision about standardized names
- 2.2 List of Processes: complete the list of processes given in XSAMS
- 2.3 List of Conventions
- 2.4 List of Quantum Numbers: complete the list of processes given in XSAMS

**Task 3:** Access Protocols and Query/Retrieval Language Documents

- 3.1 – Understand existing protocols – Leader CNRS/LPMAA with UU
  - Draft Documents M9, M11
- 3.2 – List of Requirements Application Fields in connection with User/Producer requirements of WP2 – Leader UU with CNRS/LPMAA

        o  Draft Documents M11
- 3.3 – Make Proposition – Leader UU with CNRS/LPMAA
  - Draft Documents M11

**Task 4**: Registries Documents

- 4.1 – Understand existing protocols – Leader CNRS/LPMAA with UU
  - Draft Documents M9, M11
- 4.2 – List of Requirements Application Fields in connection with User/Producer requirements of WP2 – Leader UU with CNRS/LPMAA
  - Draft Documents M11
- 4.3 – Make Proposition – Leader CNRS/LPMAA with UU
  - Draft Documents M11

The documents will be available within the consortium during the whole project on the private part of our web-site. We plan a public release of draft/final versions on our public web-site at Months 12

## 5. WP6 Tasks Reports for Period 1

### *VAMDC*
### Periodic Report Template (per Workpackage)

**Period**: 01/07/2009 – 30/06/2010
**WorkPackage:** WP6 Interoperability
**WorkPackage Leader and co-Leader**: M.L. Dubernet (CNRS: LPMAA/LUTH) /  S. Schlemmer (KOLN)
Participants in the WorkPackage: All Partners

Part 1

A summary of progress towards objectives and details for each tasks

**Task 1: Data Models and XML schema Documents**

The current data models and XML schema are a description of atomic and molecular linelists (SSLDM, http://www.ivoa.net/Documents/SSLDM/) for use in an astrophysical context and a description of atomic and molecular elementary processes (XSAMS, http://www-amdis.iaea.org/xsams/). XSAMS has been developed since 2004 by an international collaboration project currently including the following institutions: IAEA, NIST (USA), ORNL (USA), University Pierre et Marie Curie (France), Observatoire de Paris (France), and RFNC-VNIITF (Russian Federation).

On Month 4, XSAMS was presented at the VAMDC kick-off meeting and was accepted as the underlying schema for the exchange of data in VAMDC. It was agreed that the schema would be implemented or studied at few nodes in order to find bugs, missing or inadequate descriptions.

Different meetings have been organised between partners in order to work on JRA1,

JRA2, JRA3 (Paris, UU, KOLN) and these meetings have been extremely productive .

The schema was implemented (WP4-T1) on HITRAN by C. Hill (UCL), on CDMS by C. Endres (KOLN), on VALD by T. Marquardt (UU), on BASECOL by M. Doronin (CNRS/LPMAA). A software prototype was developed in WP8-T2 by L. Nenadovic (CNRS/LPMAA) that used instances of the schema including XSAMS outputs from HITRAN, CDMS and BASECOL. Implementations and software developments allowed the authors to make lists of comments and propose changes.

The most striking changes were proposed by C. Hill (UCL) proposing to replace the current molecular part of XSAMS version 0.1 by a so-called "case-by-case" approach. This "case-by-case" approach has been implemented on HITRAN, CDMS, BASECOL and in the prototype software of WP8-T2. This allowed the authors to comment on the "case-by-case" description of XSAMS description of molecular states; and mutual interaction led to the current version of "case-by-case" approach - for more details, see this report on the trial implementation of the case-by-case approach for the HITRAN database. Documentation for the quantum numbers and symmetries used in the cases themselves are at http://www.ucl.ac.uk/~ucapch0/QNdesc.html and will be accessible through the resource being developed by G. Rixon and C. Hill.

The schema was studied by INAF/Cagliari and CNRS/CESR in order to include PAH descriptions and a document has been edited in order to collect all necessary requirements to describe PAHs. Several meetings have been organised in order to discuss those aspects.

The schema was studied by UCL (C. Hill) for inclusion of molecular pressure broadening calculations and by both AOB and CNRS/LERMA for inclusion of atomic line shapes (meeting JRA in Feb. 10 in Paris) and preliminary propositions were given.

A meeting is planned in Paris on 23nd of July in order to make final decision about molecular part of XSAMS and other aspects.

**Task 2: Dictionaries Documents**

In plans of Cycle One, different dictionaries have been identified: List of Species, List of Processes, List of conventions, List of Quantum Numbers. Other lists have been identified through the course of the project in particular a list of molecular "cases".

**Task 3: Access Protocols and Query/Retrieval Language Documents**

Access Protocols and Query/Retrieval Language Documents have been discussed during the JRAs meetings.

We looked at the existing protocols currently available in other fields such as the IVOA.  Some of them could be used directly in VAMDC, i.e. TAP (Table Access Protocol) and SLAP (Simple Line Access Protocol). As mentioned in WP4 Report, TAP protocol has been implemented on few databases.   SLAP is available on BASECOL and CDMS.

Extension of those protocols include the definition of a VAMDC TAP, so called TAP-XSAMS.

A census of database queries has been created and compiled by T. Luftinger (UNIVIE). This allows to have an overview of the key words for query/retrieval language.  We  considered SQL/ADQL with standard operands  and an XML vocabulary for tightly-focused queries.  In the first stage a subset of SQL92, the so called VSS1 (VAMDC SQL - subset 1), was defined for use by VAMDC.

**Task 4: Registry Documents**

At JRAs meetings, Annual Meeting, the registries were explained by M. Doronin (Integrating BASECOL database into VAMDC infrastructure), G. Rixon (Registry) and K. Benson.  Detailed information on the VAMDC registry and links to relevant documentation have been provided by K. Benson and can be found on the wiki page: http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/RegistryInfo

Significant results (Activities and Deliverables)

### 1.1  Investigation of XSAMS

XSAMS were presented at the Kick-Off meeting and accepted as the preliminary standard for the VAMDC data model.

There was active discussion about XSAMS within the VAMDC community and exchanges are documented on the wiki.

Investigation of the molecular part of XSAMS was conducted by partners and has been discussed in various workshops (Paris, UU, KOLN).  Two different modifications were considered:

1   flattening of the xml structure of the molecular part – this was discussed at the International Atomic Energy Agency (IAEA) International Nuclear Data Committee (INDC) meeting on XSAMS in Japan in March 2010, and a modifications to the schema were proposed.

2   case-by-case by case approach – which is described in detail at http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/CaseByCase

Proposed modifications were collected and summarised.

A meeting is organised for 23-24 July 2010 in Paris, France, where the schema used by VAMDC will be finalised, based on the modifications proposal.  Complete list of modifications to XSAMS will be documented and implemented in cycle two .

### 1.2  Schema for Solid Spectroscopy

Following a meeting of an ad-hoc expert group of European solid spectroscopy data producers (held in Grenoble on 13th January 2010) who discussed the necessary (and optional) improvements and extensions of the Solid Spectroscopy Data Model (SSDM), it was first extended by:

- addition of keywords to describe additional spectroscopic techniques and their

spectroscopic products, new types of samples and new categories of materials.

- redefining structure of key-words as well as their type and format, taking into account the various recommendations of the expert group. In particular many new group of keywords have been defined to describe almost the whole set of samples that are measured by the European data producers. Major changes in the logical description were necessary to obtain a common description with a single data model.
- restructuring of the SSDM into tables organized with logical inter-dependence links in order to be implemented in a relational database.

The model is described in the SSDM specification document.

Further improvements of the structure are currently under way. The core of SSDM will be finalized, fully documented and implemented in cycle two. Several extensions of SSDM (Minerals, Solid vibration modes, …) will continue to be developed.

## 1.3 Schema for PAH – Leader INAF + CNRS/CESR
The XSAMS schema draft was studied in detail, side by side with all data in current and foreseen PAH database. An internal workshop INAF-Cagliari + CNRS/CESR was organised to converge on the apparent shortcomings of the current schema implementation (with respect to PAHs), and to propose some extensions to address them. This was summarised in a document.

## 1.4 Schema for Line Shapes related to pressure broadening
The schema was studied by UCL (C. Hill) for inclusion of molecular pressure broadening calculations (meeting JRA in Feb. 10 in Paris). Preliminary propositions were given, but no document is currently available as the main issue of the collaboration was to decide whether "case-by-case" approach was the schema to adopt definitively.

## 1.5 Schema for Atomic Line Shapes
The schema was studied by both AOB and CNRS/LERMA for inclusion of atomic line shapes (meeting JRA in Feb. 10 in Paris). As for task 1.4 above, preliminary propositions were given but no document is currently available.

## 2.1 List of Species
The list of Species was aimed at being able to query databases with standardized names. Ken Smith (QUB) worked on defining unique way of identifying species. He presented his work at the Annual Meeting, and delivered a first draft set of database tables of species extracted and converted from the UMIST database. Nevertheless in the JRAs meetings, it appeared that common view was to use IUPAC International Chemical Identifier (InChI) numbers in order to uniquely identify species. M. Doronin (CNRS/LPMAA) has implemented InChI numbers in BASECOL and has compiled a list of recommendation concerning the use of InChI numbers. In parallel, Ken Smith investigated the use of standard InChI numbers and InChIKeys , and explored various tools (e.g. Cactus, OpenBabel) to facilitate the production of these identifiers from existing data. InChI has some shortcomings, one in particular being that there is no mechanism to represent conformers, another being the difficulty in identifying the correct InChI with little more information than the chemical formula of a species.

These will need to be addressed in Cycle Two if use of InChI is agreed.

## 2.2   List of Processes
The lists of processes in version 0.1 of XSAMS is at this point considered complete and no new additions were proposed.

## 2.3   List of Conventions
The list of conventions is not yet completed as it relies on the choice that will be made for XSAMS.

## 2.4   List of Quantum Numbers
The list of quantum numbers for the "case-by-case" approach has been defined.  See http://www.ucl.ac.uk/~ucapch0/QNdesc.html for full definitions.

## 3.1   Understand existing protocols
Existing protocols currently available in other fields which have been considered for use in VAMDC are:
- Table Access Protocol (TAP)
- Simple Line Access Protocol  (SLAP).

## 3.2   List of Requirements Application Fields in connection with User/Producer requirements of WP2
Census of databases in VAMDC has been conducted to investigate the ways in which they are queried and the results are available:
- results of databases census

## 3.3   Make Proposition
TAP-XSAMS has been developed as the web-service protocol for use in VAMDC:
- TAP-XSAMS specification

A subset of SQL92 was proposed as the query language for VAMDC:
- VAMDC SQL - subset 1

## 4.1   Understand existing protocols
The registry is a database and web service holding metadata that describe the VAMDC resources. By holding the metadata in a central point, we avoid the need to hard-code and maintain them in multiple applications, thereby making the VAMDC facilities easier to sustain.

At the start of the project we resolved to use IVOA's model for a registry of metadata. This gives us immediate compatibility with the astronomical Virtual Observatory and we considered the model capable of adaptation to VAMDC's needs.

IVOA provides several documents, defining standards and recommendations for registries:
- Resource Identifier - defines format for unique resource identifiers.  Every record in registry has it's own RI and can be referenced by it.
- RegistryInterface - defines registry service interaction with client software. Other registries search over resource metadata and exchanging data between local and global registries (harvesting).

D6.2 Interoperability Report 1

- **VOResource XML schema** - basic XML schema used for describing resources.
- **Resource Metadata** - explains resource description fields available in VOResource schema and the concepts behind them.
- **VODataService XML schema extension** - provides additional fields required to describe data access service.

A test VAMDC registry has been created as part of WP4 and can is available at http://registry.vamdc.eu/ .

**4.2 List of Requirements Application Fields in connection with User/Producer requirements of WP2**
The Registry Info page contains detailed information on registry use in VAMDC, links to current implementation with relevant documentation, installation and configuration instructions and proposals for future tasks.

**4.3 Make Proposition**
A test VAMDC registry is available at http://registry.vamdc.eu/ . For more details please see the Registry Info page .

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

No deviation from Contract – apart from Leadership of WP6 handled by CNRS because of long standing experience with standards -

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

No failure to achieve critical objectives –
Some items are delayed with respect to Plans in Period 1
1.4 and 1.5 - Preliminary propositions were given, but no document is currently available as the main issue of the collaboration was to decide whether "case-by-case" approach was the schema to adopt definitively.
2.3 - The list of conventions is not yet completed as it relies on the choice that will be made for XSAMS.

Proposed corrective actions (if applicable)

NOT APPLICABLE

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