



VAMDC

Virtual Atomic and Molecular Data Centre

D8.1

–

Mining/Integration Plan

Version 0.3

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	ISРАН	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	IVIC	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 supported by funding from the xxxxx Programme under the 7th Research Framework Programme of the European Union

Document

Deliverable number: D8.1
Deliverable title: Mining/Integration Plan
Due date of deliverable: September 2009
Actual submission date:
Authors: J. Tennyson, D.J. Witherick, M.L. Dubernet
Work Package no.: WP8-JRA3
Work Package title: New Mining and Integration Tools
Work Package leader: UCL
Lead beneficiary: UCL
Dissemination level: PU
Nature: Report
No of pages (incl. cover):

Abstract	The objective of D8.1 is to describe VAMDC New Mining and Integration Tools Plan on PM3. This plan corresponds to Activities in WP8: JRA3 “New Mining and Integration Tools”. This plan is included in the VAMDC Project Plan.
----------	--

Versioning and Contribution history

Version	Date	Reason for modification	Modified by
V0.1	21/10/2009	WP8 – no plan for period 1 – no name	M.L. Dubernet
V0.2	17/12/2009	WP8 – task leaders and period 1 plan for task 3	D.J. Witherick
V0.3	22/02/2010	WP8 – Cycle 1 plan for Task 2	M.L. Dubernet

Final Version (vx.x) released by		Circulated to	
Name	Date	Recipient	Date
M.L. Dubernet	04/03/2010	Mrs Asero	04/03/2010

Disclaimer: The information in this document is subject to change without notice. Company or product names mentioned in this document may be trademarks or registered trademarks of their respective companies.

All rights reserved:

The document is proprietary of the VAMDC consortium members. No copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights.

This document reflects only the authors' view. The European Community is not liable for any use that may be made of the information contained herein.

WP8 ACTIVITIES DESCRIPTION

Work package number	8		Start date or starting event:					3			
Work package title	JRA3: New mining and Integration Tools										
Activity Type	RTD										
Participant id	1	3	7	12							
Person-months per beneficiary: (Total = EU + Node Contributions)	12	36	18	6							

Table of Content

WP8 activities description.....	5
1. WP8 Objectives.....	5
2. WP8 Milestones and Deliverables	5
3. WP8 Tasks Description.....	6
4. WP8 Tasks Description for Period 1.....	7

1. WP8 Objectives

This JRA will develop extensions to the baseline infrastructure. Key objectives are the design of advanced data mining tools and the design of cross-matching and cross-federating tools, providing sophisticated integrated science services aimed at maximising the scientific utility to the end user community of the VAMDC services.

WP8 Leader is UCL(3)

2. WP8 Milestones and Deliverables

Milestones

M8.1	Technical Meetings	WP8	UU	Months 5,10, 16, 22, 28, 34, 40, 42	Minutes. Presentations on internal Website
M8.2	Evaluation of softwares	WP8	UU	Months 10, 22, 34	

Deliverables

- D8.1 Mining and Integration Tools Plan (PM 3)**

D8.2 Mining and Integration Tools Report to be included in report to the EU – Year 1 (PM 10)

D8.3 Mining and Integration Tools Report to be included in report to the EU – Year 2 (PM 22)

D8.4 Mining and Integration Tools Report to be included in report to the EU – Year 3 (PM 34)

D8.5 Final Report of Mining and Integration Tools to be included in final report to the commission (PM41)
Annual Mining & Integration Plan revisions included in Revised Annual VAMDC Project Plans – Year 1,2,3

3. WP8 Tasks Description

WP8 Leader (co)		
Task Number	Leader	Other Partners
1	M. Doronin (CNRS: LPMAA)	RFNC-VNIITF
2	S. Schlemmer (KOLN)	CNRS:LPMAA
3	J. Tennyson (UCL)	UCL/MSSL

Description of work (possibly broken down into tasks)

Through the activities of JRA1 and JRA2, the AM resources will be searchable and will provide information in a standardised way. The following step is to build the query protocols that will access those published AM data and then to design software that will handle and process those data.

Task1: Registry Queries (lead by CNRS(1) with (12))

We will need to use protocols to query the registries at a fine level of granularity. Indeed we don't wish to only find resources having implemented a type of service such as SSAP or TAP, but rather be able to select resources according to their content through key words. The purpose of Task 1 is to implement those protocols.

Task 2: Tools for Manipulation of Data (lead by KOLN(7) with (1))

Our queries will return data organised according to schemas defined in JRA1. Those schemas will be quite complex because they will reproduce all the scientific concept attached to the data. Therefore the handling of the XML files will be complex and will require specific tools. For now we identify two main generic tools: one performing cross-matching of data and one performing cross-federation of data. These tools are particularly difficult because they require to compare the content of many fields in the schema. Those generic tools will be made available for download in SA1 to the end users and developers. Support to adapt those tools to specific applications will be provided in SA2. We plan to provide libraries to allow users to develop their own applications

Task 3: VAMDC advanced data mining services (lead by UCL(3))

With the deployment of a vast range of high value data services through the standard VAMDC infrastructure, this task will investigate optimal strategies to best mine these AM data resources to both advance the creation of new AM fundamental data, and by providing stream lined automated access to appropriate AM data targeted at specific user groups (for the astronomy community benefiting from the availability of high energy data from satellites such as Swift, XMM, Chandra, who require specific atomic data for high excitation species of elements such as iron). This task would investigate the provision of application services wrapping complex work flows combining AM data access, manipulation, and integration into user processing chains – e.g. in solar physics, astro-biology/ chemistry and so forth.

4. WP8 Tasks Description for Period 1

Full task activities are detailed at the VAMDC wiki on the WP8/JRA3 pages – see <http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/WP8>

Task 1: Registry Queries

- This task is not due to start until Cycle 2

Task 2: Tools for Manipulation of Data

- Make Prototype of Data cross-identification software based on current XSAMS schema (link with JRA1/JRA2) – Initial Test for Cycle 1 will be between BASECOL and CDMS databases.

Task 3: VAMDC advanced data mining services

- Development of use cases for workflows based around e-HITRAN
 - Identify current and future users/data miners of the HITRAN database and its upcoming successor, e-HITRAN
 - Survey the current and future HITRAN/e-HITRAN users as to how they do/would mine the information contained in the database
 - Develop a series of use cases based upon the responses of users
- Development of technical requirements of workflows
 - Begin dialogue with MSSL on the technical requirements for workflows to support the use cases