bmail

VAMDC

Virtual Atomic and Molecular Data Centre

D7.5

–

Final Publishing Tools 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	CTPM	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:	D7.5
Deliverable title:	Final Publishing Tools Report
Due date of deliverable:	December 2012
Actual submission date:	January 2013
Authors:	U. Heiter, N. Piskunov, M.L. Dubernet and WP7 team
Work Package no.:	WP7-JRA2
Work Package title:	Publishing Tools
Work Package leader:	UU
Lead beneficiary:	UU
Dissemination level:	PU
Nature:	Report
No of pages (incl. cover):	

Abstract	The objective of D7.5 is to describe VAMDC Science/Technical Report during the project. This report corresponds to Activities in WP7: JRA2 "Publishing Tools".

Versioning and Contribution history

Version	Date	Reason for modification	Modified by
V0.1	**December 2012**	**Preparation of Document**	**M.L. Dubernet**
V0.1	**Dec 2012**	**WP7 Report**	**N. Piskunov**
V1.0	**January 2013**	**D7.5 document**	**M.L. Dubernet**

Final Version (v1.0) released by		Circulated to	
Name	Date	Recipient	Date
M.L. Dubernet	12th February 2013	Mr Bodo	12th February 2013

WP7 ACTIVITIES DESCRIPTION

Work package number	7		Start date or starting event:		3					
Work package title	JRA2: Publishing Tools									
Activity Type	RTD									
Participant id	1	6	8	12	13					
Person-months per beneficiary: (Total = EU + Node Contributions)	12	12	12	5	24					

Table of Content

1. WP7 Objectives

JRA2 provides generic tools for A&M data in the VAMDC infrastructure. We envisage two publishing paths: (A) inclusion of the new data into existing databases, which are already integrated into the VAMDC, and (B) publishing data in a new database instance using an existing open source relational database software instrumented with the VAMDC interface.
The selection of specific path will be decided between the data producer and the VAMDC depending on the completeness of the new data, logical and structural similarity with existing databases, availability of the necessary resources at the producer site etc.

The process of publishing includes two steps: data quality assessment and the open offering of the new data to the VAMDC clients. In the first stage the new data will be technically included into the infrastructure but under the test version of the Portal. A group of experts working with the VAMDC will then verify the quality and integrity of the data. After the quality assessment is completed the data will be offered through the public Portal.

In order to implement this plan we have developed a concept and implemented the Node Software that allows quick and easy deployment of the new nodes based on the data model or addition of data to an existing database equipped with Node Software. The work was built on long experience accumulated by large databases handling many datasets from different data providers (e.g. VALD). The tasks in the WP7 are closely related to other work packages and were performed in close and intensive collaboration with the rest of the consortium.

WP7 Leader is UU(6).

2. WP7 Milestones and Deliverables

Milestones

M7.1	Technical meetings	WP7	UU	Months 5, 10, 16, 22, 28, 34, 40, 42	Minutes. Presentations on internal Website
M7.2	Evaluation of software	WP7	UU	Months 10, 22, 34	

Deliverables

D7.1 Publishing Tools Plan (PM 3)
D7.2 Publishing Tools Report to be included in report to the EU – Year 1 (PM 10)
D7.3 Publishing Tools Report to be included in report to the EU – Year 2 (PM 22)
D7.4 Publishing Tools Report to be included in report to the EU – Year 3 (PM 34)
D7.5 Final Report of Publishing Tools to be included in final report to the commission (PM41)

Annual Publishing Tools Plan revisions included in Revised Annual VAMDC Project Plans – Year 1,2,3

3. WP7 Tasks Description

WP7 Leader (co)	N. PIskunov (UU)	
Task Number	Leader	Other Partners
1	M. Doronin (CNRS)	All others
2	M. Doronin (CNRS)	All others
3	N. Piskunov (UU)	All others
4	P. Loboda (RFNC-VNIIT)	All others
5	A. Fazliev (IAO)	All others

Description of work (possibly broken down into tasks)

This WP developed the Node Software that is deployed with the majority of databases included in VAMDC infrastructure. This software is fully compliant with the standards developed in JRA1. The Node Software is available to the data producers and database hosts via the VAMDC website (http://www.vamdc.eu/data-providers/software) together with documentation and examples. Two alternative solutions for publishing data in VAMDC were implemented. The first offers a possibility to import new data to the existing informational resource, capable of importing data in standardized forms (tasks 2-4) developed within JRA1. The second option implies the design of a self-contained information system accessible via the Internet (tasks 3-4) instrumented with the VAMDC integration tools developed in tasks 1-3. In this variant an automatic generation of semantic metadata for uploaded information resources is realized, taking into account the restrictions imposed by formal models of molecules and atoms.

Task 1: Create/adapt tools to go from a DM/XML schema to a full database deployment with generation of automatic administrative interface. **(lead by CNRS(1) with (6))**

Task 2: Create/adapt tools to build registries from the content of databases **(lead by CNRS(1) with (6))**

Task 3: Create/adapt interfaces to easily update dictionaries **(lead by UU(6) with (1))**

Task 4: Develop software libraries using various languages allowing to easily generate output of already existing resources in standardized format (lead by RFNC-VNIIT(12) with (1), (6), (8))

Task 5: Create tools to upload, modify, retrieve, compare, visualize and publish information in molecular spectroscopy (lead by IAO(13))

4. WP7 Final Tasks Reports

Period: 01/07/2009 – 31/12/2012
WorkPackage: WP7/JRA2 Publishing Tools
WorkPackage Leader and co-Leader: N. Piskunov/U. Heiter
Participants in the WorkPackage: CNRS (1), **UU (6)**, INAF (8), RFNC-VNIITF (12), IAO (13)

Part 1

A summary of progress towards objectives and details for each task

The goal of this WP was to provide generic tools in order to help producers of A&M data to publish their data in VAMDC. The development Publishing Tools (PTs) followed two paths: constructing a new VAMDC database from scratch or adding new data to an existing database. PTs are tightly integrated with the rest of the infrastructure and must be compliant with the adopted standards. Thus the WP7 activities also included the development of the standards and the Node Software and training the data providers. The development of the PTs was completed in Period 3. Period 4 was dedicated to insuring the consistency with the latest release of the standards, completion of documentation and training/dissemination. WP7 was divided in 5 tasks.

Task 1: Create/adapt tools to go from a DM/XML schema to a full database deployment with generation of automatic administrative interface. This task resulted in development of the Node Software that provides the interface between the user portal, consumer conversion tools or user applications and VAMDC databases. The Publishing Tools are based on the Node Software combined with tutorials and examples on how to add new A&M data to VAMDC. The current version of the Node Software and PTs is consistent with the latest VAMDC standards. PTs were presented to data producers in a number of tutorials across Europe. Tutorials and examples will be added to the VAMDC web pages.

Task 2: Create/adapt tools to build registries from the content of databases. PTs are fully integrated with the Node Software and thus ingesting new data into an existing node is automatically reflected in the VAMDC registry. A procedure for registering a newly created node was detailed at the VAMDC web pages. The concept of the mirror site was implemented and tested allowing the data access even if the primary site is unavailable.

Task 3: Create/adapt interface for updating the dictionaries. Dictionaries have been updated according to the latest release of the query language. The inspection/editing/verification tool for the dictionaries was moved to the VAMDC web site.

Task 4: Develop software libraries using various languages for converting already existing resources to standardized format. The Node Software is now available in Python and Java versions. Spectr W^3 database was used as a test bench for development XSAMS generation libraries in C.

Task 5: Create tools to upload, modify, retrieve, compare, visualize and publish information in molecular and atomic spectroscopy. This task gradually focussed on the development of an ***Automatic Data Verification*** tool. The prototype of the tool compatible with the XSAMS data model was developed by IAO and tested with the VALD node. This software is complemented by a formal description of quantum-mechanical selection rules for atoms and some diatomic molecules and was tested with a number of VAMDC nodes. The final report for the new tool can found at: http://voparis-twiki.obspm.fr/twiki/pub/VAMDC/Jra2T5/Automatic_Data_Verification-Tool-V.0.20.pdf.

Significant results (Activities and Deliverables)

1. We have defined the plausible alternatives for publishing new data in VAMDC infrastructure. We have also assessed the available tools, web and database software that could be reliably supported or interfaced. The resulting plan (D 7.1 "Tools to publish A&M Data") was published on the VAMDC wiki and used throughout the project.
2. Publishing Tools software was presented to the expert group, first as a prototype and then as a fully functional subsystem of VAMDC. For more details, see "D 7.2 Summary of experts report".
3. Publishing Tools software releases are freely available through Github repository with the latest one up-to-date with the VAMDC standards and the rest of the infrastructure components. The documentation is also available on the web linked through the VAMDC web page. This completes the deliverables "D 7.3 Public release of Tools on Public Website" and "D 7.4 Revised Tools to publish A&M Data Plan".

A gentle introduction for data producers with extensive links to tutorials, examples and contact points is linked from the VAMDC website.

Deliverables to EU

D7.1 Publishing Tools Plan
See http://www.vamdc-project.vamdc.eu/public-deliverables/19-deliverables-wp7

D7.2 Publishing Tools Report to be included in report to the EU – Year 1 – DONE
http://www.vamdc-project.vamdc.eu/public-deliverables/19-deliverables-wp7

D7.3 Publishing Tools Report to be included in report to the EU – Year 2 – DONE
http://www.vamdc-project.vamdc.eu/public-deliverables/19-deliverables-wp7

D7.3 Publishing Tools Report to be included in report to the EU – Year 3 – DONE
http://www.vamdc-project.vamdc.eu/public-deliverables/19-deliverables-wp7

Annual Publishing Tools 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

Key internal deliverables were software for each of the 5 tasks. Task 1-4 deliverables are distributed on the VAMDC web site as part of the Node Software (at http://www.vamdc.eu/data-providers/software). The current standards, registry, units and data model are described at http://www.vamdc.eu/data-providers/standards.
The IAO web site (Task 5) [the general page is s http://wadis.saga.iao.ru/ and the direct link to the tool is at http://wadis.saga.iao.ru/saga2/transition/comp2/).

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

n/a

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

n/a

Proposed corrective actions (if applicable)

n/a