SUP@VAMDC

Support at the Virtual Atomic and Molecular Data Centre

D2.3

–

VAMDC Consortium Roadmap

Version 3.0

Grant agreement no: 313284

Combination of Coordination and Support Actions

Project Information

Project acronym: SUP@VAMDC

Project full title: Support at the Virtual Atomic and Molecular Data Centre

Grant agreement no.:   313284

Funding scheme: Combination of Coordination and   Support Actions

Project start date: 01/12/2012

Project duration: 24 months

Call topic: INFRA-2012-3.3 Scientific Data Infrastructure

Project web sites: http://www.sup-vamdc.vamdc.eu

                   http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/SUP-VAMDC

Consortium:

This project is supported by funding from the Scientific Data Infrastructure Programme under the 7th Research Framework Programme of the European Union

Document

Deliverable number: D2.3

Deliverable title: VAMDC Consortium Roadmap

Due date of deliverable: November 2014

Actual submission date: December 2014

Authors: The SUP@VAMDC collaboration.

Work Package no.:   WP2

Work Package title: Policies and Strategy

Leader: OBSPARIS, OU

Lead beneficiary: OBSPARIS, OU

Dissemination level: PU

Nature: Document

No of pages (incl. cover):

List of participants:

Beneficiary Number *	Beneficiary name	Beneficiary short name	Country	Date enter project**	Date exit project**
1(coordinator)	Observatoire de Paris[1]	OBSPARIS	France	Month 1	Month 24
2	The Chancellor, Masters and Scholars of the University of Cambridge	UCAM	UK	Month 1	Month 24
3	Uppsala Universitet	UU	Sweden	Month 1	Month 24
4	Open University	OU	UK	Month 1	Month 24
5	Universitaet zu Koeln	KOELN	Germany	Month 1	Month 24
6	University College London	UCL	UK	Month 1	Month 24
7	Korea Atomic Energy Research Institute	KAERI	Korea	Month 1	Month 24
8	Institute University of South Africa	UNISA	South Africa	Month 1	Month 24
9	Tata Institute of Fundamental Research	TIFR	India	Month 1	Month 24

In addition to the legal Beneficiaries, SUP@VAMDC has external partners who are supported to attend SUP@VAMDC main events and an associate member with no support. Detailed explanations are given in the implementation section of Description of Work. Below is the list of those partners.

Name	Institute	Status
Dr Yuri Ralchenko	Atomic Spectroscopy Group, NIST, USA	External
Dr Larry Rothman	The Harvard-Smithsonian Center for Astrophysics, Cambridge, USA	External
Dr Brian Drouin	The Propulsion Laboratory, NASA	External
Dr Carlos Gonzales	The Chemical and Biochemical Reference Data Division	External
Prof. Stephen Buckman	Atomic and Molecular Physics Laboratory, Canberra, Australia	External
Prof. Michael Brunger	The Electron Scattering and Modelling Group, Flinders University, Australia	External
Dr Izumi Murakami	National Institute for Fusion Science, Japan	External
Prof. Milton Fujimoto	Universidade Federal of Parana, Brazil	External
Dr Bas Braams	Atomic and Molecular Data Unit, IAEA, Vienna	Associate

Abstract	D2.3 describes the Roadmap of the VAMDC Consortium for the next 3 to 10 years

Versioning and Contribution history

Version	Date	Reason for modification	Modified by
v1.0	15-05-2014	1rst Draft	M.L. Dubernet
v1.1		Re-organisation of Document; Philosophy; res-shape business plans, include research, business, outreach, education activities, technical organisation, marketing activities	M.L. Dubernet
v1.2	23-05-2014	Annotations - Changes UCL	C. Hill & J. Tennyson
v1.3	23-05-2014	add item in intro; add section 6.1.3 TBD	CM Zwölf, U. Heiter
v1.4	25-05-2014	Final Check for Year 1 and release to EU	M.L. Dubernet
v2.0	10-06-2014	Business Plan for Industry	J. Tennyson, C. Hill
v2.1	12-06-2014	Technical Organisation	ML Dubernet, C.M. Zwölf , G. Rixon
v2.2	08-07-2014	International Collaboration	N. Mason
v2.3	10-07-2014	Consistency check	M.L. Dubernet, U. Heiter
v2.4	21-07-2014	Last check	C.M. Zwölf, M.L. Dubernet
v2.5	04-11-2014	Remove Items in Education following Remarks from Review Committee + add items related to their comments (Oct. 2014)	M.L. Dubernet
v2.6	25-12-2014	Re-arrange section on Technical Organisation to a lighter version following	M.L. Dubernet

		approval of minutes of BoD of VAMDC Consortium	
v2.7	05-12-2014	Inclusion of description of software, simplification of international section	M.L. Dubernet
v2.8	06-01-2015	Improvement Communication section	M.L. Dubernet
v2.9	15-01-2015	Inclusion of N. Piskunov text	N. Piskunov (circulated)
v3.0	23-01-2015	Polish - Final Check	M.L. Dubernet

Final Version (3.) released by		Circulated to	
Name	Date	Recipient	Date
M.L. Dubernet	26-01-2015	Mrs Kalfin	26-01-2015

TABLE OF CONTENT

Executive summary

Atomic and molecular (A+M) data are of critical importance across a wide range of applications such as astrophysics, atmospheric physics, fusion, environmental sciences, combustion chemistry, health and clinical science including radiotherapy and underpin a range of industries ranging from technological plasmas to lighting. Accordingly in the past decade the wider research community has appreciated the need to collate and make available the A+M data that describes fundamental atomic and molecular processes recognising how access to such data is central to achieving scientific breakthroughs across a range of disciplines. However such increasing demands by the research community for large amounts of A+M data present major challenges to the expert research teams in Europe, the USA, Asia and elsewhere that measure, derive and collate such data as demand outstrips supply. The interface between the producers of A+M data and the wide body of users of that data has therefore been a major bottleneck, slowing discovery and hence slowing economic growth. The VAMDC e-infrastructure (http://www.vamdc.eu), funded by the FP7 "Research Infrastructures - INFRA-2008-1.2.2 - Scientific Data Infrastructures" initiative programme in July 2009 for 42 months (project 239108), was developed to remove this bottleneck by designing/implementing interoperable protocols among a wide range of disparate A+M databases and providing a single portal through which users can access A+M data from those databases whilst providing data providers and compilers a large dissemination platform for their work.

The Objectives and Activities of the "VAMC Consortium" follow from the activities and objectives of the European FP7- VAMDC and SUP@VAMDC projects. The "VAMDC Consortium" pursues the following objectives:

➢ Cooperating on science and technology for promoting the interoperable exchange of atomic and molecular data within the VAMDC infrastructure
➢ Developing and promoting a strategy for increasing the impact of the VAMDC infrastructure towards academics, industry, education, public sector ;

The "VAMDC Consortium" will be involved with the following activities in order to fulfil its objectives:

- ➢ Maintenance and Evolution of the VAMDC infrastructure
- ➢ Conception and Innovation activities
- ➢ Support and Training activities
- ➢ Marketing and Communication activities linked to the Consortium's objectives
- ➢ Participation to Projects related to National and International Calls in order to full fill the Consortium's objectives
- ➢ On-demand Service activities
- ➢ Quality control of the VAMDC Infrastructure

Those activities concern 4 Divisions:

- ➢ Research
- ➢ Education
- ➢ Industry
- ➢ Outreach

The VAMDC e-infrastructure (see appendices) provides on a regular basis upgraded "VAMDC Service Release" which is a combined issue of VAMDC standards for data access, VAMDC nodes, each containing a database and web service following the standards. The current set of nodes is listed in the web portal at http://portal.vamdc.eu/. VAMDC provides software for creating a node and publishing data therein. Software such as client or provider tools are provided at http://vamdc.eu/software and further software developed for clients will be available in the future. Help for users is provided on the web through forums and documents and by email at support@vamdc.eu. The email address may be used for questions on operation, general enquiries about VAMDC or to report problems.

1 Introduction

The Roadmap has been written during the SUP@VAMDC project that involves only part of the historical "VAMDC Consortium" community. This Roadmap has circulated among the full "VAMDC Consortium" members for informal approuval, this Roadmap might be further amended once the Memorandum of Understanding is signed. The Roadmap will be assessed by the Scientific and Technical Board, and voted by the Board of Directors. It is expected that the Roadmap will be updated every year.

2 Roadmap Objectives

The Roadmap aims at giving a description of the political, organisational, activities, sustainability issues linked to the VAMDC Consortium within the following timelines:

- ➢ 3 years (specific objectives)
- ➢ 5 years (broader objectives)
- ➢ 10 years (foreseen objectives)

The roadmap document aims at providing the strategy on:

- ➢ Political & Technical Agreement

- ➢ Business Model
- ➢ Business Plan
- ➢ Management, Communication and Marketing
- ➢ Organisation of « VAMDC infrastructure »
- ➢ Our strategy towards Research-Users
    - o Research/Producers
    - o Research/Clients
- ➢ Other Users/Divisions
    - o Education
    - o Industry
    - o Outreach

3 The International Collaboration

3.1 General Business Plan

VAMDC aims to be an e-infrastructure that provides the international research community with access to a broad range of atomic and molecular (A&M) data compiled within a set of A&M databases accessible through the provision of a single portal. Furthermore VAMDC aims to provide A&M data providers and compilers with a large dissemination platform for their work. The primary objective of the VAMDC Consortium are the production, curation (including quality control) and dissemination of atomic and molecular data for application in research and industry.

Since both A&M data users and data providers are drawn from across the globe VAMDC requires international collaboration. The objectives of this international programme are:
  - To co-operate with existing databases from across the world in the interoperable exchange of A&M data and discuss access to such through the VAMDC Portal
  - To coordinate with other A&M e-infrastructures in the development and adoption of standards, tools and software that may be more widely shared with and adopted by the international community.
  - To provide training for new communities who wish to either construct new databases or assemble new datasets for inclusion within existing or new/planned databases.
  - To raise the awareness of the availability of A&M data through VAMDC in those communities where, to date, the culture for collation and dissemination of such data is limited.

These objectives will be met by a four point international strategy:
  - To develop formal links with existing databases and data networks (e.g. through their signing the MOU - see Appendices)
  - To provide training and support (in form of tutorials, workshops and the provision of material on the VAMDC website) for new databases and data networks in countries where, to date, no such infrastructure exists.
  - To disseminate the VAMDC facility and protocols across the international research community through publications and conferences as well as the provision of an active VAMDC website.
  - To participate in National and International (e.g H2020) projects where VAMDC can be deployed as a partner/beneficiary adding value to the project.

In order to ensure the international recognition of VAMDC e-infrastructure the VAMDC consortium will be represented at major A&M Conferences through its Outreach programme. In particular VAMDC will be present at the International Conference on Atomic and Molecular Data and Their Applications (ICAMDATA): a successful series of international conferences that promotes the use of atomic and molecular data in various fields of science and technology, provides a forum for interaction of AM data producers and users and for information exchange on AM data needs and availability, and fosters cross-disciplinary cooperation between the AM data producers and users.

3.2 International Data Centre Networks

The VAMDC e-infrastructure engages with International data centres which may be divided into two broad categories;

a) Operational and well established datacentres providing complementary and competitive structures to VAMDC. These include the Atomic and Molecular Data Unit at IAEA which chairs and organises the Atomic and Molecular/Plasma-Material Interaction (A+M/PMI) Data Centre Network (DCN); The NIST datacentre; The National Institute for Fusion Science (NIFS), Japan; The Korean Atomic Energy Research Institute; the HITRAN consortium and the International Virtual Observatory Alliance (IVOA) which includes some VAMDC partners within its structure.

VAMDC has already offered to those data centres or large consortium to become either full members or associated members of VAMDC, so that they could participate either (or both) to the Board of Directors or (and) to the Science and Technical Board.

b) Developmental datacentres/databases; These may include the Indian A&M datacentres initiated within the SUP@VAMDC project; Chinese datacentres as developing within the DCN and project based sets of databases (as in the RADAM initiative). VAMDC will provide advice/consultancy services and provide technical training to the database/centre engineers to these new datacenters. In this way VAMDC seeks to exploit its software, tools and protocols and embed them across the international A&M data community.

In those regions where the culture and infrastructure for building databases and forming data centres is still largely absent (e.g South America; Africa; The Balkans; Middle East) VAMDC will aim to initiate discussions within the nascent regional A&M communities. Such discussions will be coordinated in partnership with a local centre/provider (e.g. UNISA, University of South Africa, South Africa for Africa; Corporacion Parque Technologico de Merida, Venezuela and Universidade Federal Parana, Brazil for south America; Astronomska Opservatorija, Belgrade Serbia for the Balkans) and may include consultancy and training.

3.3 Data providers

VAMDC has engaged with A&M data providers since its inception, such data being core to the VAMDC service. A&M data providers may be divided into the following categories:
  - ➢ Experienced and mature data providers who produce large amounts of A&M data and regularly input such data into databases.
  - ➢ Developing providers who collect large amounts of A&M data but to date don't know how to ensure such data enters a database.
  - ➢ Fledgling providers who publish appropriate data but have little or no engagement with any database

VAMDC will work with all three categories tailoring its interaction to the maturity of the data provider.

3.4 Data users

VAMDC infrastructure was established to provide a service to the wider international research community and has been developed in conjunction with consultations and advice from the A&M user community (many of whom were existing users of the databases incorporated into the VAMDC Portal). VAMDC's initial clientele was largely the astronomy

community but during the course of VAMDC project the user community widened to include the Atmospheric science community (initially through inclusion of HITAN in the VAMDC database suite), fusion (through partnering with IAEA) and the plasma physics community (including lighting) a relationship further strengthened through addition of LXCAT database under SUP@VAMDC project. VAMDC has been particularly influential in the Radiation Chemistry community with the RADAM programme adopting VAMDC protocols and tools to build its ascent RADAM database of A&M data for inclusion in radiation track and nanodosimetry models for next generation radiotherapy.

The VAMDC consortium will encourage the establishment of Users Working groups to specifically engage with and inform A&M data users of the VAMDC infrastructure. The groups may be community based (Plasma, Combustion; Radiation damage) or by sector (Industry; Education; Citizen Scientists). The Users Working groups will be international in their scope and ambitions aiming to inform and attract A&M users from across the globe. The VAMDC consortium will therefore aim to have a presence at all the major relevant A&M user meetings including those bringing A&M data users and providers together and that are relevant to the thematic of both the already existing VAMDC connected databases and to the research areas of the VAMDC members.

3.5 Link to Publishers

Another key group of stakeholders are the academic publishers. As the major disseminator and repository of A+M data it is essential that any A+M e-infrastructure is acknowledged by and cooperates with the publishing industry. At present the challenges of open access publishing is a major concern for publishers but also an opportunity for database providers since open access publishing may make the accumulation, appraisal and dissemination of large amounts of data sets more practical (e.g. no longer requiring requests to publishers for inclusion of datasets published in their journals). However total open access to published material is not an economic model that will allow publishers to sustain their business. WP2 engaged with a set of publishers (representing both purely commercial ventures (Springer) and learned Societies (IOPP and RSC) and found exemplars of how e-infrastructures can establish working relationships with publishers including methodologies for ensuring that authors are recognised (cited) when their work is downloaded from the databases. The latter is important in demonstrating the importance of data to the funders of academic A+M research. As part of the roadmap the work related to Publishers was oriented towards collaboration with the Research Data Alliance and the inclusion of Digital Object Identifier (DOI) in our standards. This will allow easy citation of datasets and increase the impact of the VAMDC infrastructure. We will then be able to answer the request from Reuters (web of science) that contacted us in 2012 in order to be part of their citation databases. This DOI will be used in user publications in order to cite native VAMDC databases datasets and/or combination of datasets extracted from VAMDC.

3.6 Participation in New Projects

The VAMDC Consortium aims to support further research activities and longer term support from a range of national and international funding sources. Accordingly the VAMDC Consortium will horizon scan major calls for international funding that are appropriate to the deployment, inclusion and development of the VAMDC infrastructure, these include H2020 programme; support programmes for major space missions and other large scale infrastructure projects (e.g ITER and the ISS). In such large scale programmes the VAMDC Consortium

will be represented by the Board of Directors who will assess the benefit and viability of VAMDC involvement in particular programmes and participate in bid preparations as required. Smaller programmes, programmes supporting particular databases and projects supporting national programmes may be led by VAMDC members but the VAMDC Board of Directors should be informed of such programmes/proposals prior to submission. Similarly the VAMDC Board of Directors will review any industrial/commercial opportunities identified by members and act as the point of contact for any IPR and confidential agreements.

4 The "VAMDC Consortium" Political Organisation

The "VAMDC Consortium" is bound by a Memorandum of Understanding (MoU) that has been signed by 14 partners of the 2 EU-FP7 projects: "VAMDC" and "SUP@VAMDC" (see appendices), and that could welcome new signatures. The partners signing the MoU will become full members of the "VAMDC Consortium". Additional Partners are currently invited to join and others have already applied.

The "VAMDC Consortium" MoU handles the following aspects:
➢ Category of Memberships
➢ Governance with different bodies: Board of Directors, Executive Director Board, Scientific and Technical Board
➢ Voting Rules
➢ Entry into Force, Duration and Termination
➢ Responsibility of members
➢ Definition, Representation, Use of "VAMDC" Brand
➢ Financial Provisions
➢ Description of Activities
➢ Use of foreground
➢ Access Rights
➢ Intellectual Property
➢ List of "VAMDC" products (background)
➢ List of Members

A "Technical By Law " document that describes the technical requirements will be signed by all groups hosting "VAMDC resources". The groups' agreement on the "Technical by Law" will be a pre-requisite for their institution to become a full member of the "VAMDC Consortium". This Technical By Law will be part of the Internal Regulations set of documents that will describe the implementation of the MoU.

5 The "VAMDC Consortium" Scientific and Technical Organisation

The Scientific and Technical Organisation will be handled in the Scientific and Technical Board supported by Working Groups (WG) whenever necessary.

5.1 Composition

The Scientific and Technical Board is composed of the most prominent technical and scientific people actively involved in the VAMDC activities. Each full or associate member

will name one to 3 people who would be part of this Board. Members of the S&T Board gives a notice of 6 months if they plan to leave without replacement.

5.2   Chair

The Chair of the S&T Board is presently the executive director of the Consortium and this definition might be reviewed at any time if the load of work becomes too high. The Chair makes sure that the decision of the BoD are implemented and is the link between the BoD and the S&T Board. The chair proposes creation or closure of WGs to the BoD at the start of each calendar year and at any other time if necessary. The chair is in charge of the quality procedures within the VAMDC e-infrastructure.

5.3   Organisation

The range of actions that S&T Board members agree to carry out within the network is identified at the start of the calendar year and the chair contacts the relevant people whenever necessary.

In general the existing VAMDC products should have a maintenance volunteer, and the consortium records these volunteers formally, so they can be acknowledged. Organizations or individuals taking maintenance responsibility are expected to line up a successor if they have to drop out. They should inform the technical board about that issue 6 months in advance and indicate the name of the new maintainer.

For certain critical facilities - portal, registry, central website, species database and any other items that might be selected as such - the maintenance successor should be arranged from the start, and the primary maintainer should start training the successor when they first take responsibility.

The responsibility of some general tasks are assigned at the start of each calendar year:
- maintenance of the monitoring systems
- maintenance of the general website: content and upgrade
- communication networks
- maintenance of the VAMDC portal
- maintenance of registries
- leadership for upgrade of standards

Other Tasks :
- software/libraries are maintained by the owner following quality rules, i.e. packaging with documentation, technical notice, user guides, versioning, licencing, copyright
- tutorials are maintained by the people in charge of the software
- maintenance/upgrade of the nodes are under the responsibility of each node
- validation of data is under the responsibility of a group that will handle the link between the validation/evaluation networks/groups and the VAMDC Consortium
- contacts for including new comers: dispatched to node leaders and to identified persons
- coordinated actions towards users performed by individuals on the basis of their personal interest.

5.4 Functioning

The S&T Board will function through a collaborative distributed project management system (presently REDMINE) where every project will be recorded, through its mailing list, through teleconference using Adobe Connect (sharing documents) whenever needed and through technical/scientific meetings that will be organized as needed. The pace of teleconferences and meetings will be planned at the start of each calendar year. The Chair will follow the progresses through REDMINE and through the key people who are put into place. If some specific larger projects are put into place, an adequate organisation will be put into place.

5.5 General Business Plan

Initially, all technical work is contributed rather than paid for by the consortium. Later, the consortium might have funds to pay for selected technical work and full members might have preference for these contracts. If we plan to do work for hire to commercial customers, consortium members must set aside some priority time to do the work. We might plan this time quarterly. If there's no promise of time, then we shouldn't take the contracts. We don't plan to do work for hire before 2016 unless there is a particularly good offer.

5.6 Detailed Information about Tasks

In this section we give more detailed information about the different Tasks. These tasks might be organised in identified Working Groups if necessary.

5.6.1 Technical Operational Tasks

The activities cover the inclusion of new databases, the research and development activities and the monitoring of the components of the infrastructure. Participants to these tasks are representative of databases/services and additional developers/scientists interested into the development of the "VAMDC Infrastructure". They discuss the issues related to the databases/services tagged as "VAMDC", discuss the inclusion of new databases or services as "VAMDC" products, discuss the maintenance, monitoring and evolution of the infrastructure.

5.6.1.1 *New Databases Task*

The New Databases Task handles the inclusion of new databases. The executive director is currently the contact point for inclusion of new databases. He dispatches the newcomers to the full members that have similar data for an initial scientific assessment of the needs. Following this assessment, two cases are possible:

➢ the data are fully compliant with the existing infrastructure: the newcomer is offered a support through the Training and Support Task with 2 contact points (scientific and technical)

➢ the data are not fully compliant: the newcomer issue is discussed within the S&T Board

The communication between data providers and VAMDC is via the forum (with a first contact/alert via support@vamdc.eu). Once the new comer is sufficiently involved, his project becomes part of the management system and he has access to REDMINE: the discussions are then switched to internal means of communication.

5.6.1.2 *Research & Development Tasks*

The Research & Development Tasks handle the upgrade (small and large) of the standards and software of the VAMDC e-infrastructure, as well as the creation of new software. The

upgrades might be triggered by requirements from users' communities, by the inclusion of new data providers and/or by the security/technological updates of the software composing the infrastructure. The tasks include proposition of new standards, discussion about standards with other communities, development of software that have an interest to the different user communities. The tasks include the validation of the release of any new products and the edition of the relevant documentation. The progress of work is followed via the on-line project management tools and each product is under the responsibility of one person who supervises all activities linked to that product and who works in collaboration with other members.

Roles:
➢ The executive director checks that all documents are upgraded before a release is planed
➢ Each software is associated to a person (creator and/or maintainer)

5.6.1.3 Monitoring Task

The Monitoring Task handles the deployment and the maintenance of a monitoring system concerning every element of the infrastructure. Two independent monitoring systems are currently into place: one under the responsibility of Paris Observatory, the other under the responsibility of Cambridge University. All databases leaders get regular messages if some services are disrupted. If the anomalies persist, the executive director takes action.

5.6.1.4 Technical Quality Task

The Technical Quality Task handles the procedures, rules and good practices that should be followed while carrying out the technical activities. The quality procedures are defined by the executive director in collaboration with members of the S&T Board, and approved by the BoD. Those quality procedures are part of the Internal Regulations and are re-assessed every year. An iterative work will permit that the quality procedures are realistic and properly implemented. Non-exhaustive examples of such procedures are:
➢ definition of collaborative tools for technical management
➢ procedure to welcome new data providers
➢ duties required from resources' providers
➢ procedure for publishing new software
➢ procedure for accepting a product as "VAMDC brand"
➢ good practice for software production, e.g.
  o comments to be included in the codes
  o codes' documentation
  o repository policy and management
  o licences survey
➢ definition of Key Performance Indicators
  o for the monitoring of the resources
  o for the support and tutorials

The Task includes the assessment of the implementation of the quality procedures. To this effect each resource provider reports every 4 months about the compliance of its resource with respect to the procedures via the on-line tool. A 4 months status report is edited by the executive director in order to warn the resources' providers.

The output from this activity will be the set of the procedures and good practices, as well as an annual report about the quality status of the infrastructure.

5.6.1.5 *Training and Support Task*

The Training and Support Task handles the creation and maintenance of support and training materials, the organisation of tutorials. The developers are requested to document the software are documented, to provide tutorials and science use cases. The executive director is the contact point for recruiting tutors (a leader and participants for a specific event). The list of possible tutors is established at the start of each calendar year.

5.6.2 *"Support to Users" Tasks*

Those tasks include an education task, an industry task, an outreach task that will have initially one contact person, the scope of the tasks growing as activities grow. Eventually those tasks will become working groups as described in the sections below.

5.6.3 *Evaluation/Validation Task*

The Evaluation/Validation Task handles the integration into the "VAMDC infrastructure" of the evaluation/validation of data. The Task will propose the characterization of scientific evaluation for various sets of data. The work will be carried out via connection to the evaluation/validation networks that assess the data included in the VAMDC e-infrastructure. The task will collect their requirements and invite those networks to propose their evaluation to VAMDC. Those networks can be networks attached to a single database, or networks linked to datasets used for specific user communities. These include IAEA and IUPAC as well as standards and guidelines drawn up by other datacentres/databases such as HITRAN, KIDA, and other VAMDC databases.

5.6.4 *International Collaboration Tasks*

The International Collaboration tasks cover different scientific domains and are carried out by the scientists in their own domain. They provide the links with other datacentres, data producers and users. Those tasks handle the dissemination of the protocols and of the philosophy of the VAMDC to the international community. Those Tasks also horizon scan opportunities for VAMDC engagement in National and International projects and develop participatory projects with the VAMDC Board of Directors.

6 The Business model

The business model must be in phase with the working conditions of the different partners, i.e. must take into account that the employment situation is very different from one country to another with less and less recurrent support, and most income comes from short term projects. This means that most partners cannot promise manpower or grant support over 5 to 10 years. The partners can agree to work in some areas, taking responsibilities for some actions. In addition at this stage it is not possible to foresee the impact of « selling services ». Therefore the business model will rely for all partners mostly on their abilities to answer research calls at a national and international level.

All research activities should be sustained from projects and from local/national support to the databases/services/documents tagged as "VAMDC".

The VAMDC Consortium will delegate to its legal representative the representation of the VAMDC Consortium, protection of the "VAMDC" brand, certification of the "VAMDC brand" upon decision of the VAMDC Consortium board of director.

A small fee (less than 3kE/year) per year will be requested to "VAMDC Consortium" members in order to sustain the running costs of the VAMDC Consortium, this fee will be handled to the "VAMDC Consortium" legal representative.

The VAMDC Consortium legal representative negotiates with the Board of Directors a financial compensation in order to cover the administrative costs directly linked to its legal representation.

7 The Business Plan

7.1 Research Activities

The research activities will include maintenance and small upgrade of the infrastructure for the sake of research, it will include developing tools for our users and providing support to our users. Large upgrade and important development of software could only be carried out through projects.

In the next 5 years we intend to consolidate the infrastructure and its procedure in order to make it more robust and user friendly, as well as extending the number of its databases. Those actions will be discussed at the level of the Scientific and Technical Board and carried out by members who will be interested in sustaining such activities. If these actions impact strongly the members, an authorisation from the Board of Director will be necessary. We might plan to extend the scope of the infrastructure to fields such as nuclear physics as it is easier to do than going towards more complex systems. Of course any of those options would require personal commitment of one or more members and collaboration with new partners through common projects.

Over the next 10 years the technology will certainly evolve and the current set up might well become obsolete. Therefore there might be an effort to lift the technology above the current set up. But this could only be achieved through new projects and new partners from the ICT community.

Any upgrade impacting the nodes will be discussed in the Board of Directors.

7.2 Education

The education activities do not exist at present at the collective level. Paris Observatory is interested in leading this activity and other interested nodes would be participating if they see any interest. This activity can be initiated via prototyping and offering services, but any large initiative would be initiated through answers to calls at National or International level. During the calls we intend to hire senior level people so that they would be able to lead the activities

by themselves, thus lowering the impact on the current staff. In the next 10 years we can expect to have integrated naturally VAMDC in the teaching curriculum of some different teaching units. The place of meeting and discussion will be the "Education Working Group", and any project wanting to use the "VAMDC brand" and to be visible from our website will request the prior acceptation from the S&T Board. The possibility of collaboration and partnership will be advertised on our website.

7.3 Industry

VAMDC at present is largely academic in its orientation and in particular there is little industry engagement. In the longer term VAMDC should be providing facilities for industry and business (including SMEs) as well as research laboratories (both public and private).

UCL is willing to lead this activity as the group has a start-up company and is involved in a number of other industrial collaborations involving atomic and molecular (A+M) data. A first step for the industrial strategy is to focus on databases that are of technological interest to industries. Therefore in the next 5 years the VAMDC consortium will seek to adapt VAMDC services to data provider for industries. There are probably two main industrial sectors interested in A+M data.

1. Those involved with atmospheric processes: for ambient temperatures HITRAN covers most of the needs here. However there are specialist area such as (a) space craft re-entry plasmas and (b) hot gases in smoke stack emissions and elsewhere in which there is a need for further data provision.

2. The area of technological plasma and chemical vapour deposition are major drivers of global economic development. VAMDC is not well served for databases in these areas. One key partner is the plasma database LXCAT run from the Laplace Laboratory in Toulouse. Other databases need to be identified and the extension to A+M processes on surfaces is particularly important.

Paris Observatory and Cambridge would be involved at the level of software developments in this area. Other members might join the service activities. These industry-related activities will be carefully discussed at the Board of Directors so that the "VAMDC Consortium" could benefit from the developments and possible income that derives from them. Indeed such benefits would contribute to the sustainability of the "VAMDC Consortium". The possibilities for partnership and collaboration will be largely advertised on our website. The place of meeting and discussion will be the "Industry Working Group".

7.4 Outreach

Outreach activities are activities that will be carried out at their own pace depending on the public in contact with the partners, and depending on the partners' interest. From the consortium point of view the outreach will occur through the website, through information provided to the public on social networks, through partners normal outreach activities. No specific business plan needs to be drawn at that point since these activities can be carried out by individual members. Another aspect of outreach is the fashionable activity of "citizen projects" that lifts the awareness of science with the public. That latter activity requires a scientific and development team interested to develop the concepts and products, then to follow the project. This requires a genuine interest from the partners as these activities require

some significant professional time. The interested institutes could handle those "citizen projects" through students' projects. Our main website would be the entry point for such activities. We would invite any individual to propose ideas and assess how much support some partners are willing to bring to such ideas. The place of meeting and discussion will be the "Outreach Working Group", and any project wanting to use the "VAMDC brand" and to be visible from our website will request the prior acceptation from the S.&T. Board.

7.5 Marketing/Communication/Management

The Consortium must build a strong marketing/communication cell combining efforts from different partners, supported over the next 5 years via projects, and via a small fee, and managed by one person. The management/marketing/communication aspects of the "VAMDC Consortium" should be combined with management of the different projects so that a person could always be hired over the next 10 years.

8 The Activities

8.1 Research

Research Activities deal with activities linked to producers of atomic and molecular data and activities linked to users of atomic and molecular data.

8.1.1 Producers of Data

How to include data in VAMDC e-infrastructure ?

The current VAMDC e-infrastructure includes databases related to atomic and molecular spectroscopy and to heavy particle collisional processes, and is appropriate to the type of currently accessible data. Any producer of data can join the VAMDC infrastructure through different means:

➢ they may include their data in existing atomic and molecular databases that are partners of VAMDC
➢ they may create a new database hosted by a partner of VAMDC
➢ they may create a new node in VAMDC e-infrastructure

What is the impact of new types of data on the VAMDC e-infrastructure ?

The current underlying technology allows direct extension of its capabilities to different processes involving conservation of the nuclei, but as well to processes involving nuclei transformation.

The extension of technology to more complex systems in chemistry or biology, and the inclusion of environment data would require to be handled

➢ via a strong upgrade of the existing data models
➢ via the creation of new data models or the adoption of externally defined data models, including the creation of tools able to handle different data models

Any severe upgrade could only be carried out by external groups to which some VAMDC Consortium members might interested to bring support within the framework of common projects.

What would be the motivation for inclusion of new types of data ?

Two main motivations could be considered:

➢ a new community of data provider is interested to beneficiate from our experience and from part of our software,

➢ one of our user community needs different types of data to be combined with the set of data already available in the VAMDC e-infrastructure,

The inclusion of new types of data would certainly impact some of the "VAMDC Consortium" members, therefore the way of integration within the "VAMDC e-infrastructure" would be discussed within the Board of Directors, and the "VAMDC Consortium" members supporting such changes should make a case showing that this community is strategic for reasons such as increase of visibility, new customers, new stakeholders leading to consolidation of sustainability

Added value for data producers

Furthermore VAMDC aims to provide A&M data providers and compilers with a large dissemination platform for their work. In the near future all products related to VAMDC, portal and tools, will explicitly warn the downloading users that in their publications, besides the databases also the references to the original papers, where the data have been published and discussed, must be cited. The VAMDC cannot oblige the users to this duty, but at least a reminder of this ethical commitment, before to give access to the data during the downloading process, must be given in the web interface.

8.1.2   User Communities
Rationale

User Communities are linked to the type of databases that are included in the VAMDC e-infrastructure since the purpose of building databases is very often linked to their intended usage. Again strategy to work with one community or another will be linked to the type of support and/or funding that the user community could bring, to the visibility and impact that the VAMDC Consortium would gain, to the scientific interest of VAMDC Consortium partners. We intend that our future development with user communities be driven by science.

Up till now we have developed tools and services using scientific use cases from the Astrophysical Community and our objective is too increase our impact in this area as 80% of our current databases are useful to that User Community.

Our current databases serve other communities such as atmospheric physics, fusion, plasma physics, radiation damage community.

What can be done for user communities?
➢ we can improve our current services and tools in order to meet the users requirements.
➢ we can port the VAMDC capabilities and facilities into tools developed by institutes outside the consortium
➢ we can provide the scientific community with innovative tools for easily handling and processing results
➢ we can provide "derived products"

we can provide support so that external users can implement our VAMDC plugin in their software

Nevertheless one of the important lessons we learned while working on the VAMDC infrastructure is that the use of A&M data directly requires an expert competence in understanding the data and being able to incorporate it in various applications. These two requirements (expertise and access to tools capable of using A&M data directly) severely

limit the number of potential clients. While we can address the technical issues related with data extraction from XSAMS form according to user specifications by means of consumer tools, we do not always have the capacity and competence for developing field specific application tools. In a sense VAMDC makes the situation worse offering users a choice of data from different sources that they may not have before. Thus we are facing two issues: simplifying the use of A&M data for wider community and developing methodology and infrastructure for data evaluation. If we can address these problems in the near future we will supersede the sum of participating databases making VAMDC the main source of A.&M. data for wide range of research and application areas.

Below are sketched possible ways of addressing the two problems starting with Derived Products followed by Data Quality Assessment.

Derived Products in VAMDC

One way of reaching out to wider community without developing problem-specific applications and tools is to meet out clients half-way by converting the raw A&M data to intermediate data products often used in various tools connected to remote sensing, simulations, design etc. Examples of such products (that I will refer to as Derived Products or DPs) are opacity tables, transmission tables, emission spectra, absorption spectra etc. A DP will incorporate our knowledge of A&M data including such complex environment-related physical process as spectral line shapes and shifts into something that is much more homogeneous and directly useful for diagnostics or simulations. This concept is not mine and not new. A few participating DBs have such option: Chianti can compute line strength and simulate spectra of hot low-density plasmas, HITRAN can compute line opacity and line strength in Earth atmosphere while VALD can do similar calculations for stellar atmospheres. The need for such DPs can be illustrated by the direct access statistics to individual nodes (not through VAMDC). About 55% of all requests to VALD request line intensities in addition to the raw data. For CHIANTI this number is close to 100%. The new tool HITRANonline is gaining popularity.

It is obvious that we cannot produce one-two DPs covering all applications. We will have much better success addressing various physical conditions and, possibly, wavelength domains separately. Such approach will map well onto the expertise of different nodes taking advantage of the available building blocks. Porting tools to the VAMDC or developing new tools can be scheduled according to demand and available resources.

We have all pre-requisites to start this work already now covering atomic and molecular radiative transitions in gas/plasma phase. The latest XSAMS standard supports the description of the environment and full description of species including partition functions, ionization etc. We can also re-use the equation of state routines and the line profile recipes appropriate for different regimes. Thus we should be able to produce the absorption and transmission DPs under the assumption of LTE for a single or multi-layer environment. Combined with a radiative transfer solver this DP will be used to generate synthetic spectra and estimating emission/absorption by individual transitions. Such a relatively simple tool will go a long way in terms of data visualization. It can also be used in number of applications that do not require advanced treatment of radiative transfer and self-consistent level populations. NLTE extensions can be considered in the future development.

A built-in capability of generating synthetic spectra that can be compared with measurement will massively simplify integration of VAMDC into education on all levels. It could be a key to engage public at large in the environmental and climate issues etc.

Data Quality Control

It is obvious that the VAMDC consortium does not have enough experts and manpower in general to tackle critical evaluation and ranking of the data. This problem has many layers from cross-identification of values that refer to the same A&M parameter and then trying to rank multiple realizations. Such data may come from different sources, experimental or modeling techniques that may quote precision but no systematic errors. The VAMDC and XSAMS made a great effort in facilitating the cross-identification by forcing individual nodes to verify and homogenize the source referencing system and the inclusion of energy level identifications. The next step: the actual cross-identification and ranking will require time and close collaboration with expert centers such as NIST, various laboratory spectroscopy labs and centers for computational modeling of atoms and molecules. The exact incentives and format of such collaboration should be established in the future but the main goals should be to establish a filter for low-quality data, encourage data producers to go through such evaluation and to publish data in VAMDC nodes. We will eventually insert in the Internal Regulations a statement saying that, as a very first step, no data can be accepted by the VAMDC if not accompanied by one or more articles, published in some accredited journal, and so submitted to a refereeing process. This will ensure that both the data and the method of their production, theoretical or experimental, have been criticized by some expert referee. The VAMDC has already made compulsory the inclusion of references in the XSAMS schema.

The emphasize of the VAMDC activities should be on organizational side to relieved the data producers from logistical, proprietary and administrative aspects of the process.

8.2   Education

Education activities cover different target population and different methodologies. The target population is secondary school education, higher education and continuous education. The methodologies include the use of VAMDC in face-to-face education sessions or via on-line teaching such as the very popular MOOC.

The Education activities are linked to the national Curriculae and must be displayed in the national language at least for all level below university degrees. Nevertheless the coupling of education activities in science and the use of english is often seen as attractive.

The impact of atomic and molecular Data in Curriculae

The atomic and molecular data available through our infrastructure describe the microscopic world of atoms, molecules, particles and their interaction. At the level of Education Curriculae this world starts to be investigated in the range 15 to 18 years old and above. More relevant to any level of education are the learning area linked to our user communities. Indeed our data are used by researchers from other domains; for example:

- ➢ in astrophysics in order to understand the chemical composition of a galaxy, of an interstellar cloud, of a planet's or star's atmosphere
- ➢ in atmospheric physics in order to find the concentration of pollutants of our atmosphere and even for measuring the speed of wind

- ➤ in plasma physics
- ➤ in biophysics in order to simulate the path of electrons in living tissues for radiotherapy applications

Our Objectives for Education

Our objectives for Education are the following:
- ➤ to give easy access to atomic and molecular data and information related to these data
- ➤ to provide innovative pedagogical resources in agreement with the national curriculae in order to illustrate lectures at all level of education
- ➤ to re-inforce the link between research and education
- ➤ to create national networks, and to interconnect them at the international level
- ➤ to be partners of public institutions
- ➤ to support teachers and lecturers, and bring them our knowledge on our scientific expertise linked to e-science.
- ➤ to offer training on the developed education tools
- ➤ to train and recruit young scientists with a view to securing qualified staff for our own needs

Our Strategy for Education

Our strategy for education is based :
- ➤ on working with individual teachers in order to define the appropriate pedagogical resources for a given level of education. This includes the development of software performing some visualisation of data, the simulation of the macroscopic world, the creation of local selective databases that would full fill some pedagogical goals
- ➤ on communication and marketing actions that would
  - o highlights the examples of available education resources
  - o propose training for those resources
  - o offer collaborations and describe the modes of financing
- ➤ on offering continuous education resources and work with the continuous education departments

Our Financing and Leadership Strategy for Education

- ➤ the actions will be carried out by the "VAMDC Consortium" members who have some professional interest in those activities. Each member will have the intellectual property of their own developed products (software, database, educational documents) and be responsible for their maintenance and upgrade.
- ➤ the financing of small scale actions will be carried out within the scope of the teaching duties of the "VAMDC Consortium" members through developing pedagogical resources as part as their share of teaching hours in their teaching departments or as part of their student supervising hours.
- ➤ the financing of large scale actions will require stakeholders support at national or international level
- ➤ the external partners from the educational world would become "associated members" or "honorary" members of the "VAMDC Consortium" and therefore will be invited to participate to the "Education Working Group", as defined in the Status of the "VAMDC Consortium"
- ➤ the coordination of "VAMDC Consortium" education activities will be done at the level of the "Education Working Group" and the chair of this group being part of the

"VAMDC Consortium" Technical Board, thus reaching the highest level of decision within the "VAMDC Consortium".
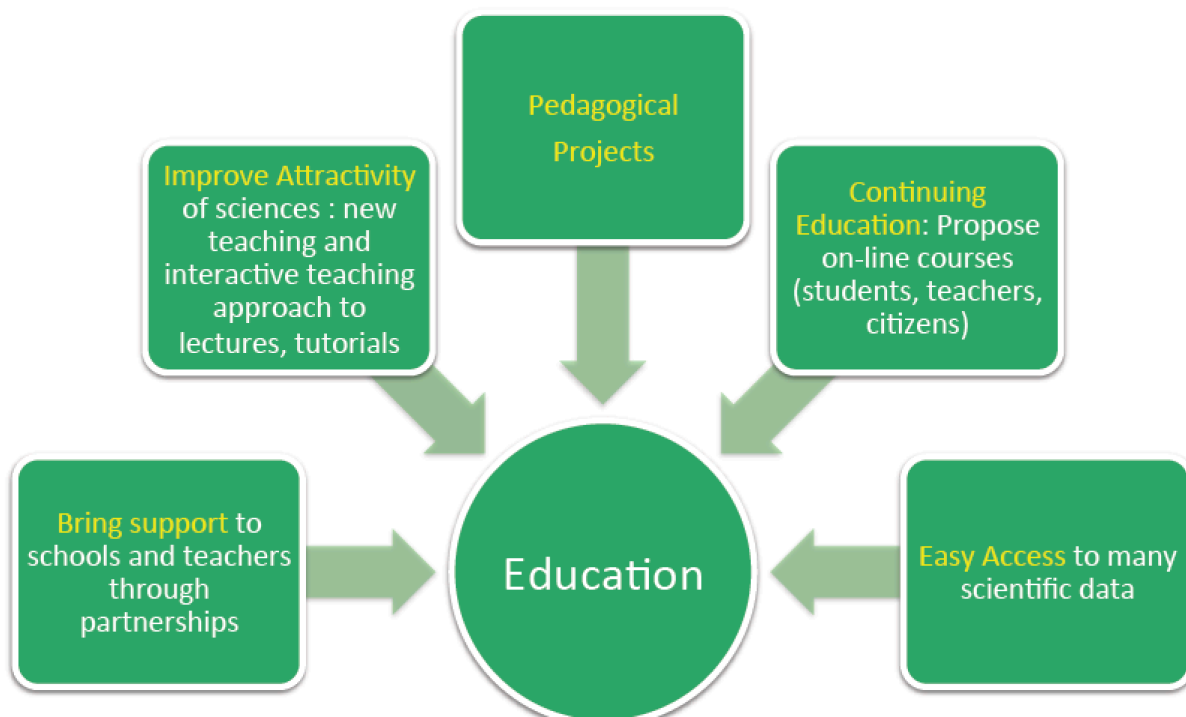
8.3 Industry

Industry covers industry and commerce, including SMEs.

The activities linked to industry can be foreseen
- as enriching our database of industries interested in atomic and molecular data
- as providing access to specific data
- as making the link between the businesses and the data provider
- as developing tools on demand to handle and transform the data for the businesses applications
- as having agreement for technology transfer and offering supports to the businesses
- as having partnerships with businesses involved in education and outreach applications
- provision of a one stop-shop location for all data needs for a sector without complete access to the academic literature or the necessary skills to use it
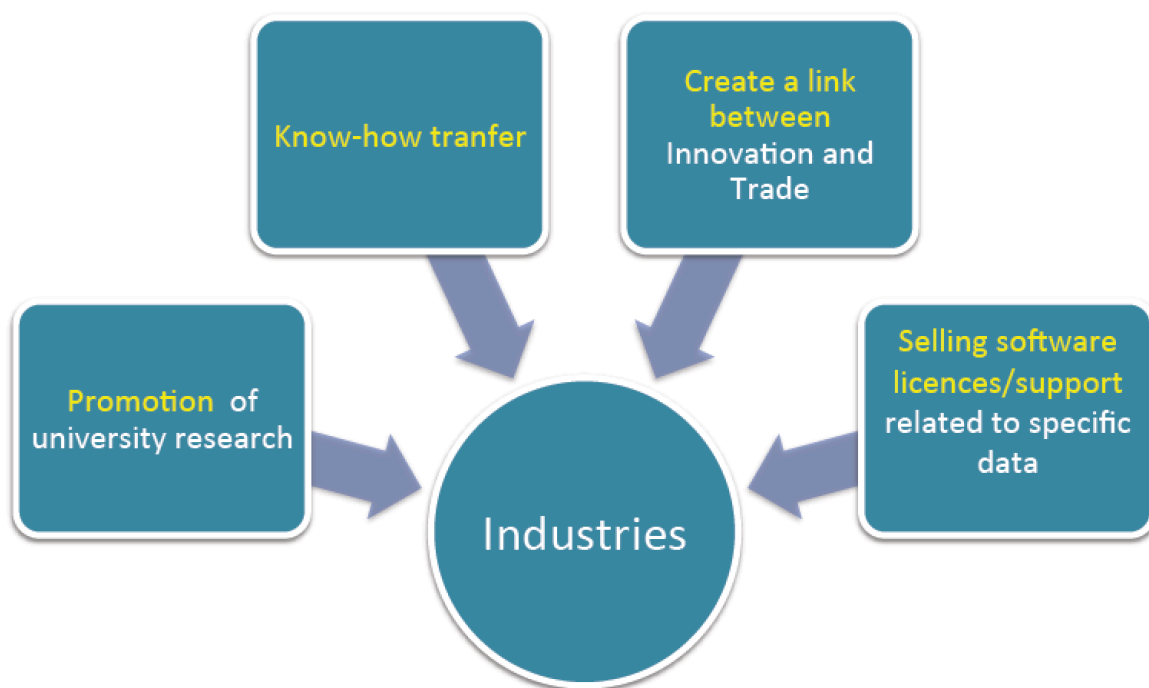
8.4 Outreach

The outreach activities are foreseen to occur through
- information provided to the public on social networks
- through participation of partners to citizen science projects
- through partners normal outreach activities such as participating to public conferences, to institutes exhibitions.

All activities will be recorded on our website and call for ideas will be advertised.

8.5 Communication/Marketing activities

The "VAMDC Consortium" communication/marketing activities will occur

- ➢ through its main website that is the entry point for all customers from research, education, business, outreach (http://www.vamdc.eu)
    - o through a virtual tour of the "VAMDC Consortium"
    - o through the availability of leaflets to downloaded
    - o through a News section for the 4 activities
    - o through a Events section for the 4 activities
    - o through forums for the 4 activities
    - o through hosting blogs in the 4 activities
- ➢ through using social networks such as Facebook and twitter
    - o Facebook is largely used by the public and would be interesting for education and general outreach. Currently the facebook account gives linked to our news and to our website
    - o Twitter has not been used recently. The twitter account will be used to mention interesting features to the VAMDC followers when we attend conferences, when we have news related to A.&M. data and their applications (either native news or re-use some other twitts).
- ➢ through using professional networks such as Linkedln, Research Gate:
    - o ResearchGate is aimed at researchers from users and producers communities. It allows to attach papers. Nevertheless it has the drawback that its concept is difficult to use for an account called "VAMDC" since "VAMDC" is not an author and an account can hold only 2 owner' s names. We have contacted ResearchGate in order to seek a solution. ResearchGate is dynamical and is a good channel to show the community the involvement of different members of VAMDC
    - o LinkedIn is a professional network that is aimed for Industry even if most of our colleagues have such account.
- ➢ through open mailing list for producers, users, and the 3 other activities that could be used by the registered persons to exchange ideas, but as well by the partners to give information
- ➢ through the edition of a Newsletter every 4 months. 4 months is an initial reasonable period that will be re-assessed after a year of existence. The Newsletter will cover all activities and will be distributed/advertised through the other existing channels of communication
- ➢ through the natural channel of dissemination in conferences and workshops. We plan one annual consortium meeting a year that will be coupled to an international conference related to A.&M. data. All partners will perform dissemination whenever they sit fit to do so, and collectively the board will ensure a presence of VAMDC in major international conferences on a regular basis
- ➢ through organising tutorials for different categories of users either through self-organisation or through joining other tutorials linked to A.&M. data or to e-infrastructure or to the application fields. The board will ensure that tutorials are offered every year.

The above communication channels and tools will be used for external customers and internal VAMDC Consortium members.

The reporting of bugs will be encouraged through the RT-support system and through the forums.

For internal communication, the vamdc.developer@sympa.obspm.fr mailing list will be used in order to discuss the technical aspects, the board.consortium mailing list will be used for political issues. The use of forums for technical aspects will only be used by the VAMDC Consortium members in order to put forward some main large issues and to expose their final results so that the public is aware of such issues. The internal communication goes through the usage of the management system, i.e. REDMINE for now where all actions/projects/bodies/products/ are reported and followed through issues, upload of documents and résumé.

9 Appendices

9.1 What is the VAMDC Infrastructure ?

9.1.1 Introduction
VAMDC has built an internet infrastructure giving researchers coordinated access to multiple databases of atomic and molecular data. This infrastructure consists in software developed by the VAMDC project and deployed as web applications and web services at member institutions. "web service" means a collection of web resources arranged for access from a software application, while "web application" means a collection of dynamic web-pages arranged to form a user interface when accessed by a user's web browser. All resources in VAMDC web services and web applications are accessible by HTTP.

9.1.2 Data nodes
VAMDC provides one web service for each database of science data. The combination of the web service and the database is called in the jargon a "VAMDC node", and the software operating the service, written by VAMDC, is called "node software".

VAMDC nodes are distributed across the member sites; centralization of the data has been explicitly avoided. Typically, a node is deployed where there are researchers curating the node's database, with a mirror of the service and database at another site to increase the service availability. Nodes and their mirrors can occasionally be moved between sites.

VAMDC nodes provide a query protocol. A client sends to a node a query to select an extract from that node's database; the node computes the extract and returns it in XSAMS[1] format. The query language is VAMDC SQL subset 2[2] (VSS2) and the query protocol is called VAMDC-TAP.[3]

Because the extracts are computed on demand, all nodes store their data in a relational database rather than in flat files. Constructing the relational form of the database is a major part of deploying a VAMDC node.

All nodes follow the same query protocol, and all present the same, relational data-model established in the VAMDC dictionary. This allows an application to send the same query to

---

1   See http://www.vamdc.org/documents/standards/dataModel/vamdcxsams/index.html
2   See http://www.vamdc.org/documents/standards/queryLanguage/index.html
3   See http://www.vamdc.org/documents/standards/dataAccessProtocol/index.html

all nodes in VAMDC and to aggregate the results. However, the kinds of data available vary from node to node. A node containing only atomic data, for example, will reject a query selecting on molecular quantum numbers. The protocol allows the client to distinguish these cases from outright failure of a node.

The VAMDC-TAP protocol is designed for direct use by applications. If the VAMDC infrastructure were reduced to just the set of data nodes, then it would be still be usable; the rest of the infrastructure helps but is not essential. This approach was chosen at the start of the project to increase the chance that VAMDC remained useful after the end of funding.

9.1.3 Registry of services

The VAMDC registry is a database of metadata describing VAMDC nodes and web applications. It allows an application to find the address of a given node; to select nodes by the kinds of data they offer; to find out which query terms are supported at a node. In short, the registry helps applications form correct queries.

A VAMDC system has exactly one registry, set up as a mirrored pair of services (for higher availability) with published, unchanging addresses. If an application knows the address of the registry, it can find all parts of the system even if those parts are migrated between sites.

The VAMDC registry follows the IVOA[4] registry protocol and is built from the AstroGrid[5] registry software. Some aspects of the format of registrations are VAMDC extensions to the IVOA standards.

9.1.4 XSAMS-processing applications

VAMDC data-nodes produce data extracts in XSAMS format, which is a good data-transfer format but needs to be translated into the native format of applications that consume the data. There is also a need to display data in XSAMS for human assessment.

VAMDC provides a suite of services that transform XSAMS into other forms. These installations are web services conforming to the VAMDC protocol for XSAMS consumers[6], and some include web applications for interactive use.

The XSAMS-processing services may be discovered in the VAMDC registry.

The commonly used XSAMS-processing applications were written by VAMDC, but other groups can write and register their own XSAMS processors.

9.1.5 Web portal

As noted above, the VAMDC nodes can be used directly from application software. For exploratory use, in connecting applications to VAMDC and as an introduction to new users, VAMDC provides one, generic application for querying the nodes and displaying the results. This is a web application, called the VAMDC web portal.

The portal is a new software written by VAMDC (http://portal.vamdc.eu).

The portal web-application handles the query process but does not handle the data resulting from the query. Instead, the portal allows users to download data directly from the VAMDC nodes, or to pass those data from the nodes to the registered, XSAMS-processing applications. The latter applications provide a way for user-communities to extend the VAMDC portal with special features.

---

4   See http://www.ivoa.net/Documents/
5   See http://www.astrogrid.org
6   See http://www.vamdc.org/documents/standards/dataConsumerProtocol/index.html

9.1.6 Database of species

Many queries select by atomic and molecular species, and the names for species vary between scientific communities. VAMDC maintains a database of names from which community-specific terms may be refined into standard identifiers understood by all VAMDC nodes. VAMDC identifiers are based on InChI codes. In most cases, the VAMDC identifier is just the standard InChI, but in special cases a suffix is added to distinguish molecular conformers. One copy of the species database is encapsulated in the web portal. Another copy is available for query by applications as a VAMDC data-node. The database is not distributed to software authors outside VAMDC.

9.1.7 Versioning the VAMDC system

VAMDC made and will continue to make coordinated[7], periodic releases of the standards noted in the architecture description above. VAMDC deployments using different versions of the standards are not compatible.[8] Therefore, VAMDC deployments are grouped into separate systems, with one system per release of the standards.

The version of a VAMDC system is the version of the ruling standards, expressed as the year and month (e.g. 12.07 standards were released in July 2012) of the release of those standards. The release date of the system lags the release date of the standards by at least one month.

VAMDC aims to keep two system versions available concurrently: one current release and the release most recently superseded. Software written outside VAMDC for a given release of VAMDC will continue after that release is superseded for at least one more release-cycle, giving the authors of the software more time to update their code to current standards.

A system release contains a set of data nodes (the URLs of the nodes change across releases), a registry, a compatible version of the web portal and a set of XSAMS-consuming services adapted to the correct version of XSAMS. Any software application using a registry from a particular release is guaranteed to see only compatible services from that release; to upgrade to the next release, the software authors must use a different registry.

9.1.8 Standards

One activity of VAMDC 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

9.1.9 Software

---

7   This is in marked contrast to other movements such as IVOA where participating installations upgrade their services piecemeal.
8   The greatest incompatibility comes from changes in the XML schema defining the XSAMS format.

All supported Software and their documentation can be found on
http://www.vamdc.eu/software

The software can be divided in software usable by the general research users such as SPECTCOL, SPECVIEW, XSAMS-CONVERTER, software for more technologically advanced users such as PDL-VAMDC, and by developers wanting to implement databases or test the output of databases or to implement access to VAMDC databases in client applications: NODE SOFTWARE in Python and Java, TAPvalidator, JAVASCRIPTPORTAL source code, the libraries

SPECTCOL

Spectcol is a tool aimed at the manipulation of data in VAMDC-XSAMS format for the purpose of extracting and merging spectroscopic data and rate coefficients data from different sources. The outputs are directly used in user software in order to solve radiative transfer problems. This tools is the example of a user community tool produced by VAMDC. The development of such tools should be pursued.

SPECVIEW/VAMDC

Specview is a tool for 1-D spectral visualization and analysis of astronomical spectrograms from STScI. It is written in Java thus can be run anywhere Java is supported. Specview is capable of reading all the Hubble Space Telescope spectral data formats, as well as data from several other instruments (such as IUE, FUSE, ISO, FORS and SDSS), preview spectra from MAST, and data from generic FITS and ASCII tables. It can also read data from Virtual Observatory servers, and read and write spectrogram data in Virtual Observatory SED format.

Specview can overplot spectral line identifications taken from a variety of line lists, including user-supplied lists. Its linelists' query form has been modified to include the VAMDC Query Module, called QueryBuilder, thus providing the full capability of querying the VAMDC databases. In particular it allows to select finely the observed species and properties of linelists. Currently about 20 spectroscopic databases are inter-connected through VAMDC and accessible through VAMDC software and libraries.

XSAMS CONVERTER

The XSAMS Converter tool is a Java application aiming at converting XSAMS file into other file formats. This application embeds the existing XSAMS processors into a standalone application that can be used without a network connection. The application provides both a graphical user interface and a common line interface so that it can be called from a script for example. The user simply chooses one or more XSAMS file to convert and the processor to be applied.

PDL-VAMDC

PDL is a standard of the Inernational Virtual Observatory Alliance for describing deploying and running scientific services and workflows
[http://www.ivoa.net/documents/PDL/20140523/index.html].

By using the PDL framework we exposed VAMDC as a PDL web service, accepting requests on radiative process. As for all the PDL services, the users can interact with the VAMDC-PDL one by using a standalone Java-Swing client or by using the Taverna Workflow Engine.

Node Software

VAMDC provides software for creating a node and publishing data therein. Two implementations are available: the node Software in Python and the node Software in Java. The version numbers of the software reflect the version of the VAMDC standards. Hence, v12.07r1 supports the 2012.07 standards and is the first revision of the software in that edition of the system. The documentation for creating a node is on the same pages as the software. The libraries published in the node software in Java can be re-used in client applications.

TAPValidator

VAMDC provides software to test the validity of the nodes outputs. VAMDC-TAP Validator (TAPValidator) is an universal tool aimed for development and testing of VAMDC data nodes. It allows to ensure correct operation of a node on low-level, by directly sending data queries and analyzing XSAMS XML response. Written entirely in Java, it is cross-platform over Linux, Mac and Windows, and requires no installation.

JAVASCRIPTPORTAL

This is a proof-of-concept application that implements access to the VAMDC infrastructure in JavaScript, thereby allowing to run an application in the users' webbrowser. The example uses the popular Query library to let the user select an element and a wavelength range. It then formulates the query to a data node, asks it for the amout of data available and presents this to the user. In a second step, the user can hand the data-URL to a VAMDC Processor service and plot the result that it gets back. Alternatively, one can also chose to download the data from the node into the browser and do a simple tabular display of it.

Since this application consists of only a HTML and a JavaScript file, it can easily be modified for different purposes and hosted on any webserver that serves static files. The source code is available at https://github.com/VAMDC/jQueryPortal

LIBRARIES

VAMDC proposes a set of libraries that can be used by producers and users to implement and access VAMDC databases:

- **QueryBuilder**

QueryBuilder is a Java Swing library which is created to reproduce the same logic of query interface and query building methode into vamdc portal. It is implemented to be integrated into other Java applications.

This tool is mainly intended for VAMDC developers who want to improve their Java applications by increasing the number of query possible for VAMDC databases.

- **VAMDclib**

This library contains several modules that implement access to VAMDC's infrastructure. It allows to query the central registry service to obtain information about registered VAMDC database nodes and to send queries to these nodes as well as to process the retrieved data. The data is made available as python dictionaries. Tools to store and manage the data in a local

sqlite3 database or based – xml database are included. For more details, please visit this link: http://vamdclib.readthedocs.org/en/latest/

- **XLS Processor**

This is a XSAMS Processor service according to the VAMDC Standards that can apply any XSLT Transformation to the XSAMS format. For many use-cases, this is a convenient way of shaping the generic XML format into something that is useful for a particular application.
This generic implementation is built upon the Python/Django software stack and allows to deploy a Processor service on multiple XSL stylesheets at the same time, with new ones being easily added. The source code can be found at https://github.com/VAMDC/Processors

- **Registry Query Library**

This is a simple PHP library to look for services in the VAMDC registry. It can be used in command line scripts or in server side script. The archive includes a sample script showing how to use it and a HTML documentation.

DOCUMENTATION

The documentation is linked to all software and standards developed by VAMDC. SUP@VAMDC has tremendously improved the documentation linked to the portal, to the software and libraries developed in VAMDC or in SUP@VAMDC. This documentation includes tutorials, user guides, technical documentation, videos, leaflets for the science use cases. Obviously this documentation will evolve over the years.
As an example is given below the snapshot of the description of the SPECTCOL software (http://www.vamdc.org/activities/research/software/spectcol/)
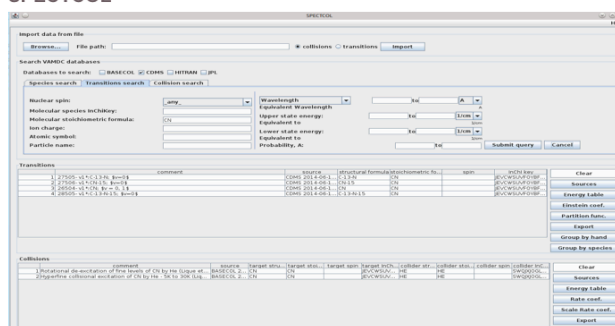
SPECTCOL

SPECTCOL is a graphical tool implemented in Java. It allows to manipulate and combine spectroscopic and collisional data coming from the databases (BASECOL, CDMS, HITRAN, JPL,…) using VAMDC technology.

XSAMS(Xml Schema for Atoms, Molecules and Solids) is the VAMDC data format and SPECTCOL is able to manipulate and provides methods to convert these data into other formats(CSV, RADEX, LTE …).

For any question or feedback use the forum link here.

Credits:

M.L. Dubernet, Y.A. Ba, L. Nenadovic, M. Doronin

BSD Licence, @VAMDC Copyright

Scientific Use Case   ⊕

Tutorial   ⊕

Version History   ⊕

Files to download

Spectcol Jar File v12.07-r2
Download

SPECTCOL user guide documentation v12.07-r1
Download

Technical information of SPECTCOL tool
Download

9.2 Partners in the FP7-VAMDC Project, FP7-SUP@VAMDC Project and "VAMDC Consortium"

The list of partners for the 2 European Projects and for the "VAMDC Consortium" can be found in the Memorandum of Understanding of the "VAMDC Consortium". The document lists the background to both EU projects.

9.3 Technical By Law (see joint PDF Document)

This document is part of the Internal Regulations of the VAMDC consortium. It defines the technical duties that each member providing physical resources has to respect for assuring the quality and reliability of the VAMDC infrastructure. This document will be submitted to the Board of Directors for internal approval.

9.4 AAA Policy (see joint PDF Document)

Contains the *Authentication, Authorisation and Accounting* strategy defined by the consortium members in a collaborative way. The proposed strategy is the result of a synthesis

between the *state-of-art* currently adopted solutions and the technical constraints related to the VAMDC infrastructure design & technology. This document is consigned to the Board of directors of the VAMDC consortium who is in charge of the implementation of the global strategy described into the Roadmap.

9.5   Background to VAMDC Consortium (see attached PDF)

Contains the list of all foreground of VAMDC and SUP@VAMDC projects that are the background to the VAMDC Consortium bound by the MoU of 1rst November 2014.

VAMDC

Virtual Atomic and Molecular Data Centre

Consortium Document
SUP@VAMDC Project

Internal technical Rules

Version 1.0

Document Information

Editor:	Carlo Maria Zwölf
Author(s):	Carlo Maria Zwölf
Contributor(s):	

Type of document:	Internal Regulations
Status:	draft
Distribution:	Public
Work package:	WP2
Version:	1.0

Directory and file name: To Be Defined

Abstract : This document describes the Technical Duties that "VAMDC Consortium" data/services providers are encouraged to respect.

Version	Date	Reason for modification	Modified by	Distribution
V 0.1	20/06/2013	Document initialization	Carlo Maria Zwölf	VAMDC extended Board
V 0.3	06/03/2014	Adding part for members providing services without data	CMZ	VAMDC extended board and sup@vamdc ept
V1.0	21/07/2014	Final Check	CMZ	VAMDC extended board and sup@vamdc ept

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.

TABLE OF CONTENT

1. INTRODUCTION

This document presents the Technical Rules that should be respected by every member of the VAMDC consortium. It explains the conditions one has to satisfy in order to share data through the VAMDC infrastructure and/or to provide services based on the VAMDC infrastructure or standards. Those Technical Rules apply to all "VAMDC" tagged products as defined in the Memorandum of Understanding (Section 5).

This document is only focused on duties and is part of the "Internal Regulations Documents". The related rights are explained in the documents ruling the consortium, i.e. the Memorandum of Understanding and other parts of the "Internal Regulations Documents".

2. DUTIES FOR DATA AND/OR SERVICE PROVIDER

The term "service" stands here for processing imbedded into web services and/or standalone software, based on VAMDC infrastructure and/or on VAMDC standards.

Each entity, structure, organization (hereafter designed as **Data/Service provider (DSP)**) wishing
•   to share its atomic and molecular data using the VAMDC infrastructure, or
•   to provide processing based on the VAMDC infrastructure or standards
must satisfy the requirements detailed hereafter and follow the quality chart described hereafter.

2.1 Data and or service publishing conditions

For joining the VAMDC infrastructure the DSP can:
1. Use the VAMDC software for publishing his/her data and/or providing services.
2. Use his/her own dedicated software solution, provided it respects all the functional and technical VAMDC standards and requirements.

In the first case, the DSP will accept the *conditions of use* and licenses coming with the adopted VAMDC software.

In both case, the DSP publishes only public data or data of which he/she holds the intellectual property and/or the diffusion rights. In any case, VAMDC cannot be held responsible for the publications through its infrastructure of private/protected data or for providing services based on protected algorithms: the legal responsibility is on the administrative organism of the DSP.

2.2 Manager designation

The DSP will designate
- a scientific manager and
- a technical manager

(A same person may have the two roles).
The scientific manager will be in charge of data quality.
The technical manager will be in charge of the computer-science aspects and will be the point of contact for the service quality.
These two managers will be the points of contact for the Technical Quality Working Group and its chair.

2.3 Documentation required for data integration and/or providing services

In the VAMDC consortium the Technical Quality Working group defines a set of documents/procedures that rule the quality of the VAMDC infrastructure. The Board of Director will validate the quality procedure.
The DSP must comply with the consortium quality procedure and will always maintain the required documentations up to date. The Technical Quality Working Group will regularly check the adequacy between what is documented and the operational reality.

2.4 Support to new DSP

The first level of support to new data-providers is based on *sponsorship*: the documents (see section 2.3) provided by a new data-provider will allow the identification of a VAMDC Consortium partner that has similar scientific and technical issues. This senior DSP may then assist the new one in deploying the VAMDC node.
Indeed every DSP, within its own limits and capabilities, is encouraged to help newcomers working in collaboration with the Technical Operational Working Group and the Training and Support Working Group.

2.5 Evolutions of the VAMDC software component

Small software and/or standards upgrades impacting the infrastructure, the existing services and/or the data-nodes must be quickly adopted by the data-provider. Their managers will coordinate the deployment of the new versions and will take care of the evolution of all the documentation produced, according to necessary changes. The Technical Quality Working Group will supervise every evolution. Large software and/or standards upgrades will only

occur through a vote of the Board of Directors that will manage how those changes could be handled at the different nodes, taking into account the local level of resources.

2.6 Technical Defaulting Party

This section recalls the conditions for the exclusion from the VAMDC infrastructure of a DSP.

A data provider might be excluded if he/she:

- Refuses to accept the conditions for data sharing (see § 2.1)
- Refuses to follow the quality procedure.
- Refuses to adopt the software and/or standard updates or refuse to update the documentation after software/technical updates.

Each defaulting party will be notified by the Technical Quality Working Group to the Board of Directors that will discuss the issue with the defaulting partner.

VAMDC

Virtual Atomic and Molecular Data Centre

Consortium Document

SUP@VAMDC Project

Authentication, Authorisation and Accounting strategy

Version v 1.0

Document Information

Editor:	Carlo Maria Zwölf
Author(s):	Carlo Maria Zwölf, Guy Rixon
Contributor(s):	Thomas Marquart, Nicolas Moreau

Type of document:	report
Status:	approuved
Distribution:	internal
Work package:	WP2
Version:	1.0

Directory and file name: /RoadMap/AAA_v1.0.docx

Abstract

AAA Strategy

Version	Date	Reason for modification	Modified by
V1.0			

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.

AAA Strategy

TABLE OF CONTENT

1. INTRODUCTION

The goal of this document is to present the Authorisation, Authentication and Accounting (AAA) strategy for the VAMDC infrastructure, while arguing the reasons for the retained choices.
We recall that

-   The authentication consists in checking the identity of the parties involved into a digital exchange.
-   The authorisation is the permission to access and/or execute a given resource.
-   The accounting refers to the tracking of the data accesses and/or resources executed by authorized users at a given time.

Before continuing, we would like to recall that the presented AAA strategy would not be mandatory. Data and software providers should be able to choose if their data/software will be freely accessible or if they would like to apply the AAA approach.

2. VAMDC REQUIREMENTS

The business plans of the VAMDC consortium require a number of AAA abilities.

-   The accounting must track the usage of the infrastructure – databases, services and software modules – by each user community. It is not considered necessary or appropriate to track usage by individual users.

- Usage of certain software modules must be restricted to authorized – licensed – users.
- If, at some future time, VAMDC has subscribers, those subscribers must be authorized to get a higher level of service from VAMDC facilities, including data nodes.
- Operators of data nodes must be allowed to restrict access to some parts of their data to authorized users or user-groups.

3. AUTHENTICATION

3.1 On the Authentication method

The authentication process could be based on different type of credentials [1],[2]: username/password pairs, digital certificates or signatures, one-time tokens.
This last method is not suitable for authentication of loyal customers, since each time they use the service, they should request a new token.
Digital certificates are a heavy solution where both service provider and users have to deal with the complex lifecycle of certificates.
The username/password pairs are the most flexible and convenient method for implementing authentication in the VAMDC infrastructure.

Remark: Digital certificate could be used as internal technical solution for encrypting client/server communications. In any case, these mechanisms will remain invisible to users and these last will have to deal just with username/password for their authentication.

Since our goal is to simplify the user experience while minimizing the implementation/maintenance costs for the VAMDC consortium
- We want to prevent a user who already has dozen of accounts (web-mails, social networks, etc.) to create a new specific VAMDC account.
- We do not want to handle the issue of new password, the registration, the cancellation of users and to code an ad-hoc VAMDC authenticator.

The *delegation of authentication* is a solution to these two last items. At present we recommend to focus on the adoption of (at least one of) these two systems:
- OpenID [3] which is a worldwide project supported by IBM, Microsoft, Google, Yahoo. Moreover the International Virtual Observatory Alliance (IVOA) started adopting OpenID for the authentication into the VO-Space-2.0 protocol. The adoption of the same technical solutions can strengthen the links between VAMDC and IVOA.
- Perun [4] which is an authority delegation system providing the management of virtual organisations for defining membership restrictions and accessible resources among its members. The resources are managed separately from the virtual organisations; therefore the resource owner can define rules for providing accesses to several virtual organisations. This feature is particularly indicated for the VAMDC infrastructure, where we would define different user groups (cf. § 3.2).

AAA Strategy

Both these systems come with a set of libraries and tools that makes easy the implementation of the delegation of authorisation into our services.

3.1.1 Discussion on authentication for accessing data

The authentication, based on a mechanism of delegation (cf. § 2.1), could be implemented on the VAMDC portal, for submitting queries and accessing data. However, since the VAMDC infrastructure is composed of autonomous nodes [5], a shrewd and experienced user could by-pass the authentication step: by using the VSS2 language [6] and the TAP protocol [7], he/she could submit the queries and retrieve the data, directly from each node.
Therefore, some authentication mechanism must operate at each data node or other service that applies authorization checks.
The software and libraries providing access to data (mainly the Portal and the QueryEditor) will implement the mechanism for the delegation of authentication, using the libraries provided by the retained mechanism of delegation (cf. § 2.1): when a user authenticates, they will get information about its membership to the different existing virtual groups. This information will be passed to each VAMDC node, while performing the requests.

The information about the membership of every user could be delegated to the same system realizing the authentication or be stored on a central VAMDC database.

The node-software could provide two distinct entry points for the incoming requests. The first one will be for the *secured* mode and will need as mandatory information the IDs of the virtual group to which the user submitting the request belongs (this information will be encrypted during the client/server communication). The second entry point will be for the non-secured mode: the IDs of the virtual group are optional.

The data providers may choose to activate/deactivate one of these two modes.

Enabling authentication in a service has costs, in both time and money (a server certificate may need to be bought for the web server). Only services that have an authorization policy are expected to enable authentication.

3.1.2 Discussion on authentication for executing software modules

VAMDC does not only provide the data, but also services and processing for handling and transforming the data extracted from the infrastructure. In what follows we will refer to those elements as ***software modules***. These software modules could be embedded into online services (typically processors accessible online from the VAMDC portal) or into standalone applications (e.g. SpectCol [8], SpecView [9]). An AAA policy could be applied for the software modules in accordance with the adopted business model and the standard and technical constraints (cf. § 3.2 and § 4.2 for details on authorisation and accounting): the delegation of authorisation will be

implemented in each software module using the libraries provided by the retained mechanism of delegation (cf. § 2.1).

Remark: Some basic and low-level software modules may remain free, with no access control.

4. AUTHORISATION

We recall that the authorisation is the permission to access and/or to execute a given resource. Following the skeleton of the previous paragraph, we will discuss in different sub-paragraphs the authorisation for accessing data and the authorisation for executing software module.

4.1 Discussion on authorisation for accessing data

Authorized privileges for access data fall into two classes: access to specific, private records in the database; or improved access to the database as a whole, such as the ability to retrieve larger data-extracts. Node operators who opt to restrict access are free to set the specific authorization rules.

Node operators are free to authorized access to specific records in any way that suits their database, perhaps using authorization annotations already set up for a non-VAMDC interface. However, the consortium strongly recommends that authorization rules be based on the (emerging) standard for *unique digital identifiers for data-sets* [11] (hereafter referred as Data-Set-ID). Support for this approach would be built into the node software.

The identities authenticated at the nodes would typically be for user groups rather than individual users. The authorization should be planned accordingly. If a node specifically needs to authorize individuals, then single-user "groups" could be defined.

4.2 Discussion on authorisation for software modules

The authorisation could be implemented for *software modules*: we will maintain a central database containing the identifications of all the *software modules* and, for each module, the list of the virtual group of users having the authorisation for its execution.
The organisation of the virtual groups and the membership of the users to the different virtual groups (a given user can be a member of multiple groups) could be managed on the same central database or delegated to the systems already used for the delegation of authentication (cf. § 2.1).
Before the execution of a given software module, the user must be authenticated. The module will then check into the central database if the user has the authority for the execution.

The mechanism for performing the authorisation could be embedded into a specific library to be used for all the *software modules*. A unique software component will be

used, reducing the global development and maintenance time. This library could be created from scratch or re-using existing products (e.g. JAAS [10] for the developments based on JAVA).

5. ACCOUNTING

Considering that the above AA is adopted, we will discuss the Accounting for data and for software modules in the following two subparagraphs. As we saw above, an Authentication/Authorisation strategy could be implemented for accessing data and on software modules. This naturally allows Accounting mechanisms: according to the local and European laws protecting the users' privacy, all the data will be collected in an anonymous form.

5.1 Discussion on the Accounting for the access data

In the most basic version of accounting, we may compile macro-statistics for knowing (for each database node composing the VAMDC infrastructure):
- The amount of data extracted as a function of time (monthly and weekly scale)
- The percentage of availability.

We can however suggest a more elaborated version, based on the Virtual Group membership information, which may be provided to VAMDC nodes while submitting queries (cf. 2.1.1): we may trace, as a function of both time and virtual groups, the number of access to a given data set.

The accounting should operate on all VAMDC services that deal in data. The majority of these services apply no authorization policy, and therefore do not enable authentication. Therefore, the accounting should be based on unauthenticated group-membership.

5.2 Discussion on the Accounting for software modules

For each software module, we will just record the number of users, as a function of time and of virtual-group membership.
This information could be stored into the same central database used for the authorisation of software modules (cf. § 3.2): when a given software module is used, just after passing successfully the authorization phase, the same software module could add an entry into the central database registering
- The identification of the software module executed,
- The execution date and time,
- The virtual-groups to which the user executing the module belongs.

As already suggested while describing the authorisation strategy, the mechanisms for performing the described accounting actions could be imbedded into the same specific library designed for centralising the authorisation procedure.

AAA Strategy

APPENDIX A: SOME GUIDELINES FOR THE IMPLEMENTATION OF AAA STRATEGY INTO VAMDC SOFTWARE MODULES

We strongly recommend embedding all the mechanism and functionality required for implementing the presented AAA strategy for software modules into a unique specific VAMDC library (one for each programming language adopted). This library (let us call it VAMDC-AAA library) should be used for implementing the AAA features in all the software modules we wish to restrict.

A standalone application or a web service may be composed of various distinct software modules. These will be launched at different stage of the execution flow. It is important that the user have to enter his/her username/password once at all for every working session (and not at every execution of a module). For this the username/password should be stored in the scope of the working session end every software module should use this stored information for contacting the central database (cf. § 3.2) for authorisation. These elements should be taken into account for the design of the library.

REFERENCES

[1] C. Rigney, S. Willens, A. Rubens, W. Simpson, "Remote Authentication Dial In User Service (RADIUS)", IETF RFC 2865, June 2000.
[2] Bernard Aboba, Jari Arkko, David Harrington, "Introduction to Accounting Management", RFC 2975, IETF, Oct. 2000.
[3] http://openid.net
[4] http://perun.cesnet.cz/
[5] Reference to a document explaining the actual VAMDC infra
[6] http://www.vamdc.eu/documents/standards/queryLanguage/vss2.html
[7] http://www.ivoa.net/documents/TAP/
[8] http://www.vamdc.org/activities/research/software/spectcol/
[9] http://www.vamdc.org/activities/research/software/specview/
[10] http://www.oracle.com/technetwork/java/javase/jaas/index.html
[11] link to Nicolas document on DOI for datasets

VAMDC

Virtual Atomic and Molecular Data Centre

Consortium Document

SUP@VAMDC Project

List of Consortium backgrounds

Version 1.0

Document Information

Editor: Carlo Maria Zwölf

Author(s):   Carlo Maria Zwölf

Contributor(s):

Type of document: Consortium document

Status: awaiting approval

Distribution: public

Work package: WP2

Version: 1.0

Directory and file name:

Abstract : This document describes the background to the "VAMDC Consortium": ownership, licences, contact, administrator. It corresponds to the foreground of the "SUP@VAMDC" and "VAMDC EU projects. The list of background included in the MoU will be based on that document.

Version	Date	Reason for modification	Modified by
V1.0	21/07/2014	Document Initialisation	Carlo Maria Zwölf

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.

1. INTRODUCTION

This document contains the list of the intellectual productions from the VAMDC and SUP@VAMDC projects. The products listed on paragraph 2 constitute the background of the VAMDC consortium.

2. LIST OF PRODUCTS COMPOSING THE CONSORTIUM BACKGROUND

In the following tables the 'Community' status indicates if the product is considered crucial for operating the VAMDC infrastructure. In this case it is designed as *common.*

Hereafter the acronyms CTOWG, CTQWG and CTSWG stand respectively for *Chair of the Technical Operational Working Group, Chair of Technical Quality Working Group and Chair of Training and support working group* (as defined in the Internal Regulation Documents).

Tableau 1 List of software product

Product Name	Ownership	Main authors	License administrator	License	Community status	Contact
VAMDC portal GUI	80% Paris 5% UCAM 5% UCL 5% OU 5% Koeln	Paris Observatory	Paris Observatory	GPL3	Common	CTOWG and/or CTQWG
SpectCol	100% Paris	Paris Observatory	Paris Observatory	GPL3	Not common	CTOWG and/or CTQWG
Django Node Software	95% UU (Marquart, Griatch) 1,8% Cambridge (Rixon) 1.7% UCL (Hill) 1% Paris 0.2% koeln (Endres)	UU (Marquart)	UU	GPL3	Common	UU (Marquart) and/or CTOWG
Java Node Software	100% Paris Obs.	Paris Observatory	Paris Observatory	GPL3	Common	CTOWG
TAP Validator	100% Paris Observatory	Paris Observatory	Paris Observatory	GPL3	Common	CTOWG and/or CTQWG
VAMDC Taverna Plugin	100% MSSL/UCL	Kevin Benson	Kevin Benson	GPL3	Not Common	Kevin Benson

	(Kevin Benson)					
Dictionary Software	100% UU (Marquart)	UU (Marquart)	UU	GPL3	Common	UU (Marquart) and/or CTOWG
XSAMS Processor "applyXSL"	100% UU (Marquart)	UU (Marquart)	UU	GPL3	Common	UU (Marquart) and/or CTOWG
XSAMS2Sme Processor	50% UU (Marquart) 50% Cambridge (Rixon)	UU (Marquart)	UU	GPL3	Not Common	UU (Marquart) and/or CTOWG
Linespec Processor	50% UU (Marquart) 50% Cambridge (Rixon)	UU (Marquart)	UU	GPL3	Not Common	UU (Marquart) and/or CTOWG
XSAMS to HTML Atomic spectroscopy processor	100% Paris Obs.	Paris Obs	Paris Obs.	GPL3	Not common	Paris team and/or CTOWG
XSAMS to HTML Molecular spectroscopy processor	100% Paris Obs.	Paris Obs	Paris Obs.	GPL3	Not common	Paris team and/or CTOWG
XSAMS to HTML Collisional data processor	100% Paris Obs.	Paris Obs	Paris Obs.	GPL3	Not common	Paris team and/or CTOWG
XSAMS Processor for spectral synthesis (upcoming)	100% UU (Marquart, Piskounov)	UU	UU	GPL3	Not Common	UU (Marquart) and/or CTOWG
Xsams2Bibtex	100% UCL (Rixon)	UCL	UCL	GPL3	Not Common	G. Rixon and/or CTOWG
XSAMSMerger	100% UCL (Rixon)	UCL	UCL	GPL3	Not Common	G. Rixon and/or CTOWG
Query Builder	100% Paris Obs	Paris Obs	Paris Obs	GPL3	Common	Paris team and/or CTOWG
Python Modules	100% Koeln (C. Endres)	Endres	Koeln	GPL3	Not Common	Endres and/or CTOWG

Tableau 2 : List of documents (standards and tutorials)

Product Name	Ownership	Main authors	License administrator	License	Community status	Contact
Tutorials	Equally distributed between members	Not applicable	Paris Observatory	CC4 with derivative work allowed	Common	CTSWG

Document defining VAMDC TAP	Equally distributed between members	Not applicable	Paris Observatory	CC4 without derivative work allowed	Common	CTOWG and/or CTQWG
Document defining standard VSS2	Equally distributed between members	G. Rixon, Nenadovic, Doronin	Paris Observatory	CC4 without derivative work allowed	Common	CTOWG and/or CTQWG
Documentation defining the Dictionar keywords	Equally distributed between members		Paris Observatory	CC4 without derivative work allowed	Common	CTOWG and/or CTQWG

The document defining XSAMS standard is a particular case. Its ownership is shared with IAEA and an external agreement is needed for stating about the licenses and the ownership percentage.

Concerning the content of the databases, of the registries and more generally all the configuration files needed for running third part software (e.g. the astrogrid registry, the tap services, the nagios monitoring, ...) standard copyright holds and a CC4 license without derivative work is adopted.