Virtual Atomic and Molecular Data Center – VAMDC and AOB Node

Present status and perspectives

Andjelka Kovačević

Department of astronomy, Faculty of Mathematics, University of Belgrade

History

- 2003: Start of Project about « interoperability of A&M Data » (Bird of Feather session at ADASS)
- 04/09: Definition of standards starts in IVOA and in NIST/IAEA/Obs. Paris/ORNL collaboration
  - Simple Line Access Protocol (SLAP) and Simple Spectra Line Data Model (SSLDM) → VERY SIMPLE, but they have already implemented tools –
    - http://www.ivoa.net/Documents
  - XSAMS: XML Schema for Atoms, Molecules and Solids
- July 09: Virtual Atomic and Molecular Data Centre (Coordinator: ML Dubernet)

Mission (official)

"VAMDC … an interoperable e-Infrastructure … exchange of atomic and molecular data. … 15 administrative partners representing 24 teams … scientists from a wide spectrum of disciplines in [AM] Physics … users of their AM data (astrochemistry, atmospheric physics, plasmas) …"

(Abstracted from the project summary on vamdc.eu)

VAMDC Partners

N. Walton (Cambridge, Euro-VO)
H. Mason (Cambridge, CHIANTI)
J. Tennyson (UCL, Hitran)
Len Culhane (UCL, Euro-VO)
N. Mason (OU)
T. Millar (QUB, UMIST)
S. Schlemmer (Koeln, CDMS/JPL)
N. Piskunov (Uppsalla, VALD)
W. Weiss (Vienna, VALD)
T. Ryabchikova (INASAN, VALD)
A. Ryabtsev (ISRAN, VALD)
M. Dimitrijevic (Belgrade Obs, Stark-B)
P. Loboda (RFNC-VNIITF, SPECTR-W3)
V. Perevalov, A. Fazliev (IAO, O3, CO2)
Y. Ralchenko (NIST)

G. Mulas (Cagliari Obs., PAH)
G. Leto (Catania Obs., LASP)
C. Mendoza (IVIC, Venezuela, TipTopBase)
V. Tyuterev, A. Barbe (GSMA, O3)
B. Schmitt (LPG, STSP)
V. Wakelam (Bordeaux Obs., KIDA)
V. Boudon (Dijon, CH4)
P. Le Sidaner (Paris Obs., VOPARIS)
C. Zeippen (Paris Obs, TipTopBase)
ML Dubernet (LPMAA, BASECOL)
C. Joblin (CESR, Toulouse)
E. Roueff (LUTH, Paris, H2 Data and Grid)

		LPMAA, VOPARIS, LERMA, LPG, LJAB, GSMA, ICB
1	CNRS (FR)	**BASECOL, KIDA, STSP, CH4, O3, PAH , TipTopBase, Technological Node** *Contact NIST, EuroPlanet , IVOA + all WP (apart WP3)* ISM, Planetology, Plasma, Atmosphere, Stellar, Solar
2	UCAM (UK)	**CHIANTI, AstroGrid (Euro-Vo) -** QA - Solar, Stellar (2 departments)
3	UCL (UK)	**HITRAN, GEISA, AstroGrid (Euro-VO)**- *Software – Support to users* (2 departments) Atmosphere, Planetology, Solar, Plasmas
4	OU (UK)	*Dissemination/Training - Software - Plasma, Industry, EuroPlanet*
5	UW-A (Austria)	**VALD** - *Standards/Software* Stellar, Solar, Plasma, Planetology, Atmosphere
6	UU (Sweden)	**VALD** – *Training/Dissemination - Interoperability* Stellar, Solar, Plasma, Planetology, Atmosphere
7	KOLN (Germany)	**CDMS and JPL** – *Software - Interoperability* ISM, Planetology, Atmosphere
8	INAF (IT)	**PAH and LASP** (2 Institutes)- *Software* ISM, Planetology
9	QUB (UK)	**UMIST Reaction Database**– *Dictionnaries/XML* ISM, Planetology, Atmosphere
10	AOB (Serbia)	**BELDATA** - *Training/Dissemination* Solar, Stellar
11	INASAN (Russia)	**VALD** - *Training Tools– User Requirements* Stellar, Solar, Plasma, Planetology, Atmosphere
12	RFNC-VNIITF (Ru	**SPECTR-W3** - *Quality Assurance – Monitoring – Software – Support to users* - Plasma, Stellar, Solar, Plasma
13	IAO (Russia)	**CO2, O3** - *Software - Training/Dissemination* Atmosphere
14	IVIC (Venezuela)	***TipTopBase***- *Training/Dissemination – GRID* - Stellar, Solar, Plasma, Planetology, Atmosphere – (2 Institutes)
15	ISAN (Russia)	**VALD** - *Quality Assurance* - Stellar, Solar, Plasma, Planetology, Atmosphere
	NIST (USA)	
	CFA (USA)	

A&M Challenges

- A&M data underpins many areas of research
  - Providing access to a wide range of users (astronomy, nuclear, climatology, biology) in academia and industry
- Data is complex and increasingly large
- Handling of data (often) involves use of applications
- Issues with ensuring data completeness & quality
- Coordination and standards – organising the A&M community

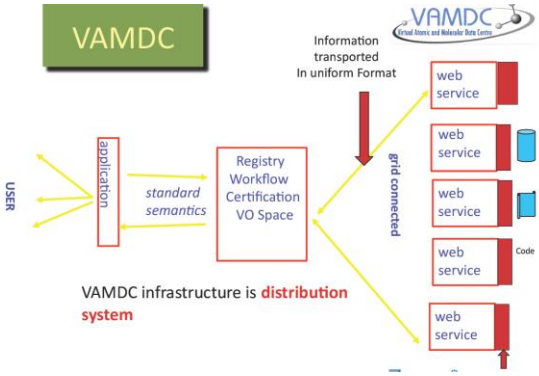
Challenge: provide data access to **all** A&M data to **all** end user communities

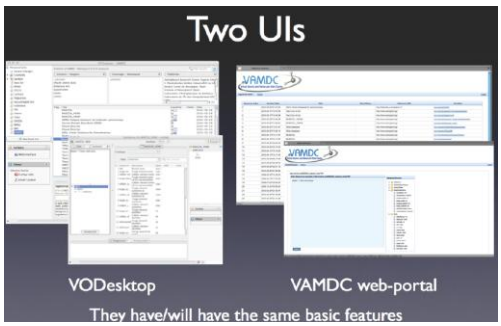
Key VAMDC Objectives

- implement VAMDC interface for accessing major existing databases containing heterogeneous data and aimed at different users
- enable data queries across multiple DBs that are focussed on specific research topic(s)
- enable data publishing/quality control process for major A&M data producers
- involve wide user and producer communities in development and use of VAMDC

Attract a manifold user community

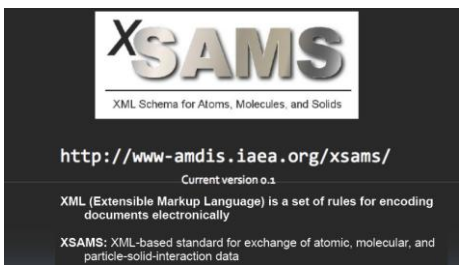
- **Astronomers** working in **far UV to radio regions**
- studying objects from **extrasolar planets to supernovae**
- **magnetic field measurements,**
- **modelling of stellar atmospheres etc.**
- **Molecular spectroscopy**
- **Atomic spectroscopy**
- **Plasma physics**
- **Atmospheric optics**
- **Gas discharge lamps**
- **Laser spectroscopy**
- **Solid state spectroscopy**
- **Quality check of materials**
- **Chemistry**





The data

- *Lists* and *tables* of:
  - Atomic/molecular states
  - Transitions between states
  - Lines arising from atomic transitions
- *Not* images; rarely spectra
- From lab measurement, or from theory
- Many "small" DBs: MB up to ~10GB



Why

- There is a need for a unified, standardized approach to A&M data exchange

Goal of XSAMS

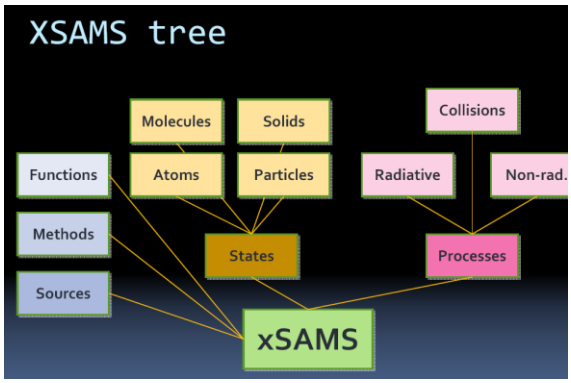
- To develop a (rather) complete set of rules (tags, document structure, relations, etc.) for description of XML A&M data sets

A&M world: $\langle \Psi_i | \hat{O} | \Psi_j \rangle$

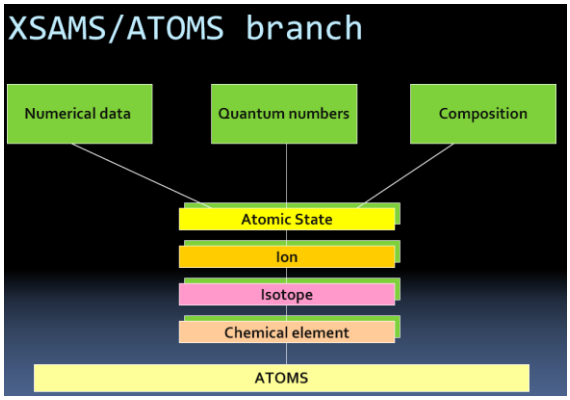
Tools for development of data exchange standards

- eXtensible Markup Language (XML)-based technologies
- XML facilitates the sharing of data across different systems, particularly systems connected to the internet
- XML is an important medium for exchanging, integrating, and storing data from diverse sources
- XML separates content from presentation
- **XML is a metalanguage, i.e., a tool for development of new languages**

XSAMS tree



XSAMS/ATOMS branch



The current Schema

The J=22, K=10 state of the $(1,0^+,0^0,2^2)$ vibrational state of NH_3

```
<ElectronicHome>
 <ElectronicComponent>
  <VibrationalHome>
   <VibrationalComponent>
    <VibrationalQuantumNumbers>
     <VibrationalNu>
      <Label>v1</Label>
      <Value>1</Value>
     </VibrationalNu>
     <VibrationalNu vibrationInversion="s">
      <Label>v2</Label>
      <Value>0</Value>
     </VibrationalNu>
     <VibrationalNu vibrationLNu-l="0">
      <Label>v3</Label>
      <Value>0</Value>
     </VibrationalNu>
     <VibrationalNu vibrationLNu-l="2">
      <Label>v4</Label>
      <Value>2</Value>
     </VibrationalNu>
    </VibrationalQuantumNumbers>
```

```
<RotationalHome>
 <RotationalComponent>
  <NonLinearPolyatomic>
   <NonLinearNoElecNoHyperF>
    <TotalAngularMomentumN>
     <Label>J</Label>
     <Value>22</Value>
    </TotalAngularMomentumN>
    <MolecularProjection>
     <TotalMolecularProjectionN>
      <Label>K</Label>
      <Value>10</Value>
     </TotalMolecularProjectionN>
    </MolecularProjection>
   </NonLinearNoElecNoHyperF>
  </NonLinearPolyatomic>
 </RotationalComponent>
</RotationalHome>
</VibrationalComponent>
</VibrationalHome>
</ElectronicComponent>
</ElectronicHome>
```

AOB INVOLVED IN

WP3 : NA2: Dissemination and Training

1) Coordination
2) Organizing an annual meeting and arranging representation at other relevant meetings
3) Organizing themed scientific workshops
4) Organizing training tutorials

STARK B DATABASE
IMPLEMENTATION: N. Moreau
Design: M.S.Dimitrijević, S. Sahal-Brechot

Currently: 43 elements

Fully documented and referenced

Povides:
Wavelength

Parameter C (for the validity condition of the isolated line approximation)

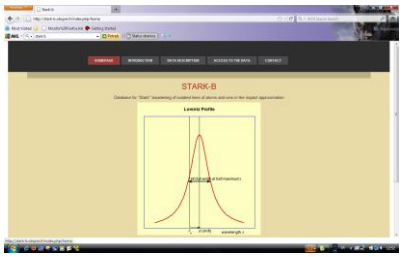
Temperature

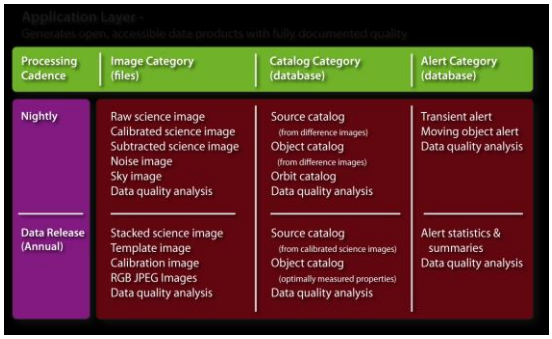
Quasistatic parameter (for Neutral atoms)

FWHI for electron ,proton colliders

Shift for electron , proton colliders

Fully checked and evaluated





The LSST scientific database will include:

Over **100 database tables**

Image metadata consisting of **700 million rows**

A source catalog of with **3 trillion rows**

An object catalog with **20 billion rows each with 200+ attributes**

A moving object catalog with **10 million rows**

A variable object catalog with **100 million rows**

An alerts catalog. Alerts issued **worldwide within 60 seconds**.

Calibration, configuration, processing, and provenance metadata

The science archive will consist **of 400,000 sixteen megapixel images per night (for 10 years),** comprising **60 PB** of pixel data.

This enormous LSST data archive and object database enables a diverse multidisciplinary research program: astronomy & astrophysics; machine learning (data mining); exploratory data analysis; extremely large databases; scientific visualization; computational science & distributed computing; and inquiry-based science education (using data in the classroom). Many possible scientific data mining use cases are anticipated with this database. The advances in these technology areas will be exported to other big data science applications (biology, remote sensing, etc) and will drive innovations in industry. Already a collaboration is forming between industry and LSST on the design of extremely large databases.