VAMDC and Taverna Workflows

Kevin Benson

Misha Doronin

http://www.vamdc.eu

http://www.taverna.org.uk

VAMDC Project origins

Many atomic and molecular databases exist in all over the world. Each of them was created for a specific scientific use case that influenced it's structure and contents. Those databases overlap and complement each other.

But each and every database has three specific features:

- Specific internal structure and output format
- Specific data extraction procedure
- Specific user and/or application interface

What is VAMDC?

Virtual Atomic and Molecular Data Centre (VAMDC) is an international project aimed to create well documented interoperable interface to the differently organised existing Atomic and Molecular databases.

VAMDC consortium involves 15 administrative partners representing 24 teams from 6 European Union member states (France, Austria, Germany, Italy, Sweden, United Kingdom), Serbia, the Russian Federation and Venezuela. The project leader institution is CNRS (France).

The project coordinator is Marie-Lise Dubernet

marie-lise.dubernet-tuckey@upmc.fr

VAMDC members

Centre National de la Recherche Scientifique - CNRS

Universite Pierre et Marie Curie, Observatoire de Paris, Universite Bordeaux 1
Universite Joseph Fourier de Grenoble, Universite de Bourgogne,
Universite de Reims, Universite Toulouse 3

Cambridge University (CMSUC)
University College London (UCL)
Open University (OU)
University of Vienna (UNIVIE)
Uppsala University (UU)
University of Koeln (UK)
Instituto Nazionale di Astrofisica (INAF)
Queen's University of Belfast (QUB)
Astronomical Observatory of Belgrade (AOB)
Institute of Astronomy RAS (INASAN)
Russian Federal Nuclear Centre All-Russian Institute of Thechnical Physics (RENC-VNITF)
Institute of Atmospheric Optics SB RAS (IAO)
Corporacion Parque Technologico de Merida (CECALCULA-CPTM, Venesuela)
Institute for Spectroscopy RAS (ISAN)

Atomic Line and Process Databases

VALD (atoms): Over 150 line lists, ~1.1 million atomic lines for detailed spectral analysis and over 250 million lines for opacity calculations (predicted lines) provided by all major spectroscopy centres across the world. Created in 1995 for detailed modelling the spectra and the atmospheric structure of stars hotter that SUN, including magnetic CP stars. Radiative data for 84 elements up to Z=92 (U) in first 9 (Fe-peak) ionization stages.

VALD (subset of observed lines): Over 150 line lists, ~1.1 million atomic lines for detailed spectral analysis.

CHIANTI: database for the analysis of optically thin collisionally-ionised plasmas. The preferred reference database in solar physics. Contains wavelengths and radiative data and rates for electron and proton collisions for highly charged ions with Z≤30 (up to Zn)

TOPbase (with OPserver): contains the radiative atomic data computed in LS-coupling) and photoionization cross-sections of elements with Z=1-14,16,18,20,26. Used for mean opacity calculations in stellar modelling, in NLTE line formation calculations

Stark-b: database of calculated widths and shifts of isolated lines of atoms and ions due to electron and ion collisions. Used for spectroscopic diagnostics of stellar atmospheres and envelopes

Atomic Line and Process Databases

Spectr-W3: For ~450000 transitions contains experimental, calculated and compiled data on energy levels, wavelengths, transition probabilities. Parameters for analytical approximation of electron cross-sections. The completeness is for multicharged ions.

LUND laboratory spectroscopy database: Experimental data for transitions and lifetimes (currently data for Nb II and Nb III)

DESIRE: Contains updated information about wavelengths, energy levels, transition probabilities, Lande-factors and radiative lifetimes for neutral singly and multiply ionized elements belonging to the sixth row of periodic table. Original DB is in Liege, DESIRE-VAMDC node is temporarily in Moscow site (currently contains data for Os I and Os II).

Molecular Line and Process Databases

HITRAN: **High-**resilution **TRAN**smission molecular absorption database. The largest database for atmospheric reasearch and planetology. The line-by-line portion of the DB contains spectroscopic parameters for 42 molecules. It also contains absorption cross-section parameters and aerosol indices of refraction. In total, HITRAN currently contains data for about 2×10^6 individual transitions.

CDMS: The Cologne Database for Molecular Spectroscopy provides recommended values of transition frequences and intensities for molecules. Used for the studying the Earth atmosphere in 0 – 340 cm^{-1}. CDMS is cross correlated with its US counterpart, the JPL submillimeter catalogue, which is now the member of VAMDC consortium as well.

BASECOL: provides excitation rate coefficients for ro-vibrational excitation of molecules by electrons, He and H_2. BASECOL is mainly used for the study of interstellar and circumstellar medium and of cometary atmospheres

PAH: polycyclic aromatic hydrocarbon spectral database provides a number of properties (general energetics, ground state optimised geometries, harmonic vibrational frequencies, IR activities and photoabsorption cross-sections) for a sample of about 60 species in four charge states: anion, neutral, cation and dication. Used for diagnostics of interstellar medium

Molecular Line and Process Databases

CDSD: Carbon Dioxide Spectroscopic Databank contains calculated spectral line parameters for 7 isotopologues of carbon dioxide in the 5-12784 cm^{-1} wavenumber range. CDSD contains 419610 lines.

S&MPO: (Spectroscopy & Molecular Properties of Ozone) database contains a well-established data on spectral line parameters for the ozone molecule (experimental UV cross-sections, for example, and calculated linelists for $^{16}O_3$, $^{16}O^{18}O^{16}O$, $^{18}O_3$)

UMIST: database for astrochemistry provides a fundamental set of reaction rate data and related software for use in chemical kinetic modelling of astronomical regions (star formation, for example)

KIDA: **KI**netic **D**atabase for **A**strochemistry that will contains all chemical reactions used in the modelling of the chemistry in the interstellar medium and in planetary atmospheres.

ICB Dijon Metane: Vibration-rotation calculated linelists with wavelengths and line intensities in 0 – 6200 cm^{-1} region ($^{12}CH_4$, $^{13}CH_4$, $^{12}CH_3D$)

Molecular Line and Process Databases

IDEADB: Innsbruck Dissociative Electron Attachement Database contains data on interaction of low-energy electrons with the molecules

W@DIS: Water Internet Accessible Database Distributed Information System provides energy levels, wavenumbers, transition probabilities of the water lines.

GSMA Reims Ethylene: calculated data for $^{12}C_2H_4$ vibration-rotation energy levels, wavelengths and line intensities in 500 – 7500 cm^{-1} region

Solid State Spectroscopy Databases

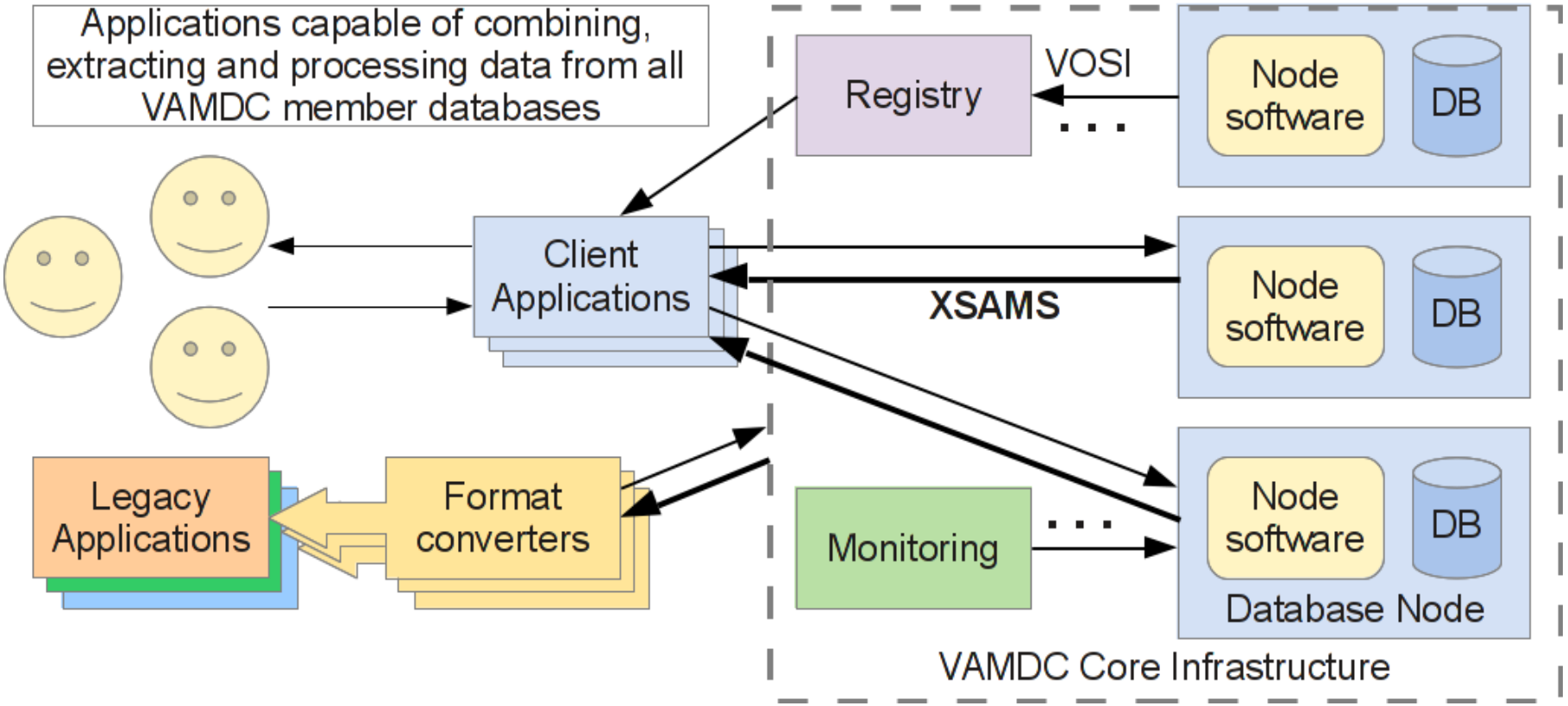
OACT-LASP database: IR spectra of molecules in solid state

GhoSST database: spectroscopy of thin films and solids

VAMDC portal
http://portal.vamdc.org/

- First entry point for new-coming users
- Overview of accessible databases
- Query generator
- Available data estimate
- Download or process result XSAMS documents
- Persistent query log

VAMDC Infrastructure

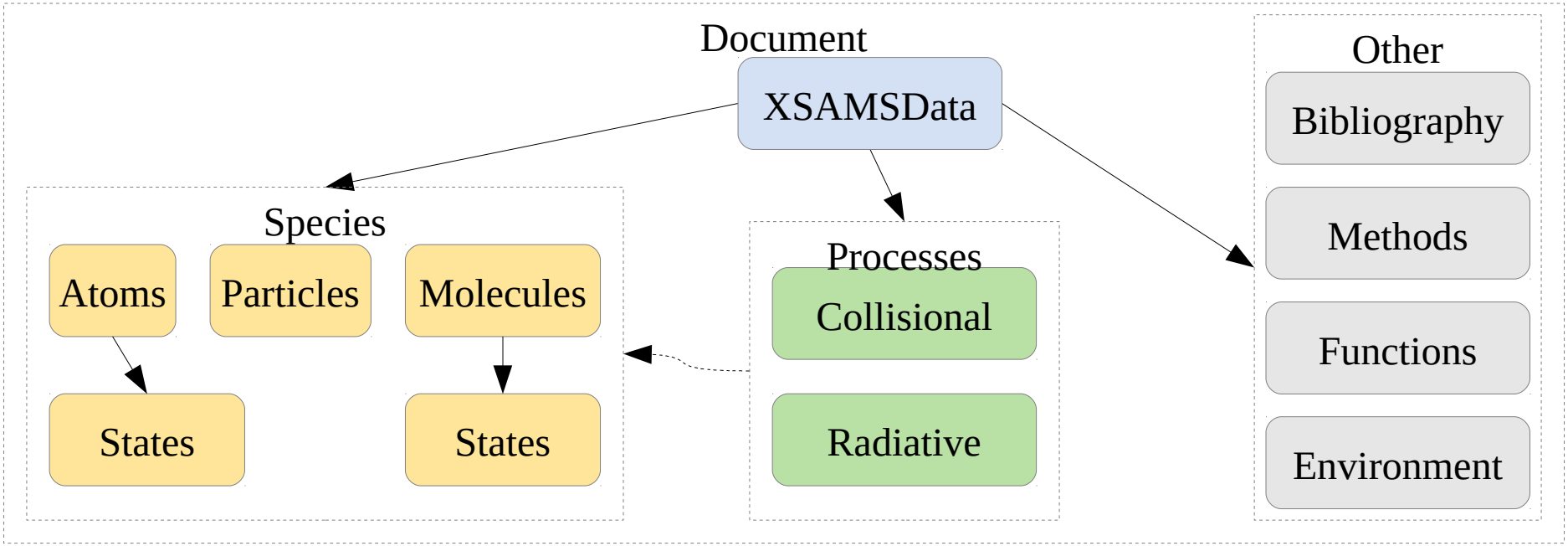


Registry

- VAMDC uses the standard IVOA Registry to describe all the VAMDC Resources.
- Extensions were developed to describe TAP Nodes and their abilities on what could be returned and what restrictions there are to queries.
- Clients query the Registry to get details on the Resources such as Descriptions, Locations, what is available to query for Tap Nodes and which version of the standards is implemented.

```
</content>
<capability standardID="ivo://vamdc/std/VAMDC-TAP" xsi:type="ns4:VamdcTap">
  <interface xsi:type="ns5:ParamHTTP">
    <accessURL use="base">http://dev.vamdc.org/basecol/tapservice_12_07/TAP/</accessURL>
    <accessURL use="base">http://batz.lpma.jussieu.fr:8080/tapservice_12_07/TAP/</accessURL>
  </interface>
  <versionOfStandards>12.07</versionOfStandards>
  <versionOfSoftware>Java VAMDC-TAP node implementation v.12.07r1</versionOfSoftware>
  <sampleQuery>select species where atomsymbol like '%';</sampleQuery>
  <sampleQuery>Select * where reactanta.MoleculeStoichiometricFormula="CO" and reactantb.AtomSymbol = "H";</sampleQuery>
  <returnable>MethodCategory</returnable>
  <returnable>MethodDescription</returnable>
  <returnable>MethodID</returnable>
  <returnable>CollisionComment</returnable>
  <returnable>CollisionID</returnable>
  <returnable>CollisionMethod</returnable>
  <returnable>CollisionCode</returnable>
  <returnable>CollisionRef</returnable>
  <returnable>SourceAuthorName</returnable>
  <returnable>SourceCategory</returnable>
```

Data model: VAMDC-XSAMS



Self-contained XML documents
Full description of processes
Full description of involved species, if available
Links between blocks through unique IDs/IDREFs
Bibliographic references for all data
Reference data producers, not only the aggregator database

XSAMS physical quantity container: DataType

- Value, Units
- Accuracy information
- Evaluation (Recommendation)
- Bibliography references
- Reference to method of data acquisition (experiment/theory, with own bibliography)

Data access protocol: VAMDC-TAP

- Based on IVOA TAP
- RESTful protocol: each document has an unique URL
- XSAMS as an output
- SQL-like query language
- HEAD request for the result estimation
- HTTP status codes to identify missing data
- HTTP header information for data availability estimation
- Typical query: select * where

  ((AtomSymbol = 'C' AND IonCharge >= 3 AND IonCharge <= 4))

http://www.vamdc.eu/documents/standards/dataAccessProtocol/vamdctap.html

VAMDC-TAP Node access

- Construct VSS query string

  (look at http://dictionary.vamdc.org/restrictables/ for keywords)
- Determine from the registry what nodes could understand the query
- Try all relevant nodes with HTTP HEAD request
- Refine query or HTTP GET data in XSAMS format
- Do not forget gzip transfer compression

HITRAN OCS query

Node

http://vamdc.mssl.ucl.ac.uk/node/hitran/tap/sync?

TAP

LANG=VSS2&REQUEST=doQuery&FORMAT=XSAMS&

Query

QUERY=select+*+where+%28RadTransWavelength+%3E
%3D+2.0E7+AND+RadTransWavelength+%3C
%3D+3.0E7%29+AND+%28%28InchiKey+%3D+
%27JJWKPURADFRFRB-UHFFFAOYSA-N%27%29%29

Wavelength in A

select * where

(RadTransWavelength >= 2.0E7 AND
RadTransWavelength <= 3.0E7) AND

OCS main isotope

((InchiKey = 'JJWKPURADFRFRB-UHFFFAOYSA-N'))

XSAMS Processor

- Processor services that take VAMDC-TAP URL queries of XSAMS data and translate the data to some other format.
  - HTML
  - CSV
  - SME – Spectroscopy Made Easy Application format.
  - Line Image Plotting
  - BibTeX

XSAMS processing: two ways

- Load as DOM, handle programmatically as related objects tree
- Use XSLT (eXtensible Stylesheet Language Transformations) to convert to other (legacy) formats.
  - Stylesheets can easily be shared as XSAMS Processor services

Taverna and VAMDC

- VAMDC has constructed a plugin for the Taverna Workflow System.
- Instructions and User Guide, including movies on Installation and Workflow building, can be found here:
  - http://vamdc.mssl.ucl.ac.uk/taverna/vamdc
  - http://voparis-twiki.obspm.fr/twiki/bin/view/VAMDC/UserGuides

Service panel

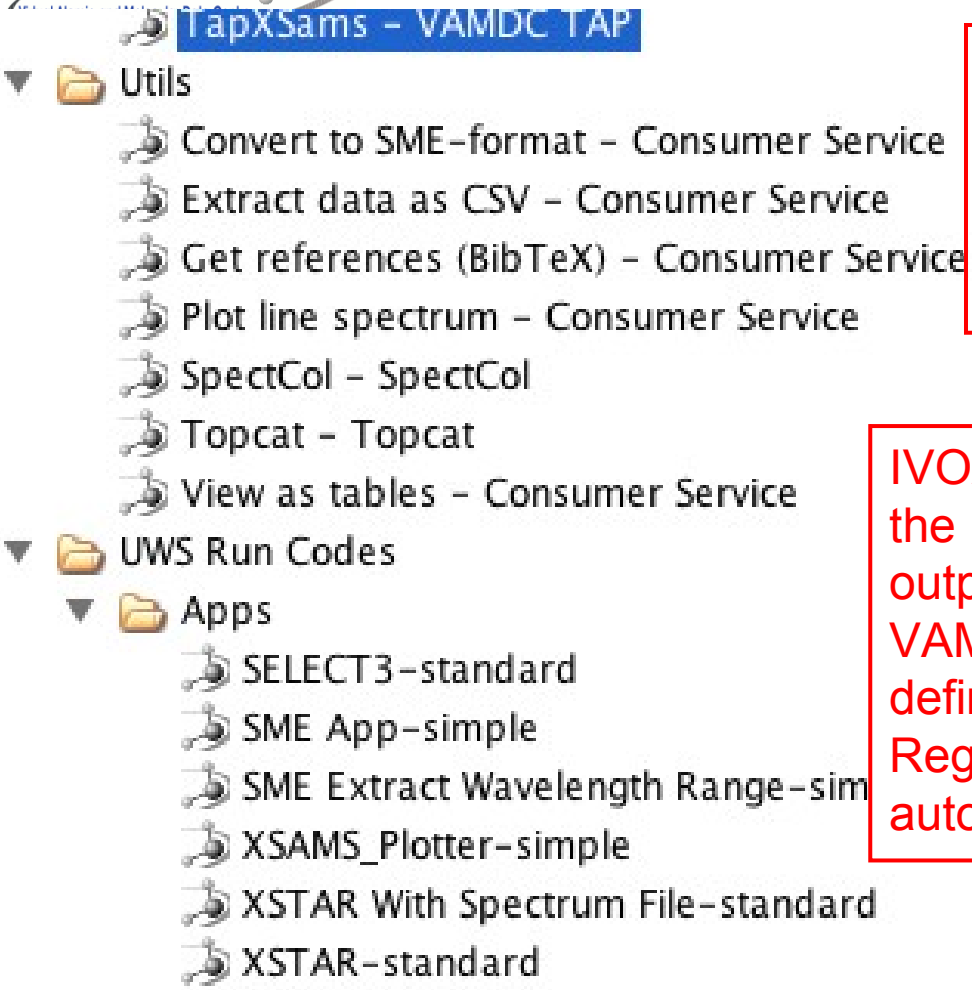
Filter:

Import new services

- Available services
  - Service templates
  - Local services
  - Astro tools
  - Tools decribed @ http://taverna.nordugrid.org/sharedRepository/xml.
  - VAMDC
    - Registry
    - TapXSAMS
      - Combo Operator
      - Nodes
        - TapXSams - VAMDC TAP
    - Utils
    - UWS Run Codes
      - Apps
  - WSDL @ http://eutils.ncbi.nlm.nih.gov/entrez/eutils/soap/eutils.wsdl

- GSMA Reims S&MPO
- HITRAN-UCL resource
  - InchiKey
  - MoleculeChemicalName - Molecule name
  - MoleculeInchiKey
  - MoleculeStoichiometricFormula - Molecular stoichiometric formula
  - RadTransProbabilityA
  - RadTransWavelength - Wavelength Units in 'A'
  - RadTransWavenumber
  - URL - Can be used for TapXSams Service
- ICB Dijon Methane
- IDEADB - Innsbruck
- JPL database: VAMD
- JPL: VAMDC-TAP ser
- KIDA: 12.07 VAMDC
- KIDA: VAMDC-TAP i
- Lund laboratory spec
- OACT - LASP Databa
- Spectr-W3
- SpEctroScopy of Atoms and Molecules
- Stark-b

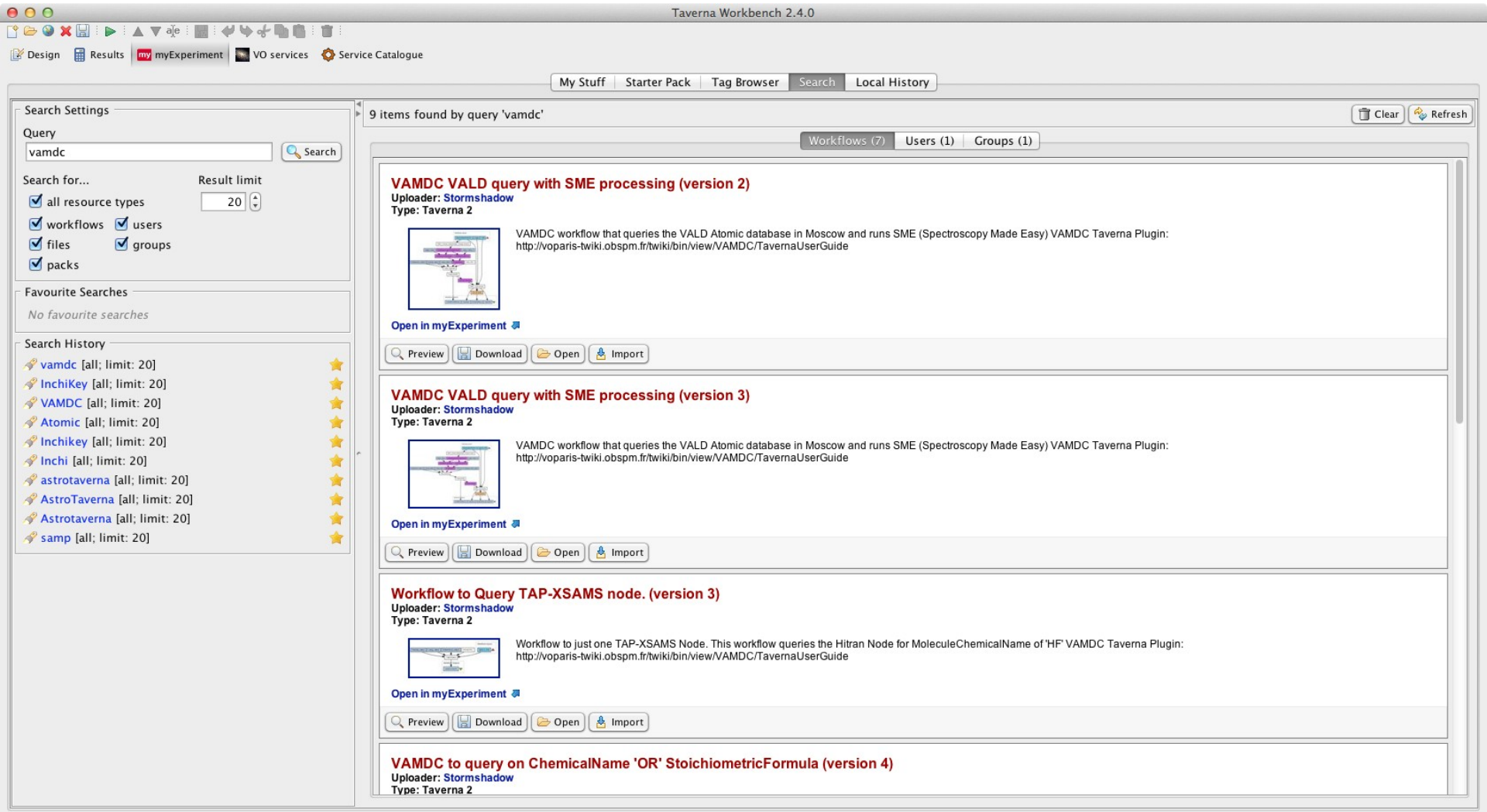
Tap Nodes From the Registry are shown
As Available services to query
Along with the Restrictables.
Users can input a Full Query or build it using the Restrictables.

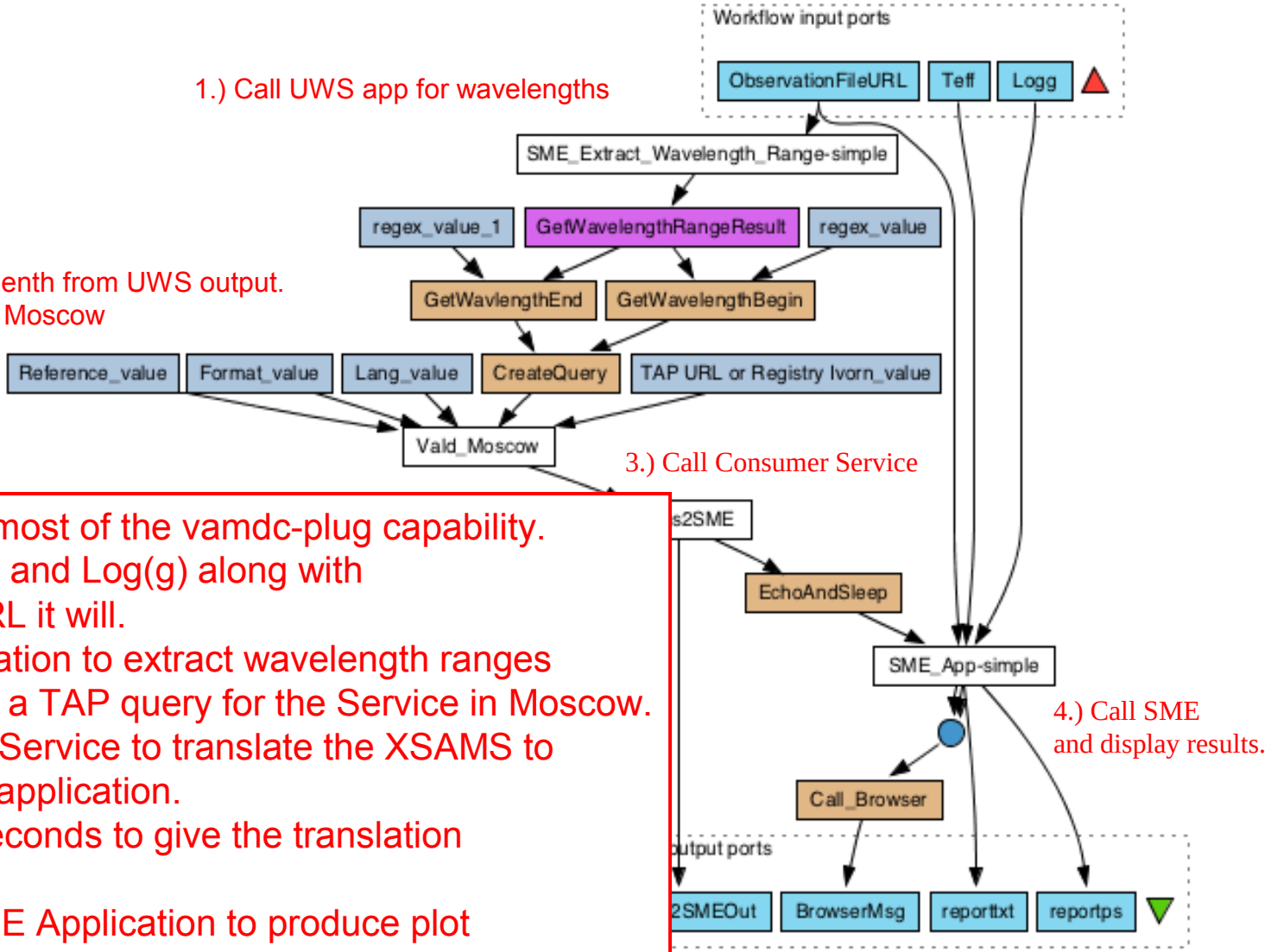


XSAMS Processors and Special Applications such as Topcat and SpectCol are bundled automatically with the plugin.

IVOA-UWS Applications can be read from the Registry for all the needed inputs and outputs to call a particular Applications. VAMDC also has SoapLab applications defined, but currently not read from the Registry and shown as Available services automatically in the plugin.

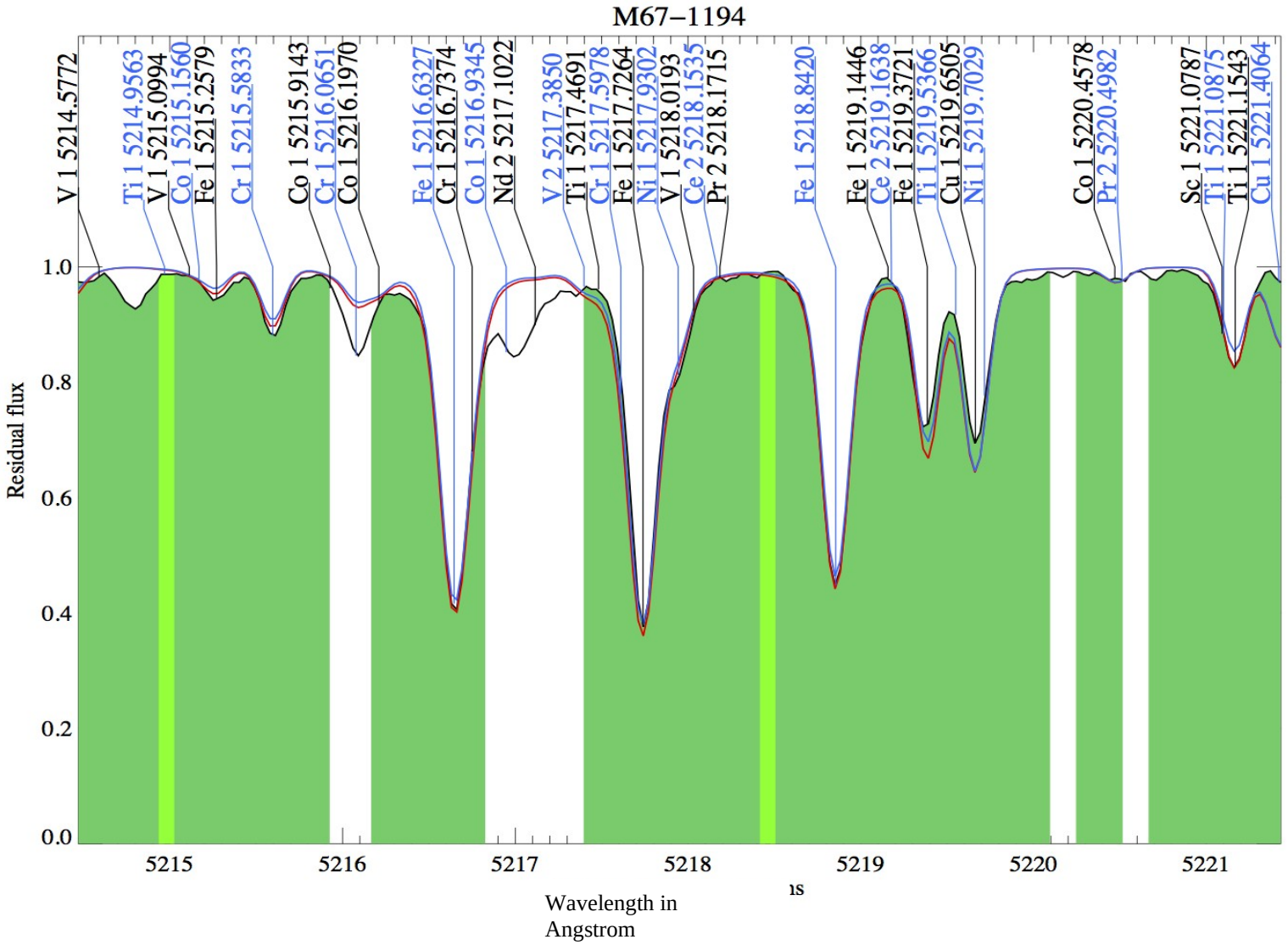
Workflows to allow users to interact with the VAMDC services are Uploaded to myExperiment. All with several tags, but 'VAMDC' is always tagged for the workflows and can be used to see all the workflows Contained in myExperiment. All workflows have example inputs set.





Workflow that shows most of the vamdc-plug capability.
Given a Temperature and Log(g) along with
an observationFile URL it will.
1.) Call a UWS application to extract wavelength ranges
2.) Call and Construct a TAP query for the Service in Moscow.
3.) Calls a Consumer Service to translate the XSAMS to
A format for the SME application.
*Waits for about 10 seconds to give the translation
Service time to finish.
4.) Finally calls an SME Application to produce plot
And report. Send results to the browser.

Result of SME workflow



Recap of VAMDC-Plugin

- A TAPXSams Service is the primary service in the plugin for querying TAP Nodes. Mostly used for VAMDC at this point, but has been tested on IVOA and Helio-Tap Nodes as well.
  - This is primarily due to the fact that it simply builds up a URL to the particular Service.
  - It also does not care about the output hence Votable or XSAMS. A 'Reference' boolean input tells the service if you wish to construct only a URL as output or return the output back to Taverna.
- Looks up the Registry for all the VAMDC-TAP Nodes.
  - Constructs Available Services to each Node including the Restrictables (think of columns) that can be queries on.
- Looks up the Registry for all the available Consumer Services to translate XSAMS outputs.
- Looks up IVOA-UWS registered to construct an Available Service for querying a UWS service.

Future

- SAMP– SAMP could be used to pass the TAP results to another application on the user's desktop. Similar to AstroTaverna ability with Votables.
- SoapLab – Read a soaplab url in the Registry for processing and showing as an available service. This should be similar to importing a soaplab service that Taverna already handles.
- UWS – Is using an older Astrogrid CEA service with its own job description language. A change to a different style UWS and/or PDL is foreseen in the near future
- The plugin does not make clear the versions of XSAMS which are being returned.
- Possible XSAMS Renderer similar to AstroTaverna for Votables. Note though most of the workflows are sending URL's around.
- A start on cross-domain activity with Helio in the workflows has begun and needs to be completed.