VAMDC tutorial for prospective data-providers

Guy Rixon

SUP@VAMDC meeting, IPR, November 2013

Agenda

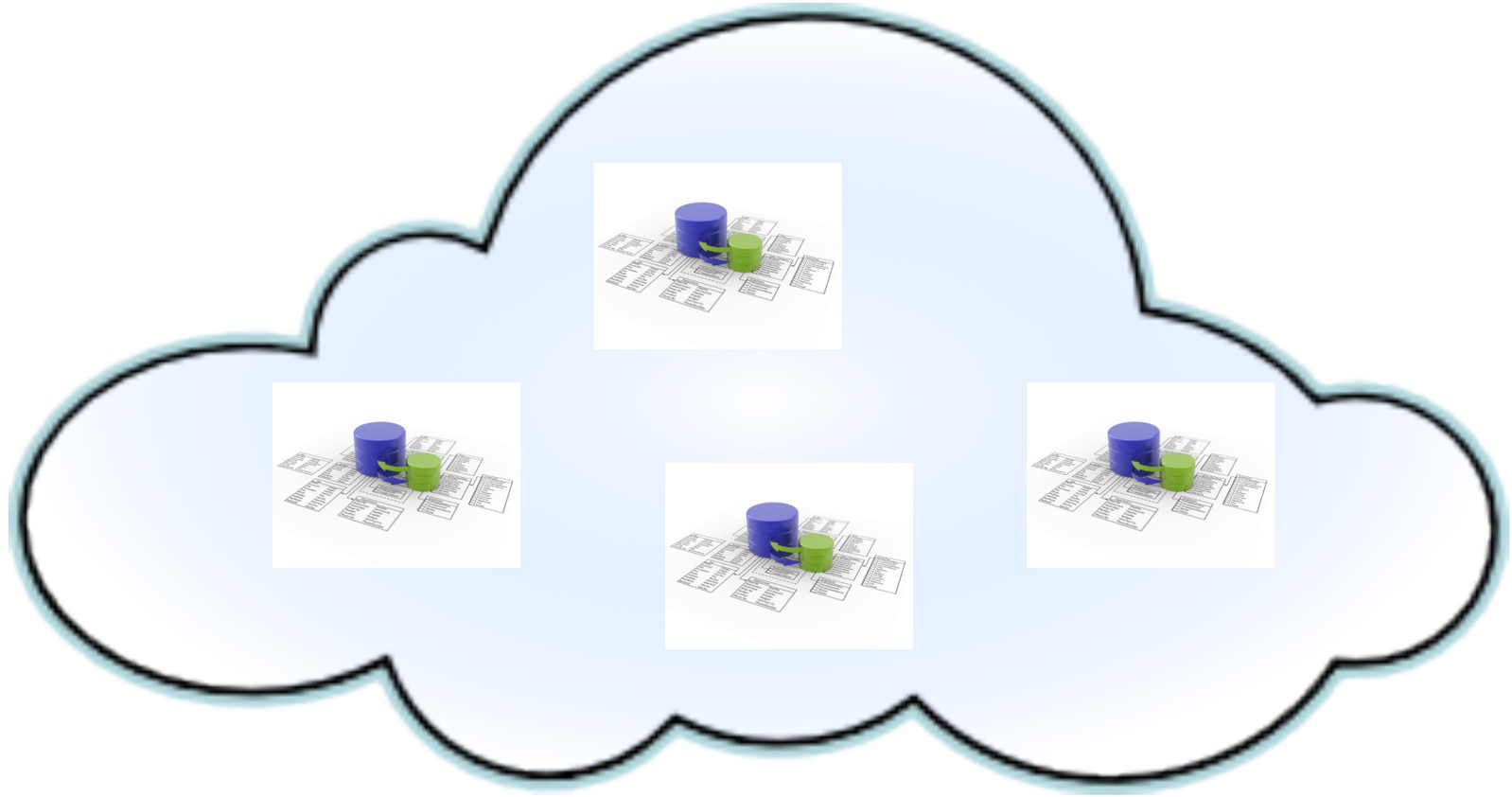
- VAMDC orientation
- Introduction to node building (technical)
- Self-paced investigation with on-line tutorial material

Orientation

VAMDC is in the cloud

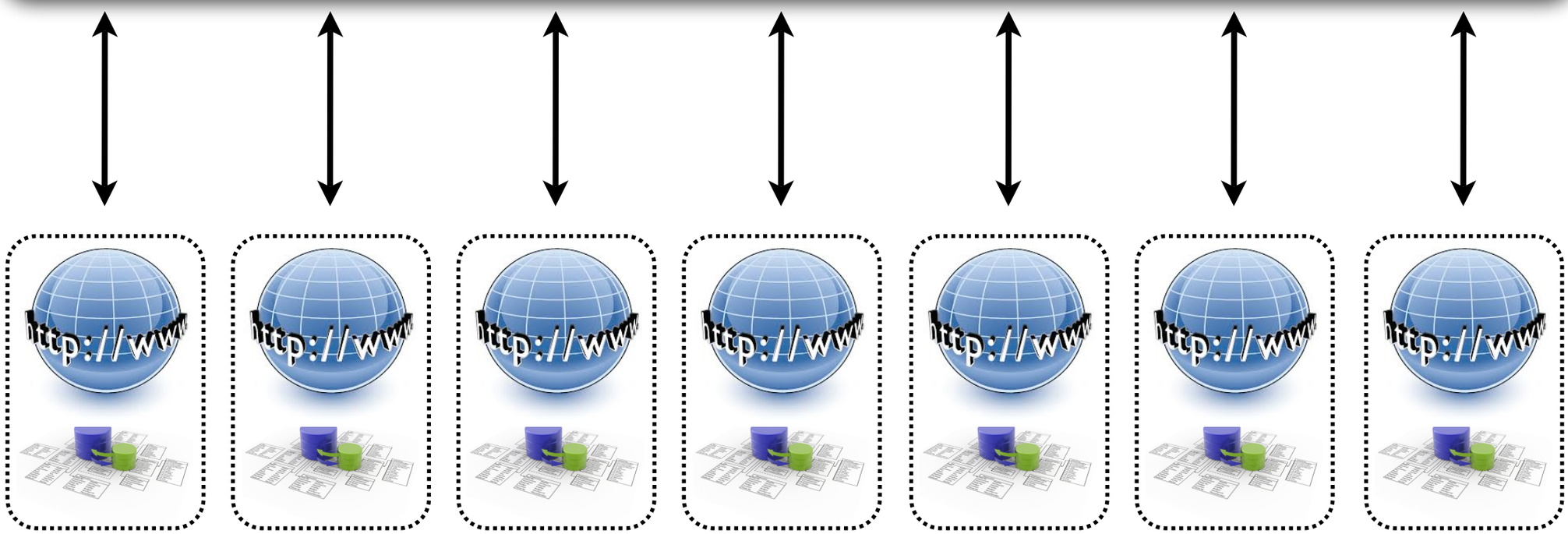
Query

Extract



A flock of databases

Science application



VAMDC data nodes

For list of databases see:

http://portal.vamdc.eu/vamdc_portal/nodes.seam

VAMDC
Virtual Atomic and Molecular Data Centre

Home VAMDC databases Query Saved queries | Info Known issues Login Register

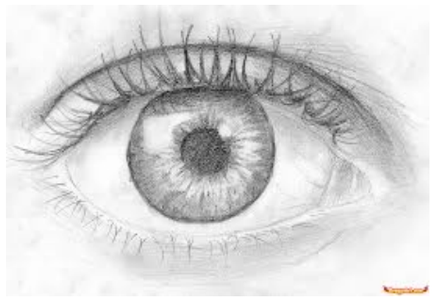
Name	Description	Maintainer	Status
Theoretical spectral database of polycyclic aromatic hydrocarbons	The Cagliari/Toulouse PAH database is a collection of theoretical spectroscopic data about Polycyclic Aromatic Hydrocarbons and carbon clusters. It provides basic geometric characteristics, energetics, harmonic analyses and electronic photoabsorption data. It is maintained by the Astrochemistry group at INAF-Observatory of Cagliari and the Institut de Recherche en Astrophysique et PlanÃ©tologie in Toulouse.	gmulas@oa-cagliari.inaf.it	OK
Chianti	Chianti consists of a critically evaluated set of up-to-date atomic data, together with user-friendly programs written in Interactive Data Language (IDL), to analyse the spectra from astrophysical plasmas. The VAMDC interface presents just the data from the Chianti-v7 release.	gtr@ast.cam.ac.uk	OK
GSMA Reims S&MPO	Calculated line lists for ozone (16O3, 16O18O16O and 18O3). The data on methane contain the vibration-rotation energy levels, line positions and line strengths in the range from 0 to 8000 cm-1.	ylb@iao.ru, vladimir.tyuterev@univ-reims.fr	OK
ECaSDa - Ethene Calculated Spectroscopic Database	Calculated data of ethylene (12C2H4). The data on ethylene contain the vibration-rotation energy levels, line positions and line intensities in the range from 500 to 7500 cm-1	ludovic.daumont@univ-reims.fr, maud.rotger@univ-reims.fr	OK
GhoSST	The GhoSST database ("Grenoble Astrophysics and Planetology Solid Spectroscopy and Thermodynamics" database service) provides laboratory data on spectra (from UV to FIR) of natural and synthetic solids (ices, molecular solids, minerals, salts, inorganic materials, organic materials, meteorites, adsorbed molecules, hydrated solids,?) of space sciences, Earth sciences and astrophysical interest. It is completed with band list data (NIR to FIR) on molecular solids and adsorbed/hydratation molecules. The GhoSST data come from laboratory experiments performed since 1989 at IPAG (and formerly at LGGE and LPG) with different spectroscopy techniques (transmission, bidirectional reflection, micro-spectroscopy, ATR, Raman, Fluorescence, ...).	damien.albert@obs.ujf-grenoble.fr	OK
Lund laboratory spectroscopy	Experimental data for transitions and lifetimes	hampus@astro.lu.se	OK

Two-stage selection

Select

XSAMS

Filter and extract



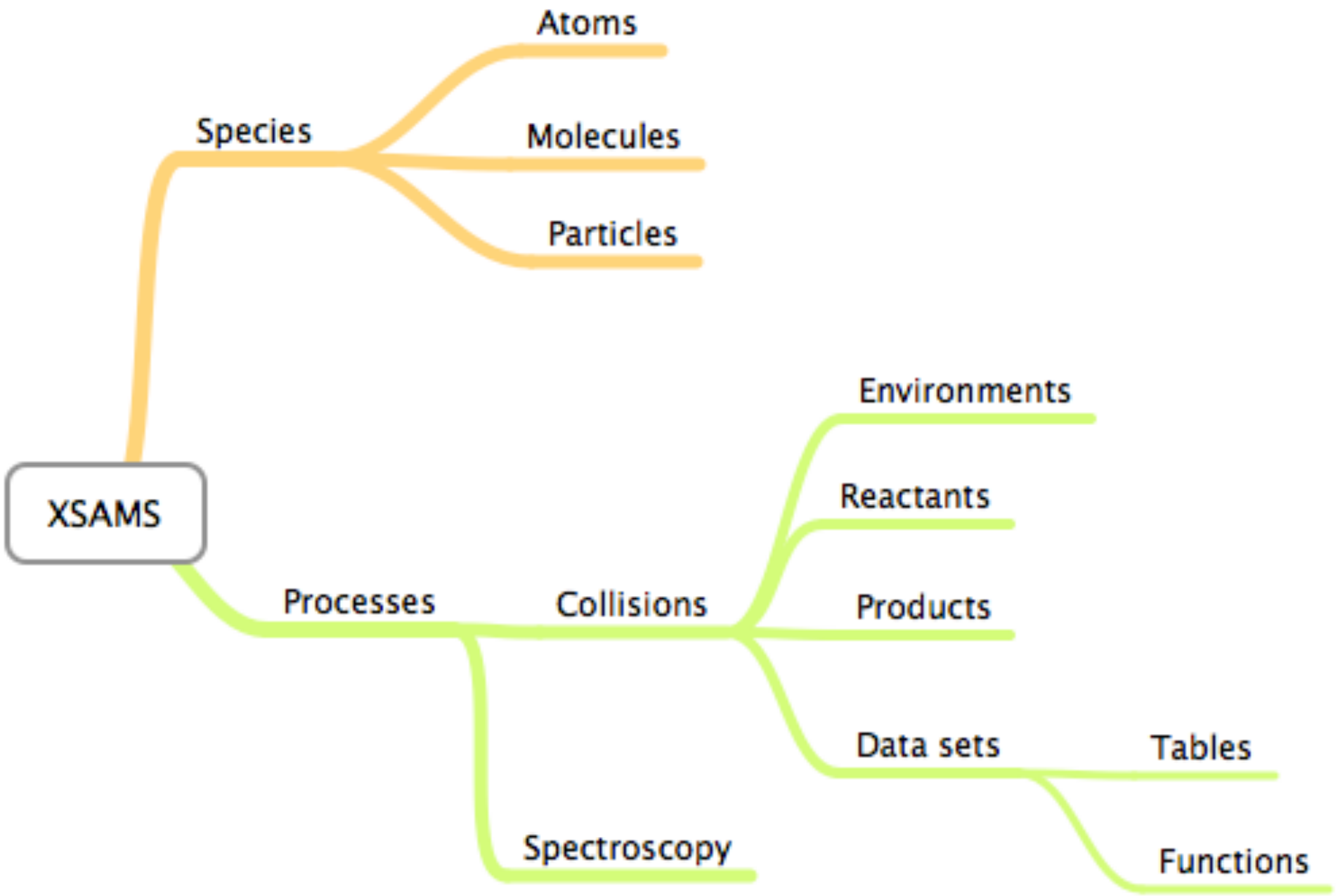
OR

Science code

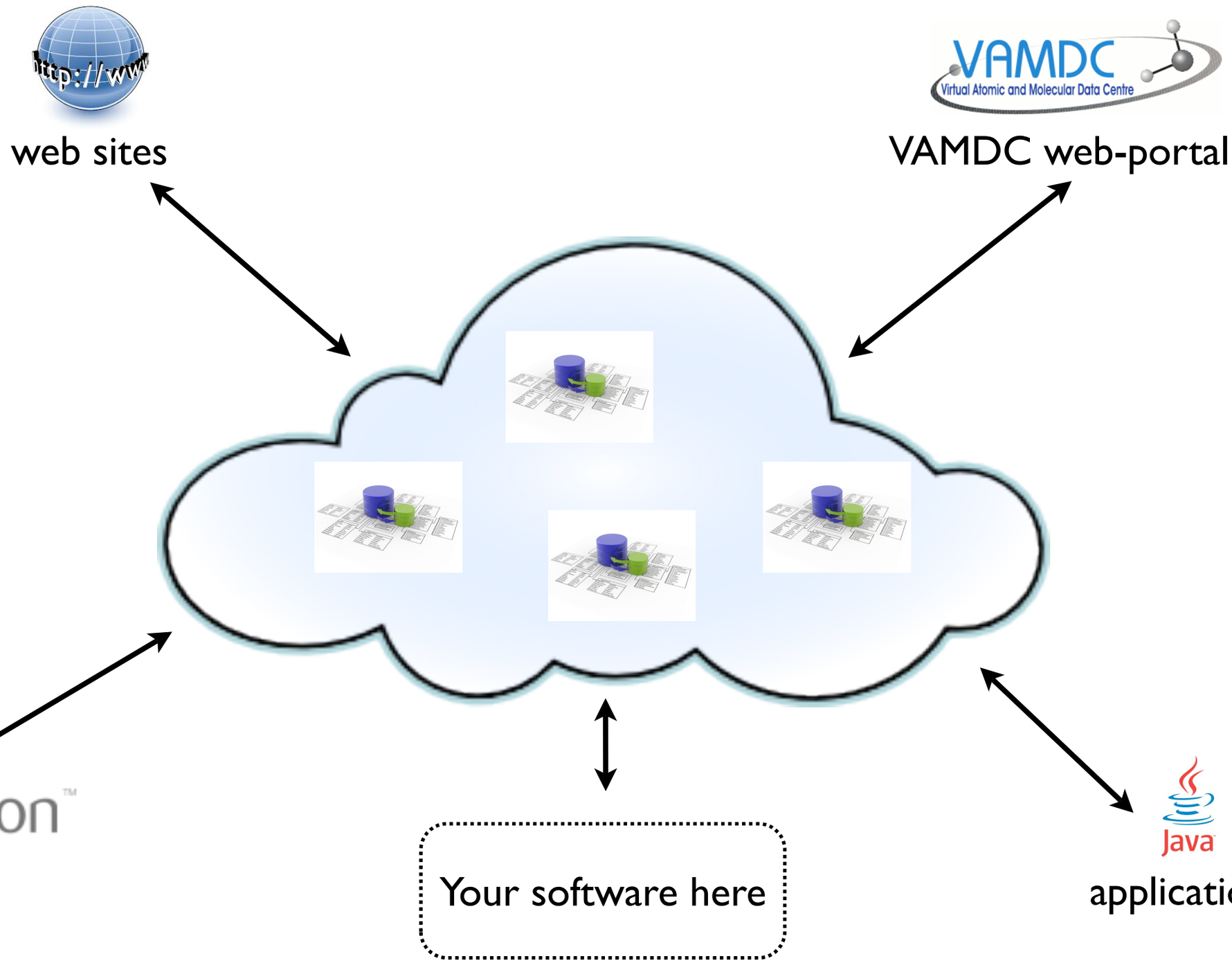
XSAMS

- XML Schema for Atoms, Molecules and Solids
- IAEA originally; developed by VAMDC
- Rich ⇒ good for transforming to other formats
- See http://www.vamdc.eu/documents/standards/dataModel/vamdcxsams/index.html

E.g. XSAMS for phys. chem.



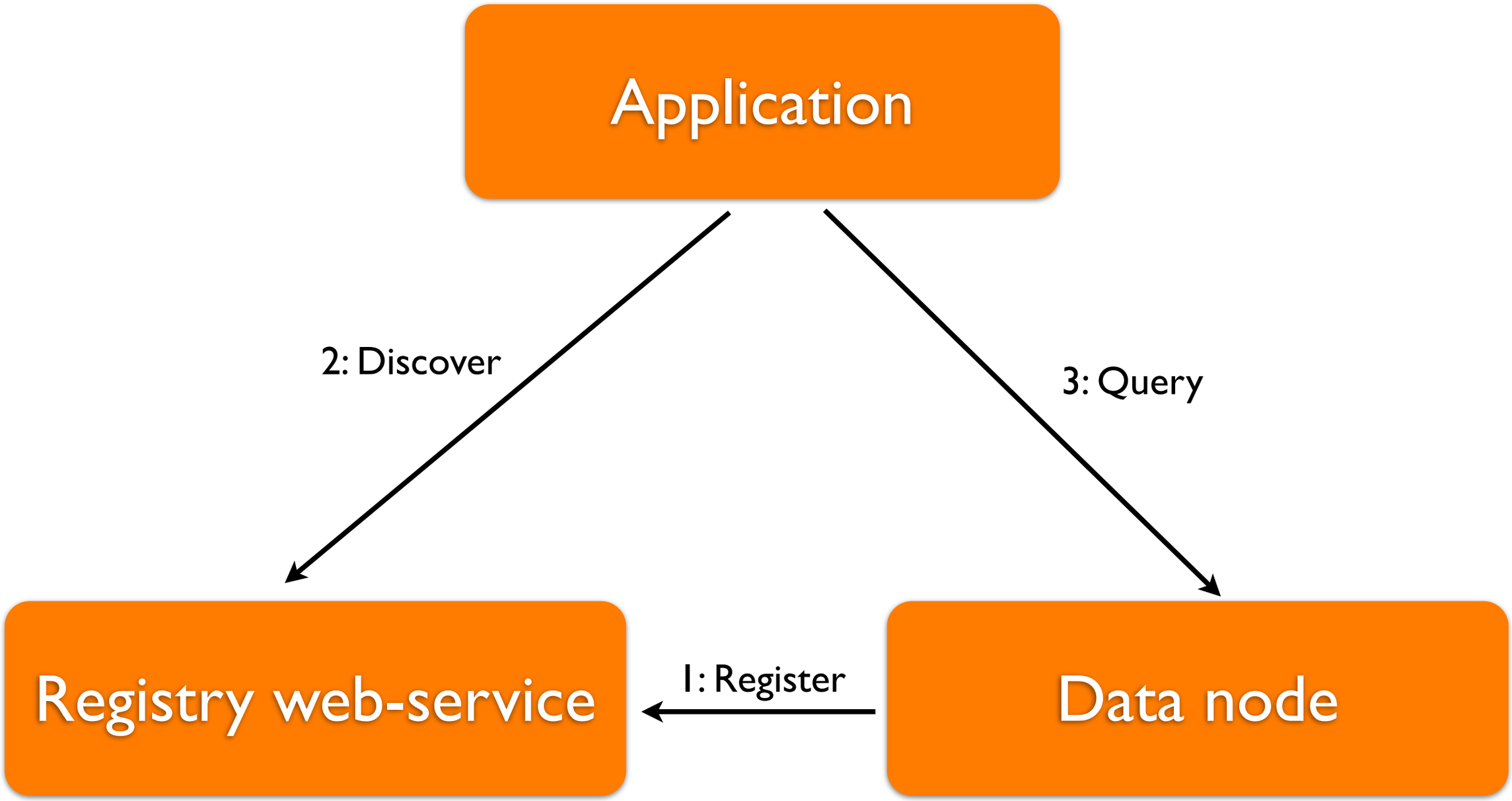
Many UIs


web sites
VAMDC
Virtual Atomic and Molecular Data Centre
VAMDC web-portal
python
scripts
Your software here
Java
applications

Some UIs and applications

- VAMDC web portal - the starting point
- SpectCol - combine spectroscopy and collisions
- Specview - STScI's spectrum viewer with VAMDC support
- Query Builder - app to generate queries for scripting
- VAMDC as IVOA PDL service - astronomy integration
- Taverna - workflow engine with VAMDC plug-in
- Selection of Python scripts from VAMDC

Finding things: registry



Avoids hard-coding addresses: data nodes may move

VAMDC web portal: query

Atom symbol H

Mass number to

Nuclear charge to

Ion charge to

InChIKey

State energy to 1/cm

Equivalent to 1/cm

Molecules (target) Clear Remove «

Chemical name Carbon Monoxide

Stoichiometric formula CO

Structural formula CO

Spin isomer

Standard InChIKey UGFAIRIUMAVXCW-RGIGPVFXSA-N, UGFAIRIUMAVXCW-ZDOIIHCHSA-N, UGFAIRIUMAVXCW-HQMMCQRPSA-N,

Select All None Search by stoichiometric formula if no inchikey is selected.

Formula	InChIKey
(12C)(16O)	UGFAIRIUMAVXCW-UHFFFAOYSA-N
(13C)(16O)	UGFAIRIUMAVXCW-OUBTZVSYSA-N
(12C)(18O)	UGFAIRIUMAVXCW-HQMMCQRPSA-N
(12C)(17O)	UGFAIRIUMAVXCW-VQEHIDDOSA-N
(13C)(18O)	UGFAIRIUMAVXCW-RGIGPVFXSA-N
(13C)(17O)	UGFAIRIUMAVXCW-ZDOIIHCHSA-N

Collisions Clear Remove «

Process name

Process description

Process code

IAEA process code

Species	Role
Atoms	Role Collider
Molecules	Role Target

Legend

available, can answer
available, don't support query
unsupported keyword

- Cologne Database for Molecular Spectroscopy: VAMDC-TAP service
- MeCaSDa - Methane Calculated Spectroscopic Database
- VALD (atoms)
- OACT - LASP Database
- BASECOL: VAMDC-TAP interface
- TOPbase : VAMDC-TAP interface
- Theoretical spectral database of polycyclic aromatic hydrocarbons
- IDEADB - Innsbruck Dissociative Electron Attachment Database
- Chianti
- TIPbase : VAMDC-TAP interface
- GSMA Reims S&MPO
- ECaSDa - Ethene Calculated Spectroscopic Database
- Carbon Dioxide Spectroscopic Databank - 296K
- GhoSST
- Carbon Dioxide Spectroscopic Databank - 1000K
- Lund laboratory spectroscopy database
- Stark-b
- Spectr-W3
- Water internet Accessible Distributed Information System
- HITRAN-UCL resource
- VALD sub-set in Moscow (obs)
- KIDA: VAMDC-TAP interface

VAMDC web-portal: results

Done

Modify query | Stop waiting | Save query

select * where (RadTransWavelength >= 5000.0 AND RadTransWavelength <= 5010.0)

Comments

XSAMS processors

- BibTeX from XSAMS
- Table views of XSAMS
- Xsams2SME

Process

Name	Response	Download	Species	States	Processes	Radiative
VALD sub-set in Moscow (obs)	OK	XSAMS	81	1157	701	701
Water internet Accessible Distributed Information System	OK	XSAMS	1	591	601	601
TOPbase : VAMDC-TAP interface	OK	XSAMS	70	619	353	353
HITRAN-UCL resource	OK	XSAMS	8	64	149	149
Chianti	OK	XSAMS	18	81	43	43
Spectr-W3	OK	XSAMS	9	26	13	13
GSMA Reims S&MPO	OK	XSAMS	0	0	5	5
Stark-b	OK	XSAMS	4	6	3	3
VALD (atoms)	TRUNCATED (9%)	XSAMS	78	10076	65459	65459
Cologne Database for Molecular Spectroscopy: VAMDC-TAP service	EMPTY		0	0	0	0
Carbon Dioxide Spectroscopic Databank - 1000K	EMPTY		0	0	0	0
Carbon Dioxide Spectroscopic Databank - 296K	EMPTY		0	0	0	0
MeCaSDa - Methane Calculated Spectroscopic Database	EMPTY		0	0	0	0
Lund laboratory spectroscopy database	EMPTY		0	0	0	0
ECaSDa - Ethene Calculated Spectroscopic Database	EMPTY		0	0	0	0

VAMDC web-portal: display

Data for single collision

$CO + {}^1H \rightarrow CO + {}^1H$

- M.-L. Dubernet, BASECOL database, , 2013
- N. Balakrishnan, M. Yan and A. Dalgarno, *Quantum-Mechanical Study of Rotational and Vibrational Transitions in CO Induced by H Atoms*, apj, **568**, 443-447, 2002

rateCoefficient

(K)	(cm3/s)
5.0	1.78E-10
10.0	1.93E-10
20.0	2.02E-10
30.0	2.09E-10
40.0	2.15E-10
50.0	2.2E-10
60.0	2.25E-10
70.0	2.28E-10
80.0	2.32E-10
90.0	2.34E-10
100.0	2.37E-10

Collisions with data sets

PBASC48t2T1c1C1

$CO + {}^1H \rightarrow CO + {}^1H$

M.-L. Dubernet 2013; N. Balakrishnan et al. 2002

rateCoefficient

Table (CSV): Rate coefficients

PBASC48t3T1c1C1

$CO + {}^1H \rightarrow CO + {}^1H$

M.-L. Dubernet 2013; N. Balakrishnan et al. 2002

rateCoefficient

Table (CSV): Rate coefficients

PBASC48t3T2c1C1

$CO + {}^1H \rightarrow CO + {}^1H$

M.-L. Dubernet 2013; N. Balakrishnan et al. 2002

rateCoefficient

Table (CSV): Rate coefficients

PBASC48t4T1c1C1

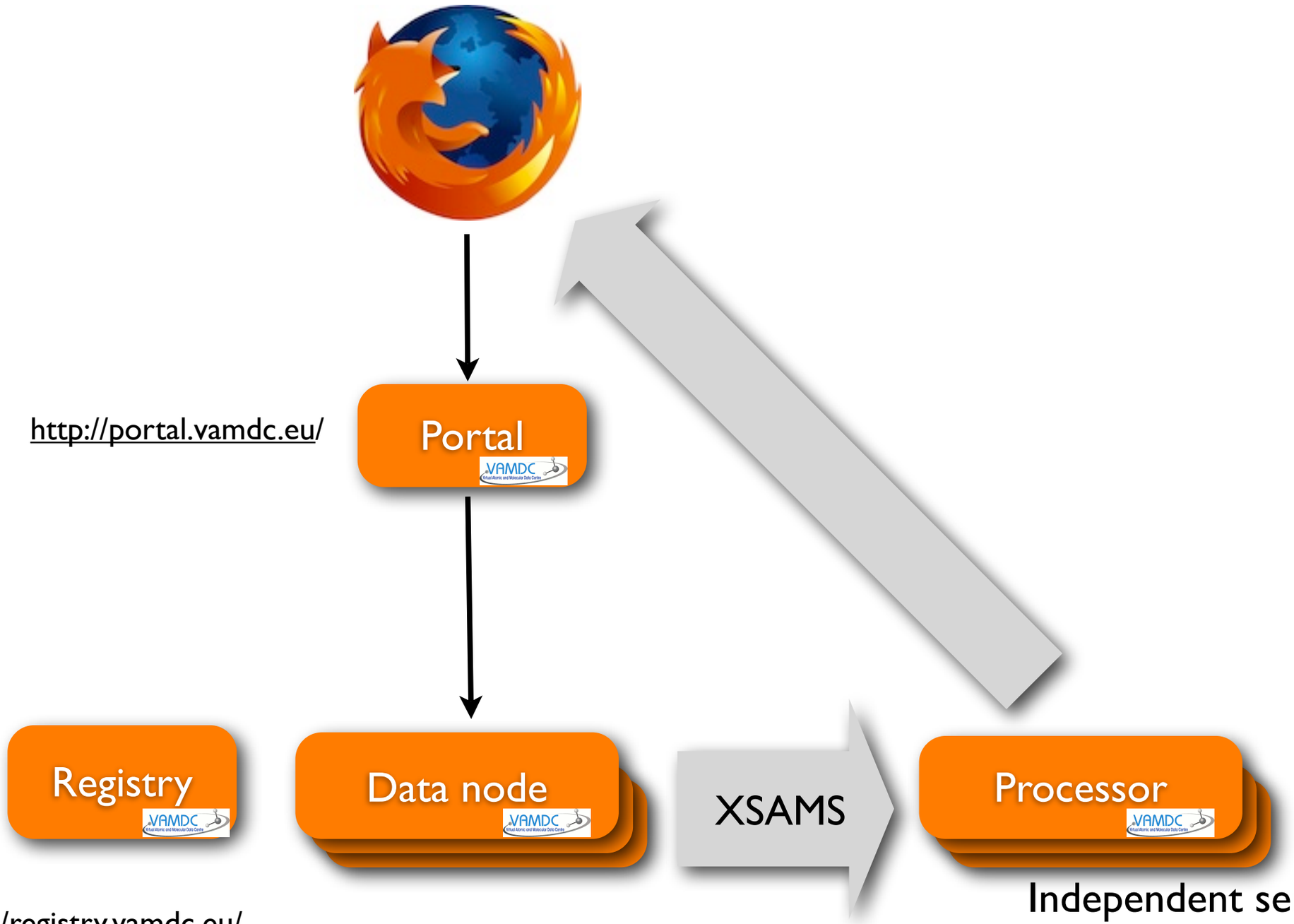
$CO + {}^1H \rightarrow CO + {}^1H$

M.-L. Dubernet 2013; N. Balakrishnan et al. 2002

rateCoefficient

Table (CSV): Rate coefficients

Portal, nodes & processors



http://portal.vamdc.eu/

http://registry.vamdc.eu/

Independent services,
can be used with any UI

SpectCol application

SPECTCOL

Help

Import data from file

Browse... File path: collisions transitions Import

Search VAMDC databases

Databases to search: BASECOL CDMS HITRAN JPL

Species search | Transitions search | Collision search

Nuclear spin: _any_

Molecular species inChiKey:

Molecular stoichiometric formula: CO

Ion charge:

Atomic symbol:

Particle name:

Submit query Cancel

Transitions

	comment	source	structural formula	stoichiometric formula	spin	InChI key
1	30502-v1:CO-18; $v=0$	CDMS 2013-09-06 10:44...	CO-18	CO		UGFAIRIUMAVXCW-HQMMCQRPSA-N
2	28512-v1*:CO; $v=1,2,3$	CDMS 2013-09-06 10:44...	CO	CO		UGFAIRIUMAVXCW-UHFFFAOYSA-N
3	31502-v1:C-13-O-18; $v=0$	CDMS 2013-09-06 10:44...	C-13-O-18	CO		UGFAIRIUMAVXCW-RGIGPVFXSA-N
4	28503-v1:CO; $v=0$	CDMS 2013-09-06 10:44...	CO	CO		UGFAIRIUMAVXCW-UHFFFAOYSA-N
5	30503-v1:C-13-O-17; $v=0$	CDMS 2013-09-06 10:44...	C-13-O-17	CO		UGFAIRIUMAVXCW-ZDOIIHCHSA-N
6	29501-v2*:C-13-O; $v=0$	CDMS 2013-09-06 10:44...	C-13-O	CO		UGFAIRIUMAVXCW-OUBTZVSYSA-N

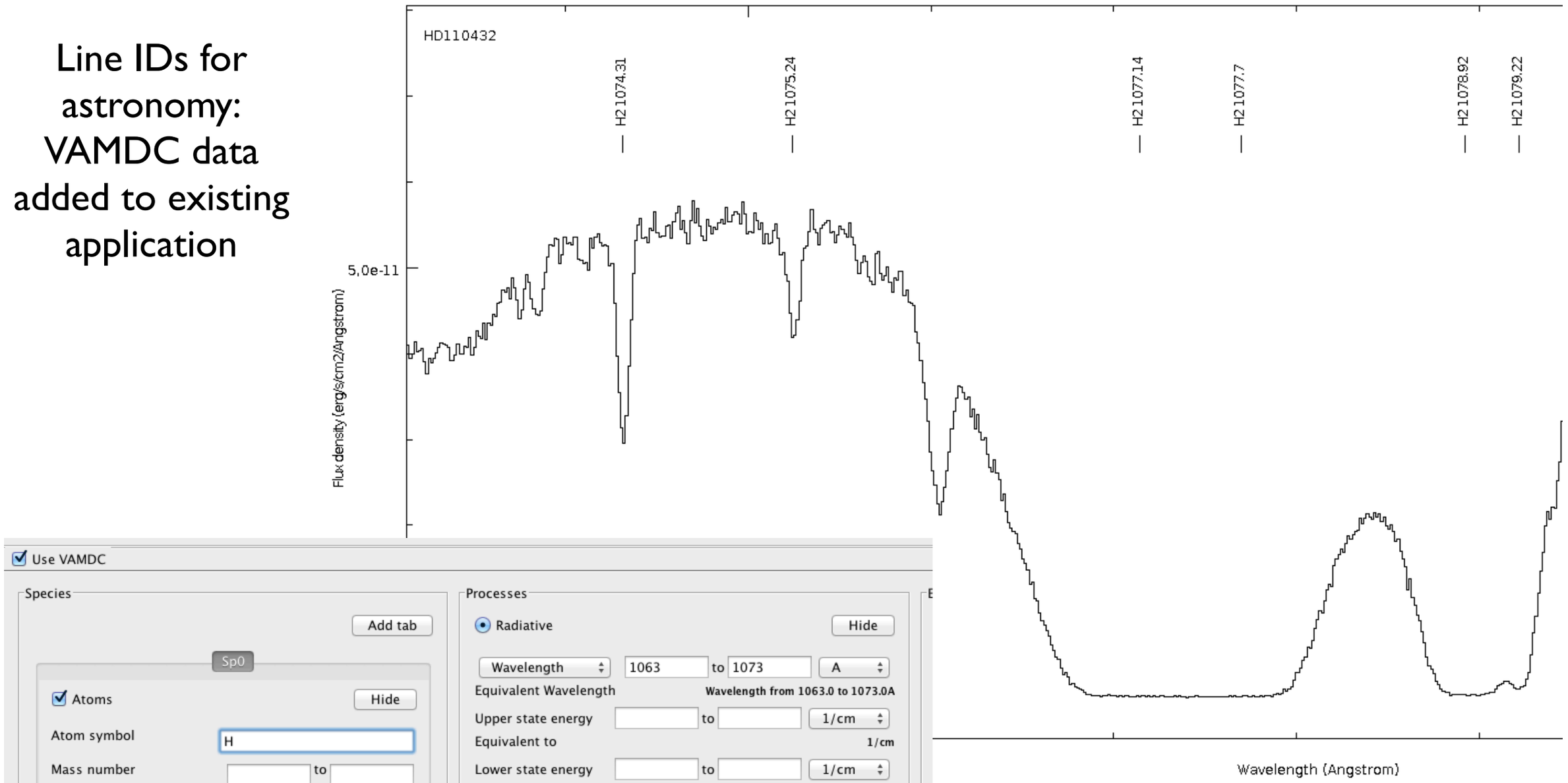
Clear | Sources | Energy table | Einstein coef. | Partition func. | Export | Group by hand | Group by species

Implements the original use case for matching spectroscopic and collisional data

See http://www.vamdc.eu/software

Specview application

Line IDs for astronomy: VAMDC data added to existing application



This query UI available as a Java library

See http://www.vamdc.eu/software

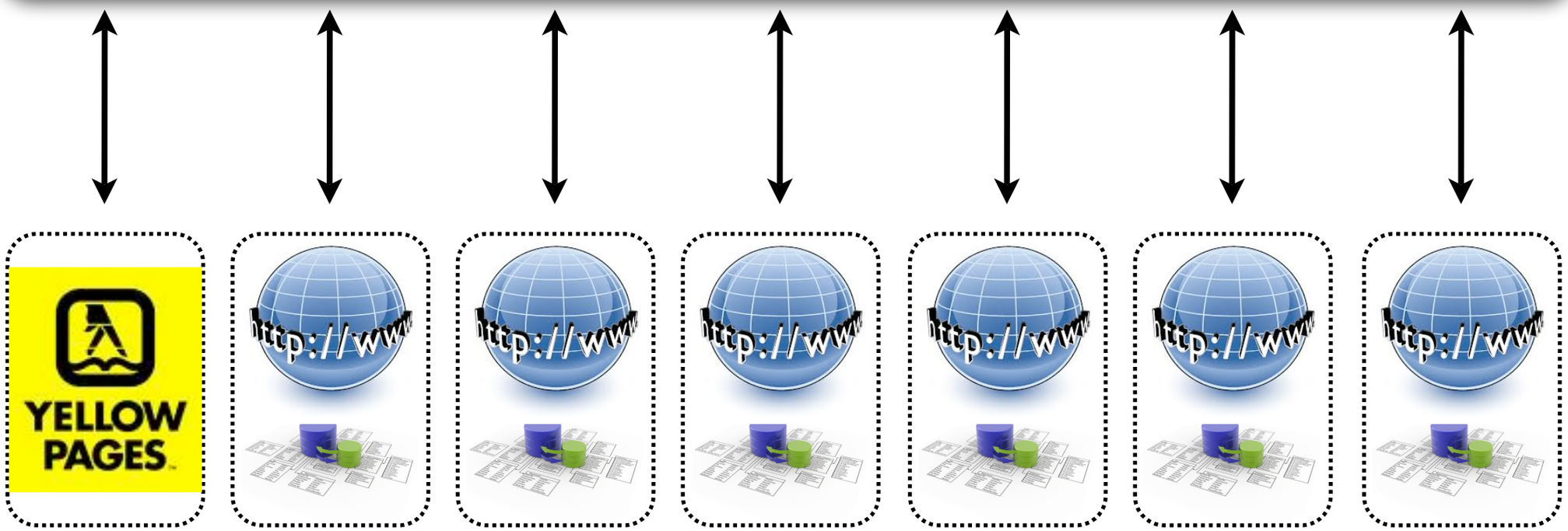
Introduction to node building

Options for data providers

- Publish your data into VAMDC by:
  - adding your data to an existing node, *or*
  - build a new node around your data and run it, *and*
    - host the node at your site *or*
    - have the node hosted at another VAMDC site

Database → data node

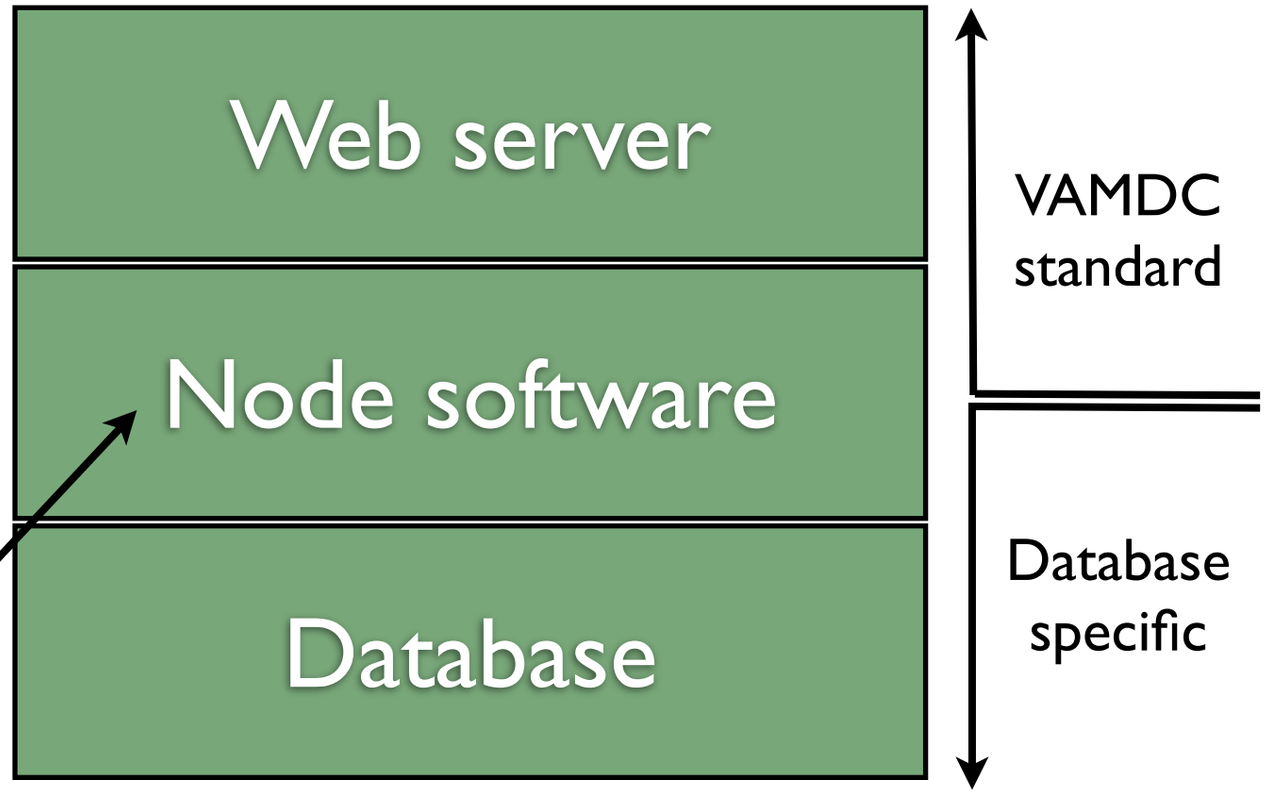
Science application (e.g. portal)



VAMDC data nodes

Parts of a data node

=



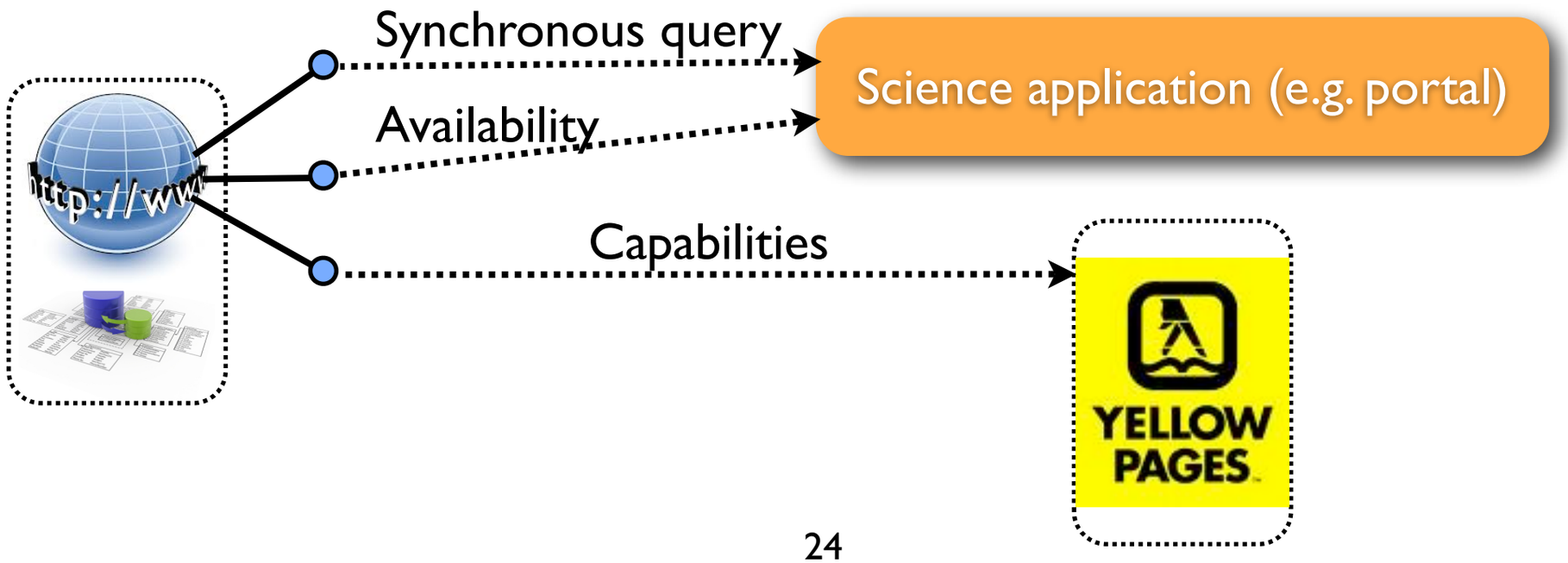
Web service,
runs in web server

Qualifying as a data node

- A web service is a VAMDC node if it:
  - implements the VAMDC-TAP protocol
  - is publicly visible on port 80
  - is registered in the VAMDC registry
  - actually emits data
- (No constraint on how you achieve that)

VAMDC-TAP

- VAMDC Table Access Protocol
- Based on IVOA Table Access Protocol
- Specifies a facade for queries to DB via web service
- Also ancillary interfaces for registration, availability checks



VAMDC-TAP: query

http://some.server/tap/sync?
LANG=VSS2&FORMAT=XSAMS&QUERY=SELECT+*+...

- Synchronous-query URL within TAP service
- HTTP HEAD → statistics, no data raised
- HTTP GET, POST → data raised, returned in HTTP response
- Query language, result format variable
  - VAMDC requires VSS2 and XSAMS
  - could add others

VSS2 query-language

- VAMDC SQL Sub-set #2
- ANSI SQL with much of the detail excluded
- E.g. SELECT * WHERE collider.AtomSymbol='He' AND target.MoleculeInchiKey=...
- Operates on a virtual, single table with columns defined by VAMDC dictionary
- See http://www.vamdc.eu/documents/standards/queryLanguage/index.html

VAMDC dictionary

- Lists, defines:
  - RESTRICTABLES: columns to constraints in query
  - RETURNABLES: columns that can be in the results
  - REQUESTABLES: columns/structures desired in results
- See http://dictionary.vamdc.eu/

VAMDC-TAP: capabilities

http://some.server/tap/capabilities

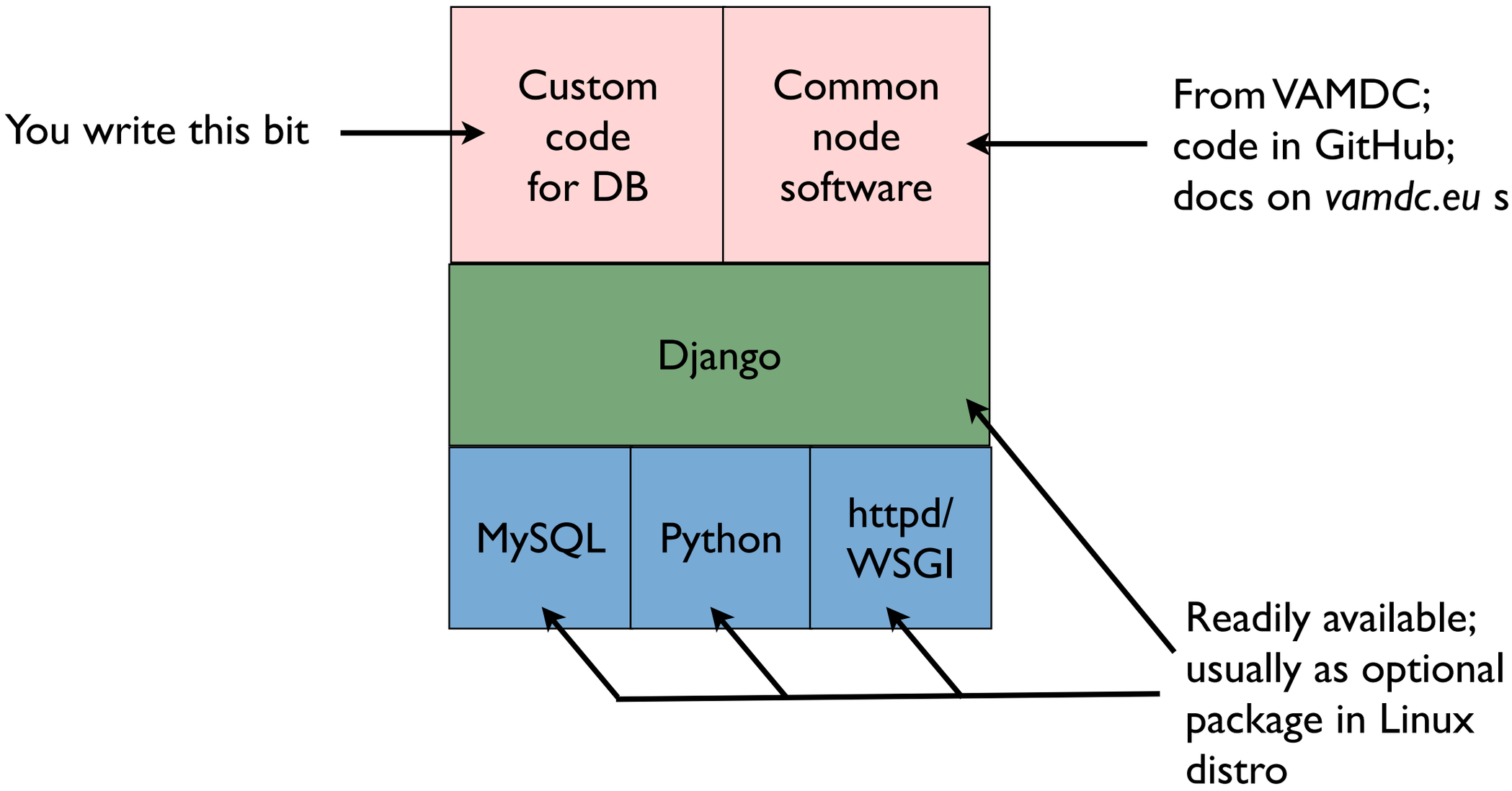
- Describes service interfaces in a form that the registry understands
- Responds to HTTP GET
- XML document
- Capability for VAMDC-TAP lists:
  - version of standards
  - version of software
  - search terms supported in query
  - sample queries
- E.g. http://ag02.ast.cam.ac.uk/chianti/tap/capabilities

VAMDC-TAP: availability

http://some.server/tap/availability

- Check that web-service is up
- XML document (XSL for browser display)
- E.g. http://ag02.ast.cam.ac.uk/chianti/tap/availability

VAMDC standard "stack" for node

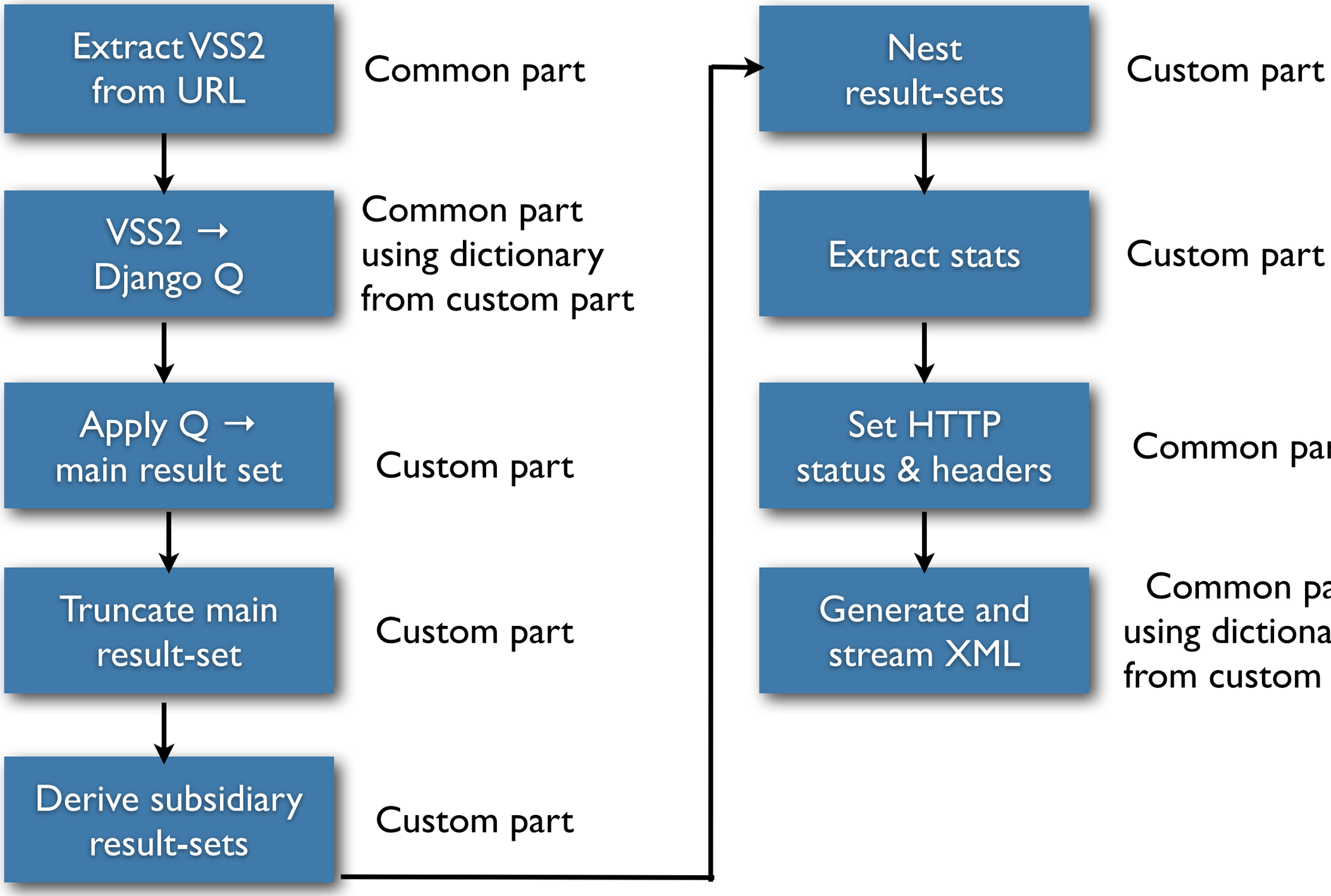


See http://www.vamdc.eu/software

What does Django do?

- Represents DB tables as "Model" objects
  - Handles joins
  - E.g. http://ag02.ast.cam.ac.uk/tutorials/_downloads/models.py
- Represents queries as "Q" objects
  - E.g. http://ag02.ast.cam.ac.uk/tutorials/_downloads/queryfunc.py
- Represents query results as "query-set objects"
  - Cursors on DB query
  - ⇒ lazy evaluation
  - E.g. http://ag02.ast.cam.ac.uk/tutorials/_downloads/queryfunc.py

Processing a query



Therefore, you write:

- *models.py*: define table structure to Django
- *dictionaries.py*: define mappings to VAMDC
  - *RESTRICTABLES*: VSS2 → Django Q
  - *RETURNABLES*: Django query-set → XSAMS
- *queryfunc.py*: implement query flow as per previous slide

Pause to digest that information...
...possibly reviewing examples:

- http://ag02.ast.cam.ac.uk/tutorials/_downloads/queryfunc.py
- http://ag02.ast.cam.ac.uk/tutorials/_downloads/models.py
- http://ag02.ast.cam.ac.uk/tutorials/_downloads/dictionaries.py

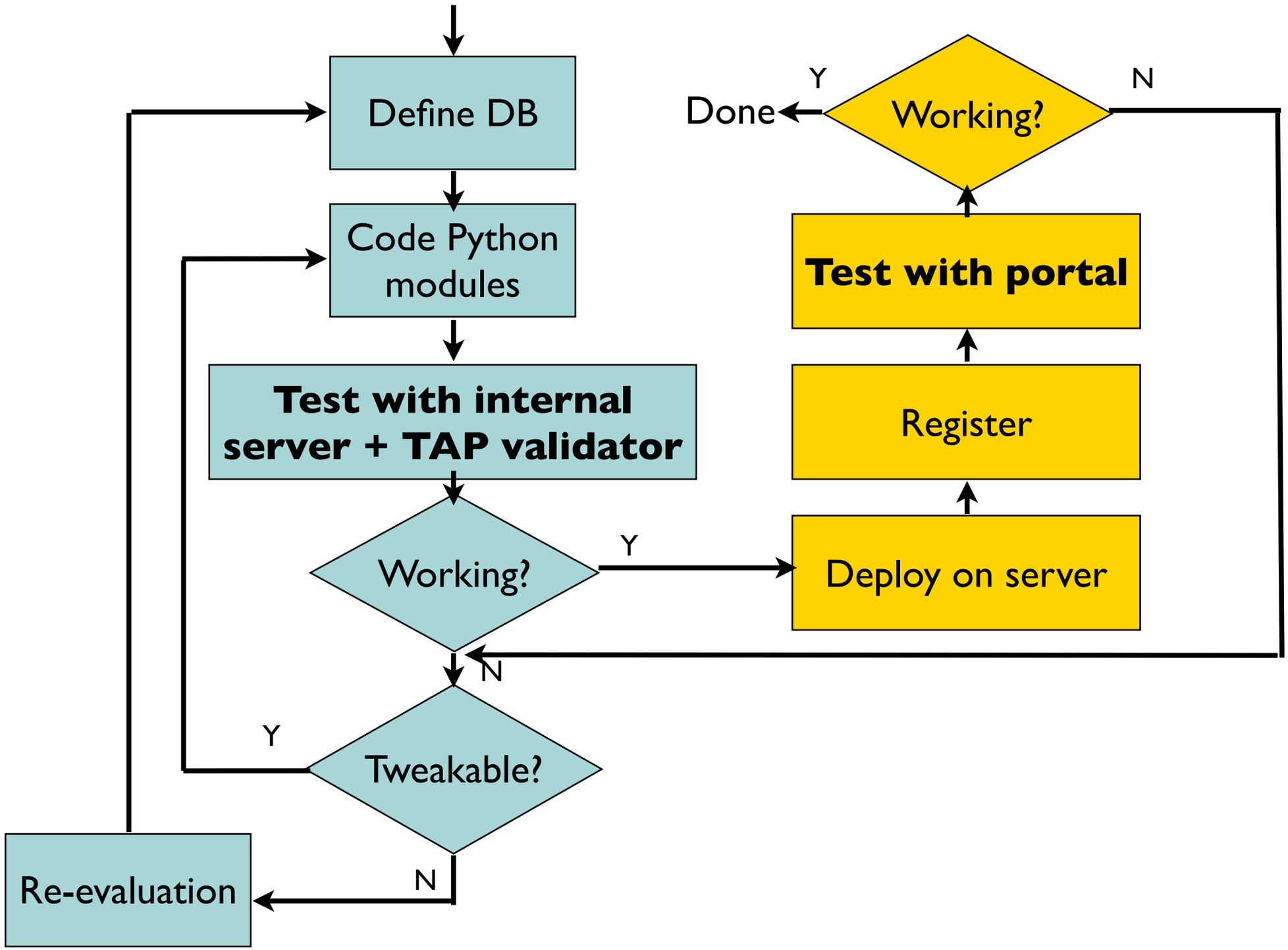
Design sequence

- Choose DB tables; define as Django models
- Design query strategy (queryfunc.py) for models
- Choose search terms (RESTRICTABLES dictionary)
- “Wire up” models to XSAMS (RETURNABLES dictionary)
- Test; iterate, refine

Database ingestion

- Node software doesn't care how you load data
- MySQL can read either SQL scripts or ASCII files
- ASCII inputs have to match chosen DB-schema
- Node software includes code to re-arrange ASCII files:
  - see *imptools* package: https://github.com/VAMDC/NodeSoftware/tree/master/imptools
  - docs at http://www.vamdc.eu/documents/nodesoftware/importing.html

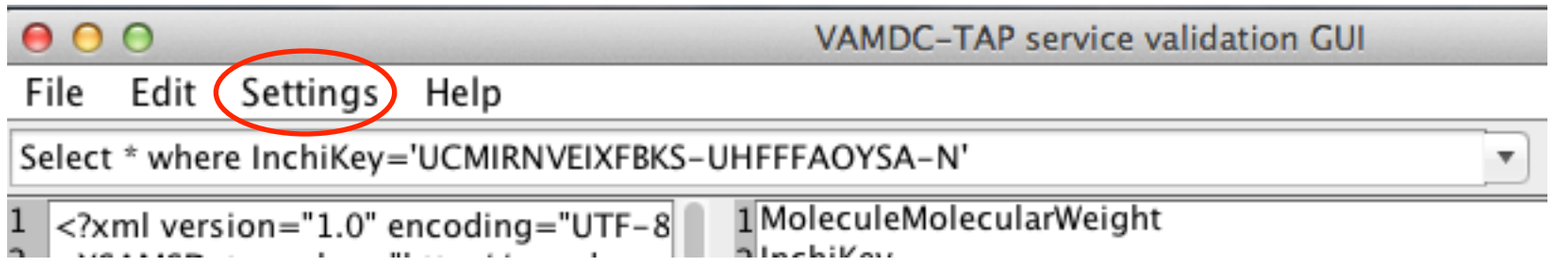
Testing sequence



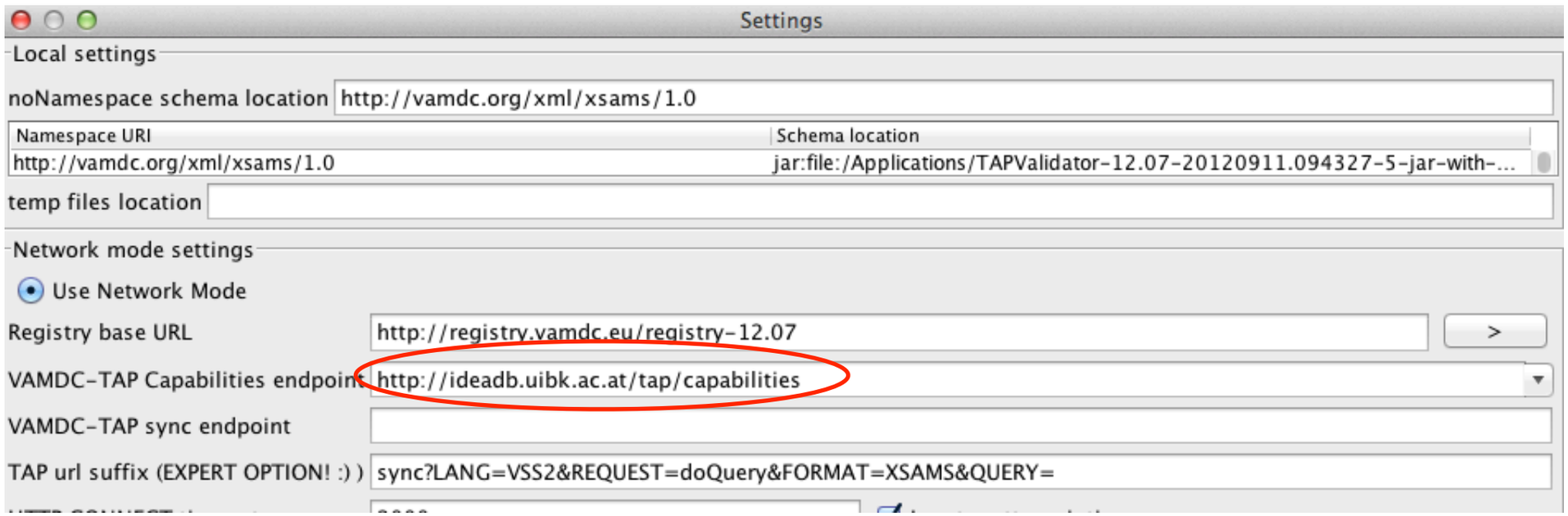
TAP validator

See http://www.vamdc.eu/software

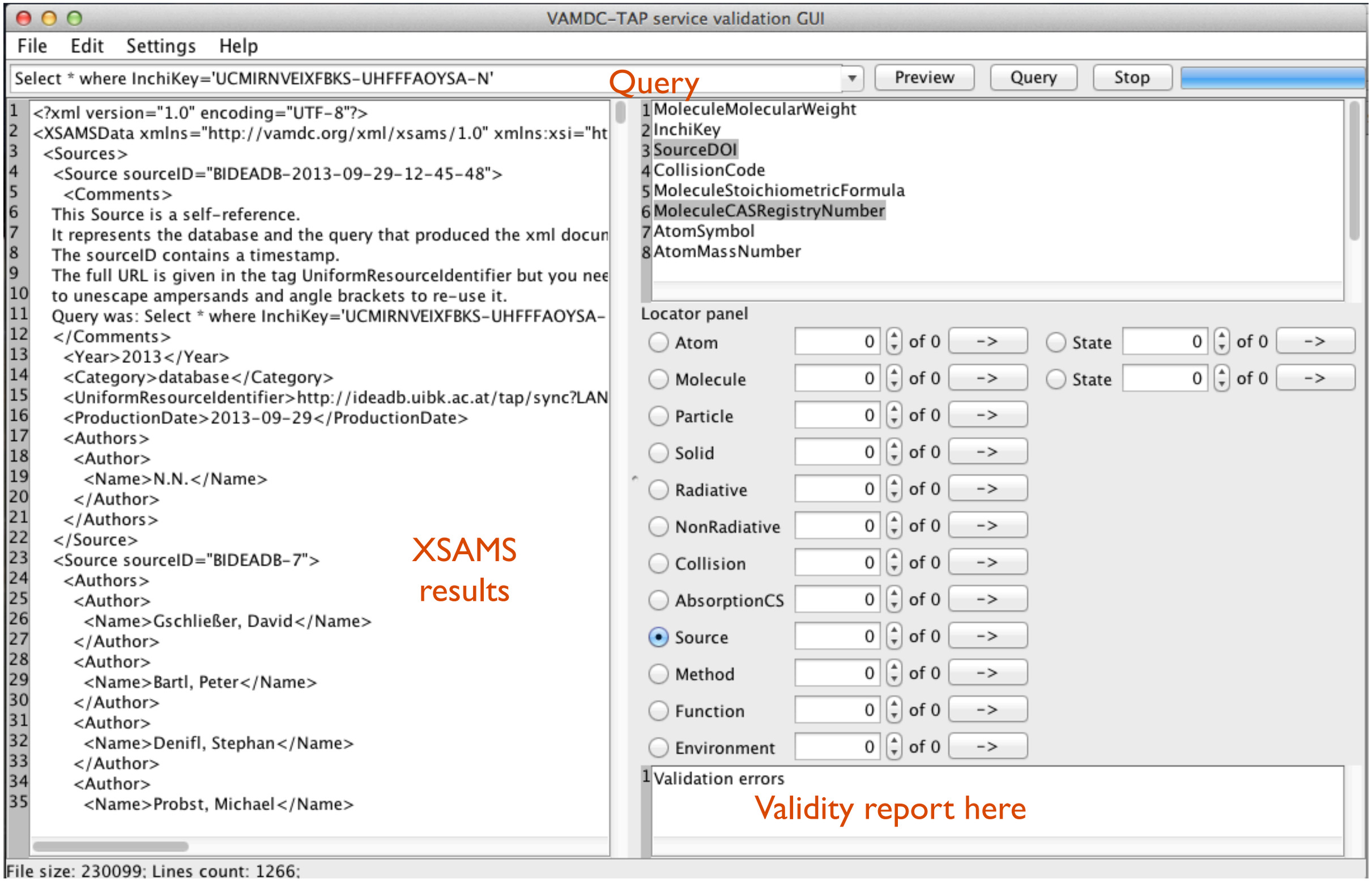
Download and run locally



Enter Capabilities URL for your node in settings page...

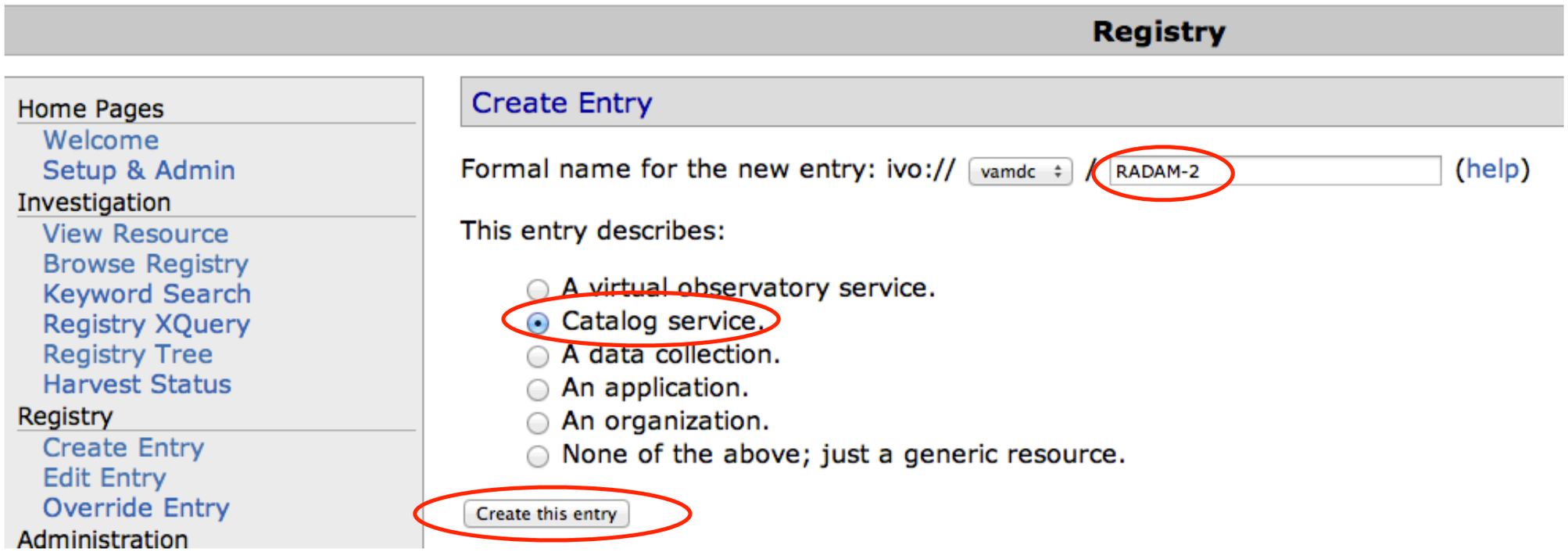


TAP validator (2)



Registration, step 1

- Go to http://registry.vamdc.eu/ and select *production** registry
- Select "create entry" from side-bar
- Fill out name of service; select "catalog service" type



*or use dev registry for practice: http://casx019-zone1.ast.cam.ac.uk/registry/

Registration, step 2

- Fill out "core information" on next form

Core metadata: editor

IVO identifier	ivo://vamdc/RADAM-example
Resource status	active
Title	Example node for RADAM network
Publisher's name	RADAM
Publisher's IVO identifier	
Creator's name	D. Provider
Creator's IVO identifier	
URL of creator's logo	
Release-date of resource	
Version of resource	
Name of contact person	J. R. Techie
Postal address of contact person	
Email address of contact person	techie@whatever.edu
Telephone number of contact person	
Keywords describing this resource	RADAM electron molecule collisions
Text describing this resource	A database of electron-molecule collisions concentrating on fast-particle effects in biological media
Source of the resource content	
URL for web page describing this resource	http://whatever.edu/radam
Type of the resource content	Other
Intended audience	Research
WebBrowser Capability URL	http://whatever.edu/radam/search

Record this information in the registry

Registration, step 3

- Select "edit" for this registry entry (use "browse registry" to search for entry if necessary)
- Select "Edit metadata ... by VOSI"
- Paste in the capabilities URL for your node and submit

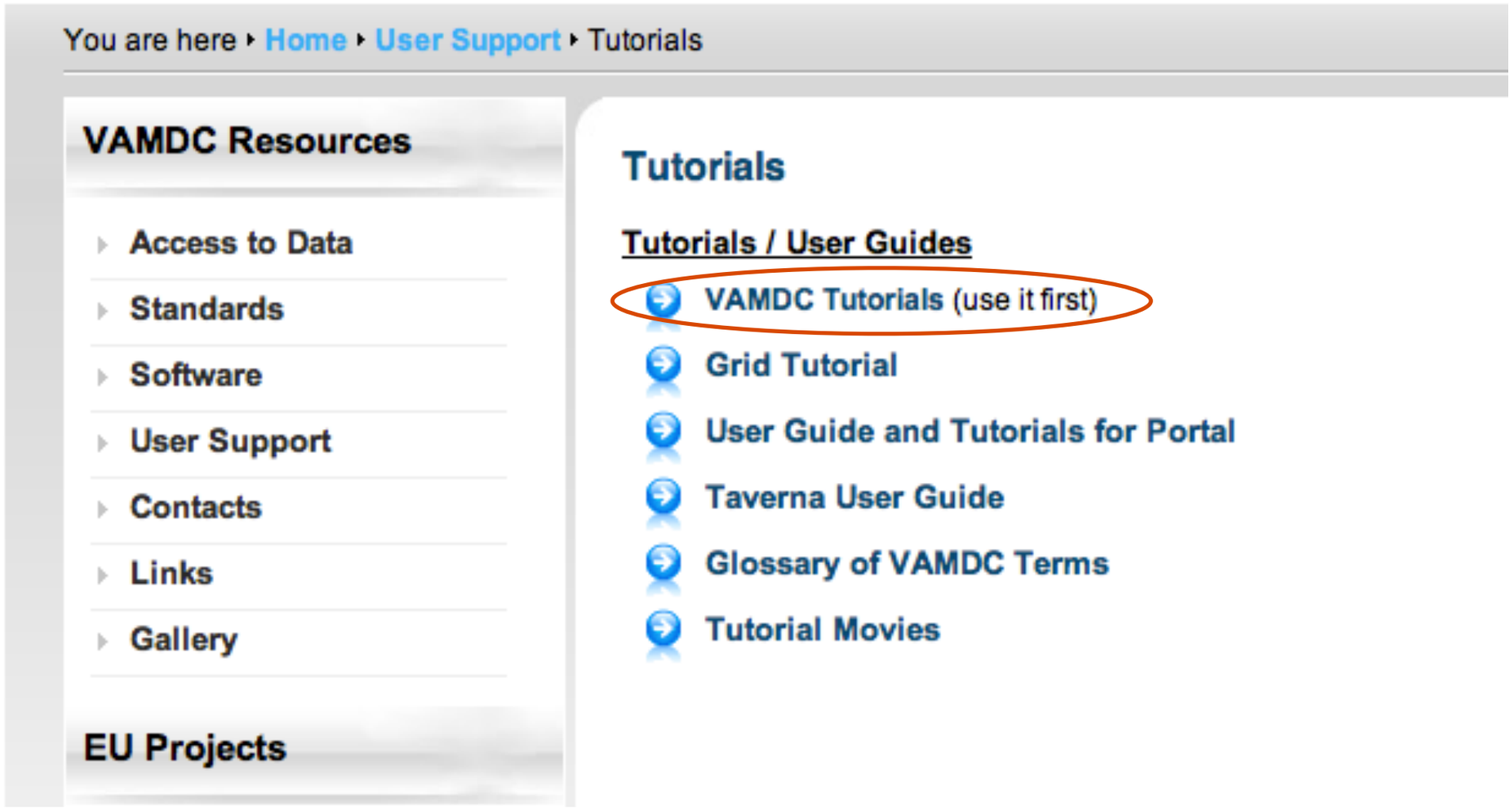
More information

- Node-software manual: http://www.vamdc.eu/documents/nodesoftware/
- VAMDC standards: http://www.vamdc.eu/standards
- Node-software video tutorials: http://ag02.ast.cam.ac.uk/tutorials/self-study/data-provider-self-study/index.html

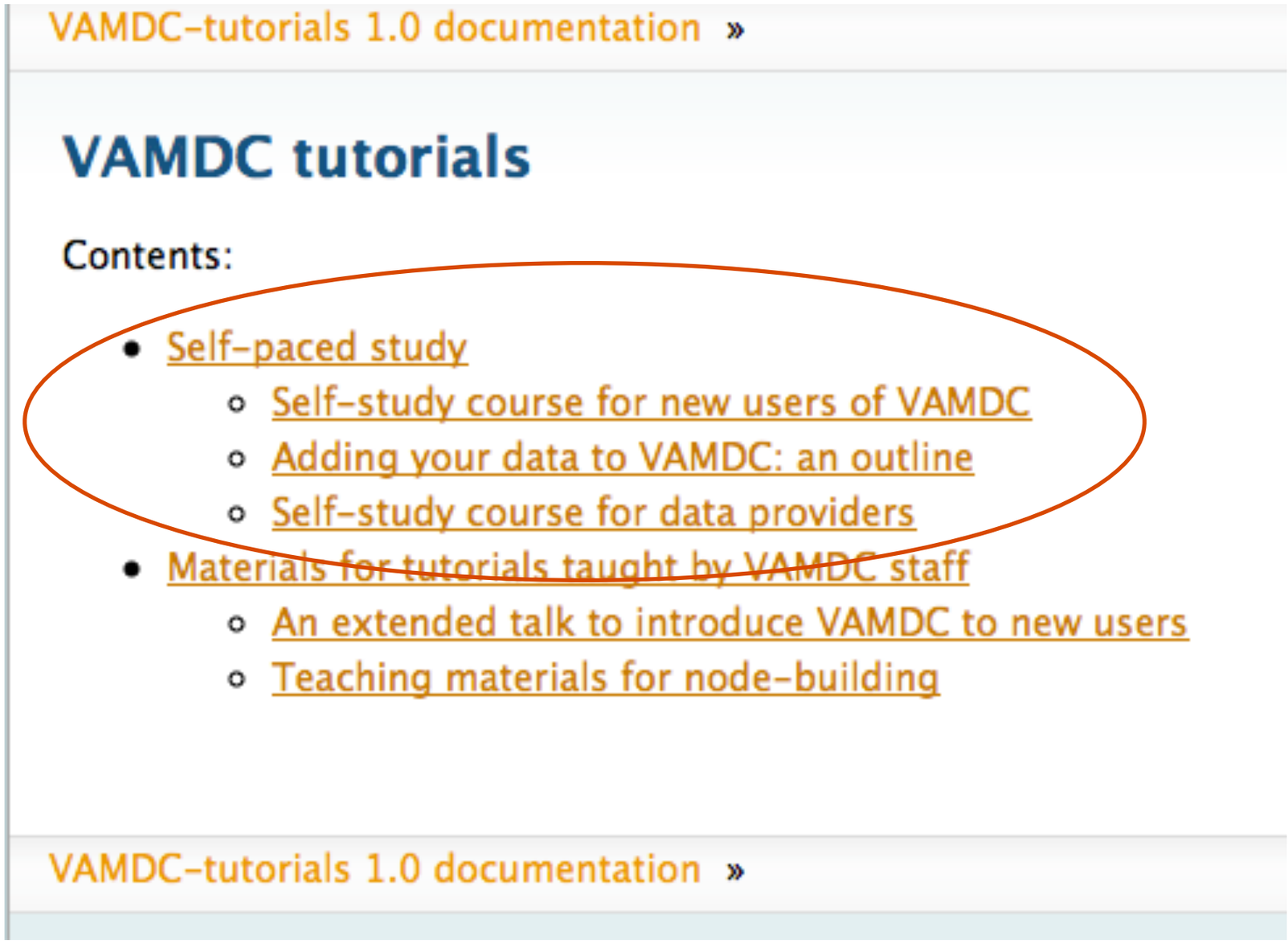
Self paced tutorials using VAMDC's on-line material

On-line tutorial suite

http://www.vamdc.eu/usersupport/tutorials



Self-paced tutorials



Examples of node building