VAMDC architecture

VAMDC implementation tutorial

VAMDC aims

Easy for consumers of VAMDC data:

- **easy to find data;**
- **easy to select data;**
- **easy to download data;**
- **easy to use data in applications**

Plan A

Dump every database into a file and put on web.

Pro:

- "The simplest thing that could possibly work"
- Everything you can get has its own URL

Con:

- Data-sets too large (up to 10GB)
- No easy way to make data extracts

Plan B

Pre-compute all possible data extracts and dump on web

Pro:

- Selection now easy
- One URL for each possible extract

Con:

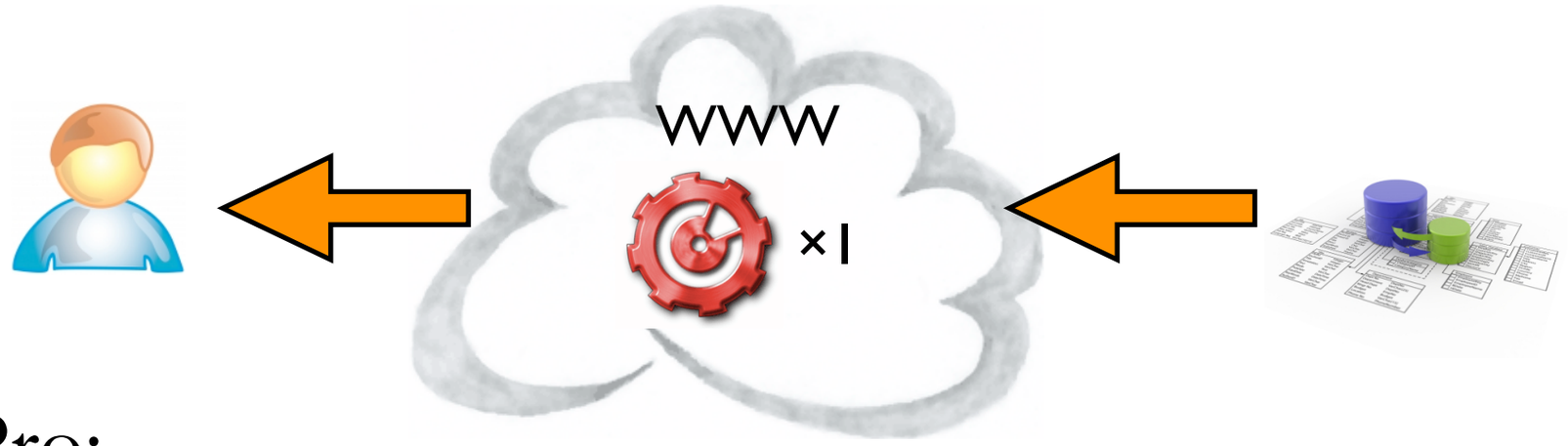
- Impossible to implement!

Plan C

Compute data extracts on demand

but

index them on the web as if pre-computed



- Pro:
  - Implementation now feasible
  - Still have a URL for every data-set
- Con:
  - Some assembly required
  - Need to define standards for services, queries etc.

VAMDC standards

- **Data model and format ⇒XSAMS**
- **Web-service protocol ⇒ TAP-XSAMS**
- **(Data-transfer protocol ⇒ HTTP)**
- **Query language ⇒ VSS1**
- **Standard terminology ⇒ VAMDC dictionary**

Data-extract URLs

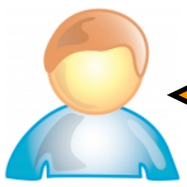
http://some.where/web/service?LANG=VSS1&QUERY=...

Identifies a database

Identifies a data extract

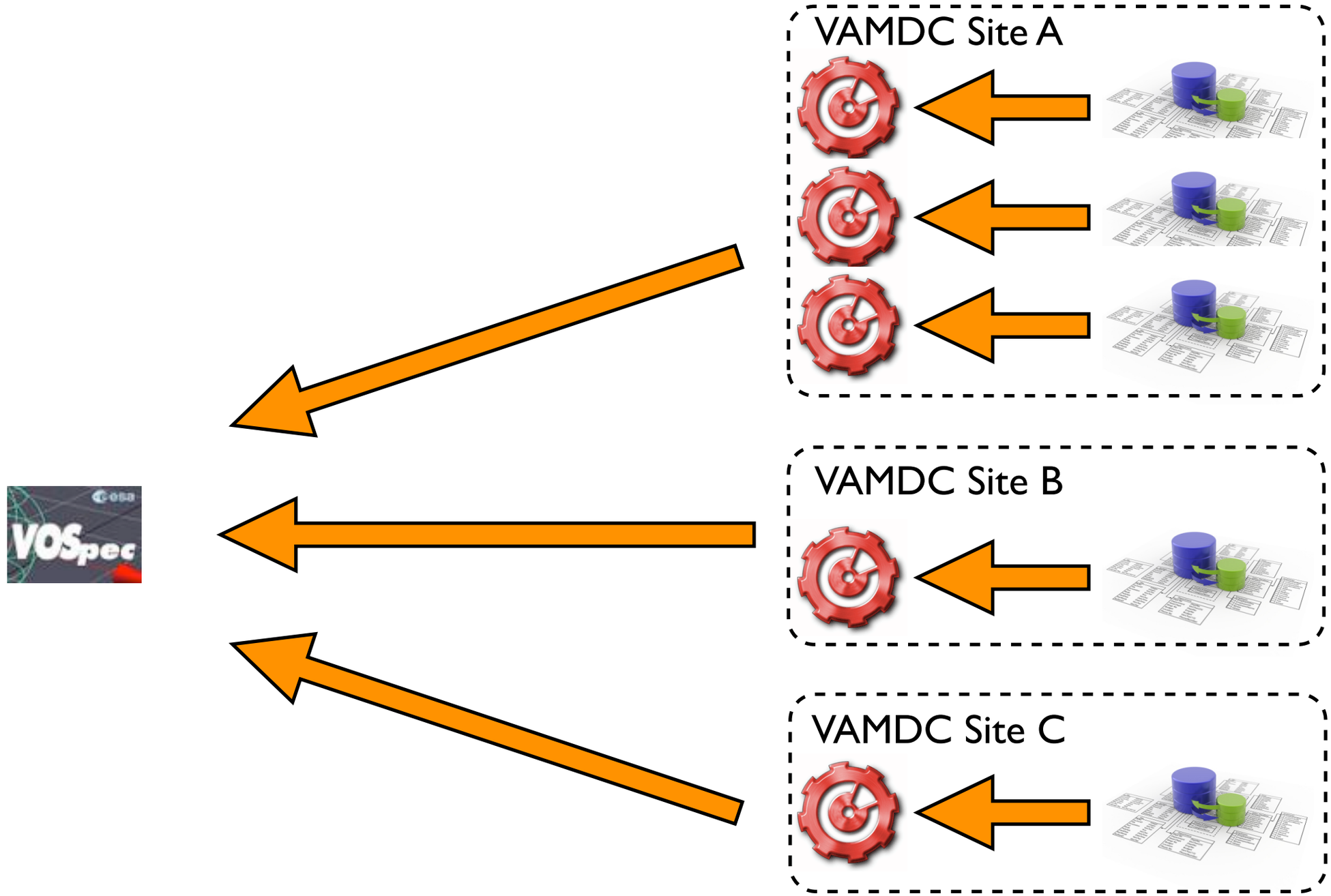
Web services & web pages

Offers
web *pages*



Offers
web *services*

Distribution of services



XSAMS

- **"XML schema for Atoms, Molecules & Solids"**
- **Proposed 2003, at IAEA DCN meeting**
- **Developed continuously by consortium (IAEA, NIST, ORNL U. Pierre & Marie Curie, OPM, RFNC-VNIITF)**
- **First release 2009**
- **See http://www-amdis.iaea.org/xsams/**

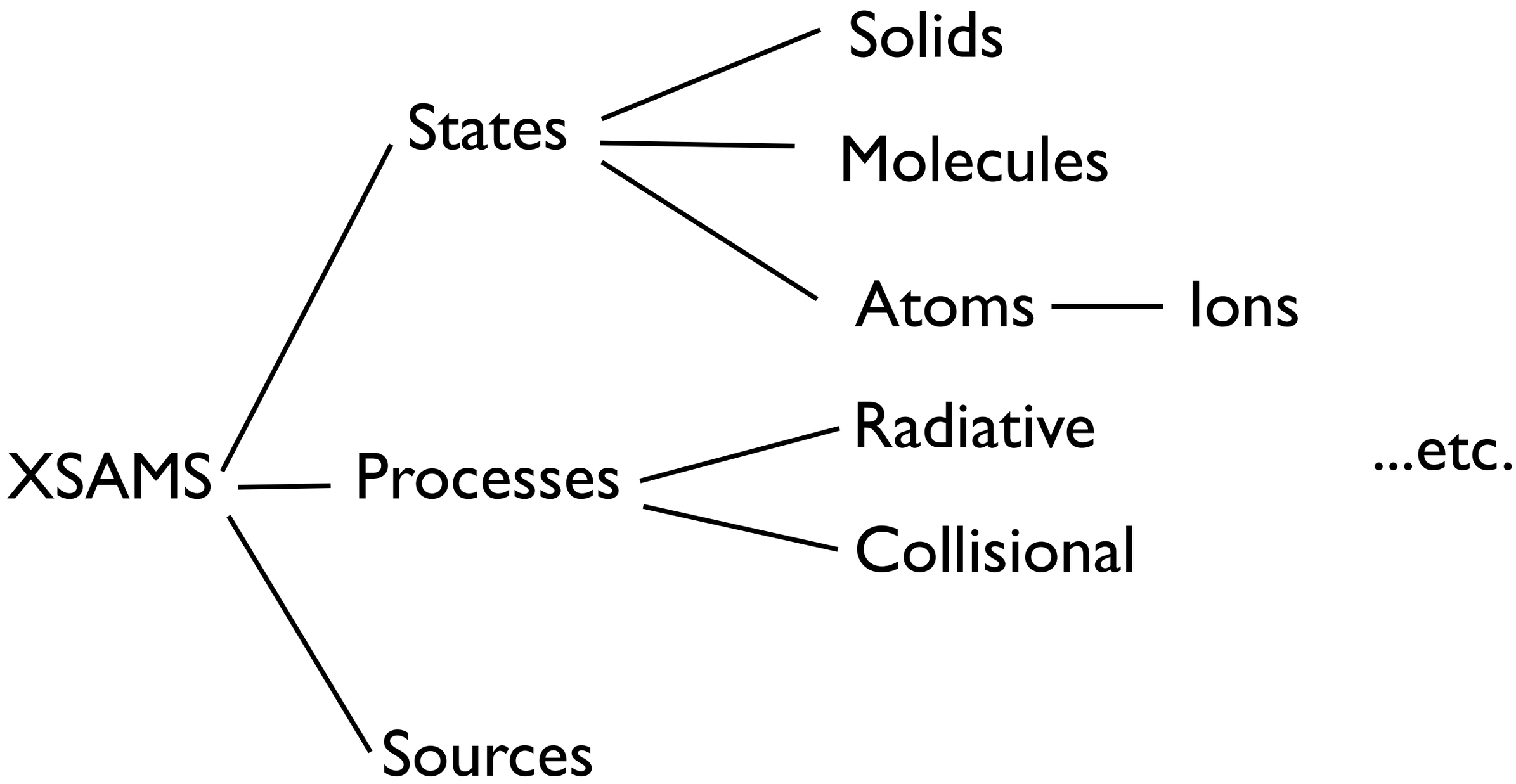
XSAMS

XML Schema for Atoms, Molecules, and Solids

Not ASCII tables?

- **Tables hard to make general and self-describing**
- **Data naturally include one-many & many-one links**
- **XML can be *easier* to process than tables**

XSAMS structure: top



XSAMS structure: bottom

- **All quantities have units**
- **All values can have associated uncertainties**
- **All values can have a source reference**
- **XML ⇒ no encoding issues for numbers**

XSAMS for molecules

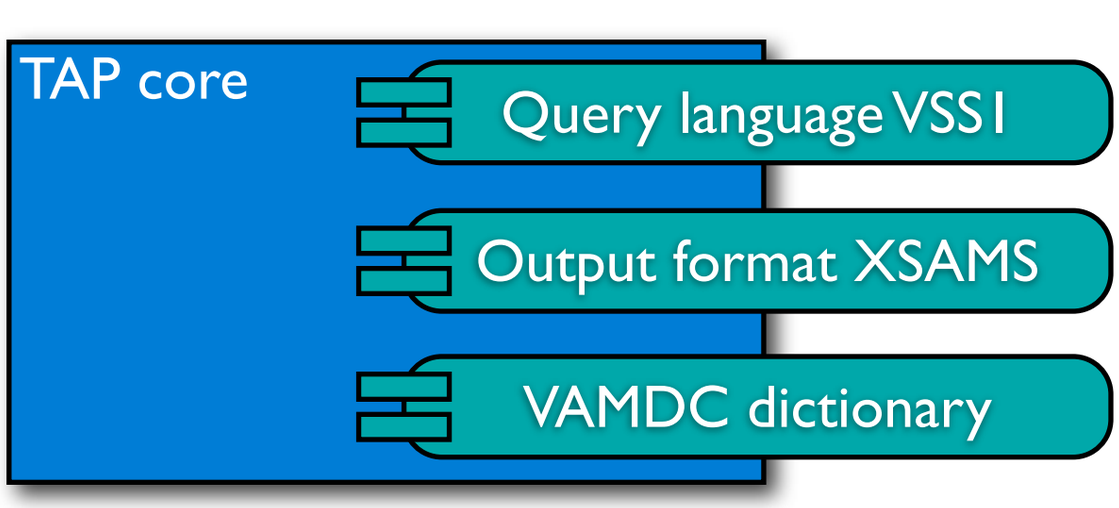
- **"Case-by-case" XSAMS:**
- **Separate, additional schema for each class of molecule:**
  - **1. Diatomic closed shell (`dcs`): CO, N2, NO+**
  - **2. Hund's case (a) diatomics (`hunda`): NO, OH [for low J]**
  - **3. Hund's case (b) diatomics (`hundb`): O2, OH [for high J]**
  - **4. Closed-shell, linear triatomic molecules (`ltcs`): CO2, HCN**
  - **...etc up to at least 12 cases**
- See http://www.ucl.ac.uk/~ucapch0/XSAMS/cbc-doc-0.2.1.pdf

XSAMS status

- **Not done yet**
- **Atoms part stable, molecules part less so**
- **XSAMS will use case-by-case**
- **VAMDC working standard February 2011**
- **“Final” standard later in 2011**
- **⇒ expect software updates as we switch versions**

TAP-XSAMS

- **Web-service protocol based on IVOA Table Access Protocol**
- **Defines:**
  - data-access URLs
  - output formats
  - query languages
  - dictionaries
- **See <u>specification doc</u>**



TAP-XSAMS parameters

Parameters to every data URL:

- LANG - selects query language
- FORMAT - selects output format
- QUERY - selects data extract
- REQUEST - cruft for IVOA compatibility

http://server/service/sync?
LANG=VSS1&
FORMAT=XSAMS&
QUERY=SELECT ALL WHERE ... &
REQUEST=doQuery

VSS1

- **VAMDC SQL Sub-set #1**
- **Sub-set of SQL92**
- **SELECT only - no modification of DB**
- **Assumes single-table view of DB where columns taken from VAMDC dictionary**
- **See specification document**

SELECT ALL WHERE AtomSymbol= 'Fe' AND IonCharge=6

VAMDC dictionary

- **List of names for data + descriptions, types, units**
- **Names taken mainly from XSAMS structure**
- **E.g. MolecularSpeciesIonCharge**
- **Like IVOA "utypes"**
- **Uses:**
  - Data labelling
  - Query language: operands in VSS1

Summary

- **VAMDC supplies data via HTTP URLs**
- **Each URL embeds a VSS1 query**
- **Each URL returns (usually) XSAMS**
- **TAP-XSAMS services respond to URLs**
- **We're here to implement those services**

Questions?

Introduction to implementation

VAMDC implementation tutorial

Three ways to implement

	Web application	Web framework	Language	Web server	Relational database
Canonical:	VAMDC	django	python	APACHE	
Alternative:	LPMAA	?	Java		

DIY: Any implementation you like, using any type of database

Tasks for this workshop

- **Get data into relational database**
- **Customize VAMDC node-software for DB**
- **Deploy node-software in web-server**
- **Register deployed service**
- **Write little script/app to use new service**

Next: details of node software by Thomas Marquart

Introduction to registry

VAMDC implementation tutorial

What is registry for?

- **Discovering data sets: list of available databases, with references to science descriptions**
- **Finding data sets: lists the parts of the data URLs identifying the services**
- **Describing services: lists the variations from common base line; e.g.**
  - RESTRICTABLES
  - RETURNABLES
  - XSAMS molecular cases used

How the registry is made

- **Web service with (XML) database of metadata**
- **One per system; in VAMDC**
  - production: http://registry.vamdc.eu/vamdc_registry
  - development: http://casx019-zone1.ast.cam.ac.uk/registry
- **AstroGrid Java web-application + eXist**

Using the registry

Read metadata

Publish metadata

- **All *reusable* applications should use registry**
- **Don't keep hard-coded lists of services in s/w**
- **To view registry interactively, use VAMDC portal**

Registry contents

- **One registration per "resource"**
- **⇒ one registration document per service**
- **Human-readable part: "Dublin core"**
- **Machine-readable part: "capabilities"**

Dublin core example

Registry entry for ivo://vamdc/cdms

IVO identifier	ivo://vamdc/cdms
Resource status	active
Title	CDMS
Publisher's name	University of Cologne
Publisher's IVO identifier	
Creator's name	University of Cologne
Creator's IVO identifier	
URL of creator's logo	
Date of publication	
Version of publication	
Name of contact person	Christian Endres
Postal address of contact person	
Email address of contact person	endres@ph1.uni-koeln.de
Telephone number of contact person	
Keywords describing this resource	
Text describing this resource	
Source of the resource content	
URL for web page describing this resource	http://www.cdms.de/
Type of the resource content	Other
Intended audience	Research

```
<title>CDMS</title>
<identifier>ivo://vamdc/cdms</identifier>
<curation>
  <publisher>University of Cologne</publisher>
  <creator>
    <name>University of Cologne</name>
  </creator>
  <contact>
    <name>Christian Endres</name>
    <email>endres@ph1.uni-koeln.de</email>
  </contact>
</curation>
<content>
  <subject></subject>
  <description></description>
  <referenceURL>http://www.cdms.de/</referenceURL>
  <type>Other</type>
  <contentLevel>Research</contentLevel>
  <relationship>
    <relationshipType>service-for</relationshipType>
    <relatedResource>ivo://vamdc/cdms/ceaApplication</relatedResource>
  </relationship>
</content>
```

IVORNs

e.g.: ivo://vamdc/chianti/django

- **IVORN = IVO Resource Name**
- **Formal name for registered resource**
- **Publishing authority is alway vamdc for us.**
- **You choose the path part (“resource key”)**
- **Must be unique in whole registry**

Capability example

```
<capability standardID="ivo://vamdc/std/TAP-XSAMS" xsi:type="tx:TapXsams">
  <interface xsi:type="vs:ParamHTTP">
    <accessURL use="base">http://vamdc.fysast.uu.se:8888/node/vald/tap/</accessURL>
  </interface>
  <returnable>AtomStateLandeFactorRef</returnable>
  <returnable>SourcePageBegin</returnable>
  <returnable>SourcePageEnd</returnable>
  <returnable>AtomIonizationEnergy</returnable>
  <returnable>AtomNuclearCharge</returnable>
  <returnable>AtomSymbol</returnable>
  <returnable>AtomConfigurationLabel</returnable>
  <returnable>SourceURI</returnable>
  <returnable>SourceID</returnable>
  <returnable>AtomStateEnergy</returnable>
  <returnable>SourceAuthorName</returnable>
  <returnable>RadTransBroadWaalsAlpha</returnable>
  <returnable>RadTransWavelengthExperimentalUnits</returnable>
  <returnable>RadTransBroadWaalsRef</returnable>
  <returnable>RadTransProbabilityLog10WeightedOscillatorStrengthAccuracy</returnable>
  ...
  <restrictable>AtomStateEnergy</restrictable>
  <restrictable>AtomNuclearCharge</restrictable>
  <restrictable>RadTransLogGF</restrictable>
  <restrictable>AtomSymbol</restrictable>
  <restrictable>RadTransWavelengthExperimentalValue</restrictable>
  <restrictable>AtomIonCharge</restrictable>
</capability>
```

Registration process

1. **Get dev-registry password from Guy Rixon**
2. **Install TAP-XSAMS on a web-visible server**
3. **Go to registry web UI, type in dublin core**
4. **Ask registry to load VOSI data from your service**
   - VOSI URL: http://your.server/tap/capabilities

- **Repeat steps 2..5 as necessary.**

Questions?

Registry demonstration with CHIANTI follows