Semantic tags for species and lines identification

N.Moreau, C-M. Zwolf and
VAMDC Consortium
Paris Observatory, LERMA

VAMDC infrastructure

- e-infrastructure to access A&M data
- Databases connected through a middleware
- Standards :
  - for querying databases (VAMDC-TAP)
  - for exchanging data (XSAMS xml schema)
  - to describe services (VOResource extension)

The species database

- http://species.vamdc.eu
- Central repository of species
- Quick discovery of databases content
- Provides new features to other elements of infrastructure :
  - species name suggestion
  - dynamically test availability of a species in a DB
- Queryable with a JSON API

Main problems

- Chemical species naming is not homogeneous
- For atoms : we use symbol (+ and ion charge)
- For molecules, it can be ambiguous
- VAMDC uses InChi standard (International Chemical Identifier)

InChi / InChiKey

- Species description in InChi format :

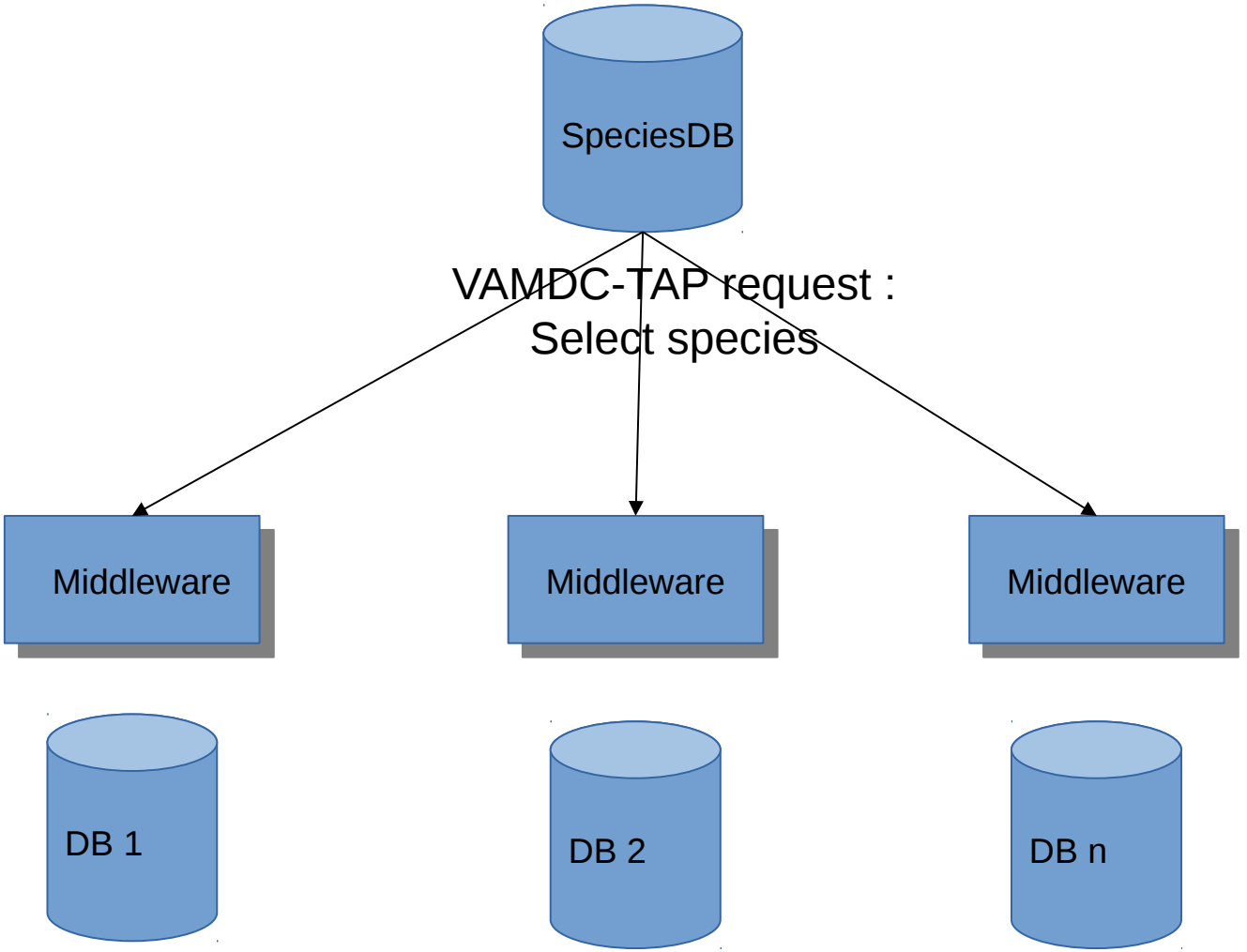
  Ex : 1S/H, 1S/He, 1S/C/q+1

- Key is a 27 characters hash of InChi (SHA-256) :

  GKDCRJWYAGBLFY-UHFFFAOYSA-N

- This identifier is used to bind the different "versions" of a species

Filling the database



Result in web interface

BASECOL: VAMDC-TAP interface (Atomic states, Atoms, Collisions, Molecular states, Molecules)

Name	Stoichiometric formula	Formula	InChI	Mass number	InChIKey	Charge
C3	C3	C$_3$	InChI=1S/C3/c 1-3-2	36	NVLRFXKSQ QPKAD-UHFFFAOYS A-N	0

CDMS (Atomic states, Atoms, Molecular states, Molecules, Radiative transitions)

Name	Stoichiometric formula	Formula	InChI	Mass number	InChIKey	Charge
Propadienediyl idene, tricarbon	C3	C3	InChI=1S/C3/c 1-3-2	36	NVLRFXKSQ QPKAD-UHFFFAOYS A-N	0

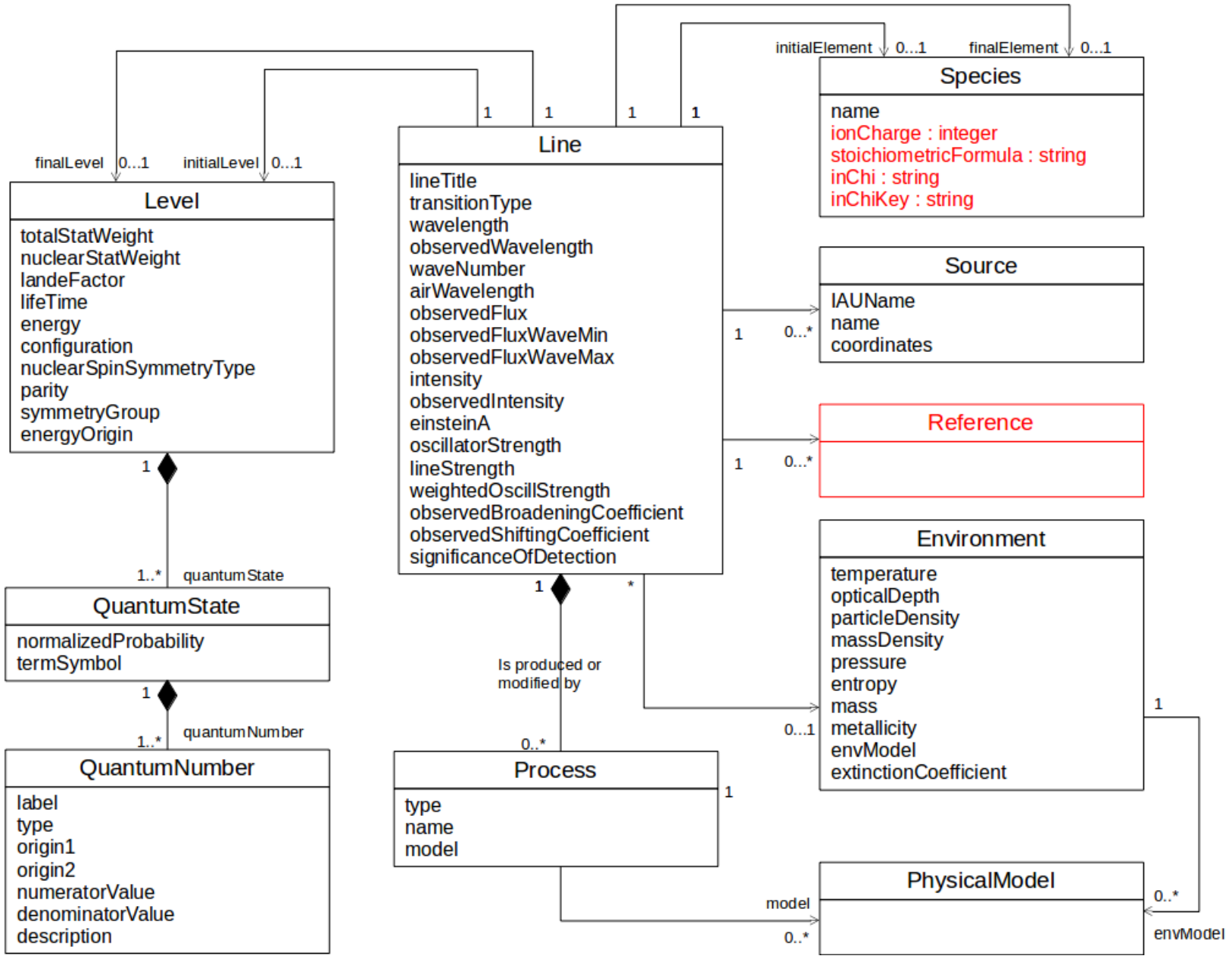
LXcat (Atoms, Collisions, Molecules)

Name	Stoichiometric formula	Formula	InChI	Mass number	InChIKey	Charge
C3 radical	C3	C3	InChI=1S/C3/c 1-3-2	36	NVLRFXKSQ QPKAD-UHFFFAOYS A-N	0

UMIST Database for Astrochemistry (Atoms, Collisions, Molecules)

Name	Stoichiometric formula	Formula	InChI	Mass number	InChIKey	Charge
None	C3	C3	InChI=1S/C3/c 1-3-2	36	NVLRFXKSQ QPKAD-UHFFFAOYS A-N	0

SSLDM proposal



Databases description

- Dbs are available in a standardized way

- Each one describes its own capabilities
  - returned data
  - possible request parameters

- They can be listed with a getCapabilities request

  Ex : CDMS

- They are chosen from terms in a dictionary

Databases description

- Descriptions are processed when data are included in the DB for the 1st time
- Tags are used to describe content
- Users know immediately type of data they can expect

name	category	prefix
Atoms	Species	Atom
Atomic states	Species	AtomState
Molecules	Species	Molecule
Molecular states	Species	MoleculeState
Collisions	Process	Collision
Cross sections	Process	CrossSection
Radiative transitions	Process	RadTrans
Radiative transitions shifting	Process	RadTransShifting
Radiative transitions broadening	Process	RadTransBroadening
Non radiative transitions	Process	NonRadTran

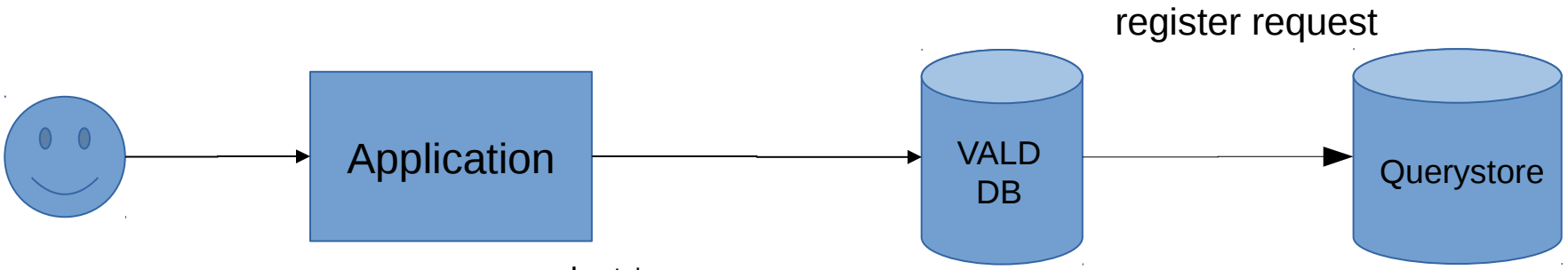
Line identification

- Well known need
- No mechanism in VAMDC
- Requests are generally done on :
  - Interval (wavelength, frequency)
  - Species name and charge
- Finding one specific line can be tedious
- Transitions have an id but their persistence is not guaranteed

Line identification

- A solution : using the query store
- It contains already executed requests
- Require a request description mechanism
- For well-known transitions, ucd can be used :
  - em.line.Halpha
  - em.line.Hbeta
  - em.line.Hgamma
  - …
- Additional list of keywords to be defined

Line identification



Looking for Hα

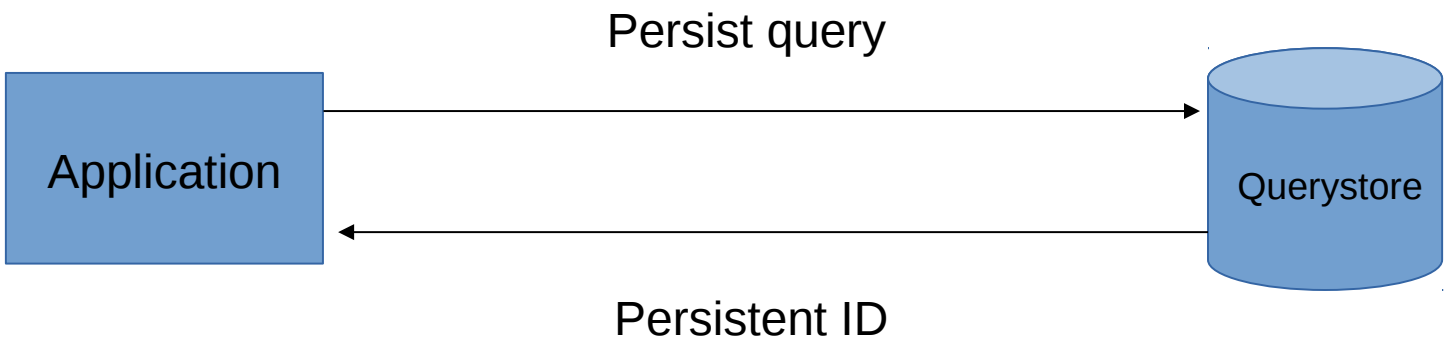
select *
where RadTransWavelength >= 6560
AND RadTransWavelength <= 6570.0
AND AtomSymbol = 'h'

Line identification

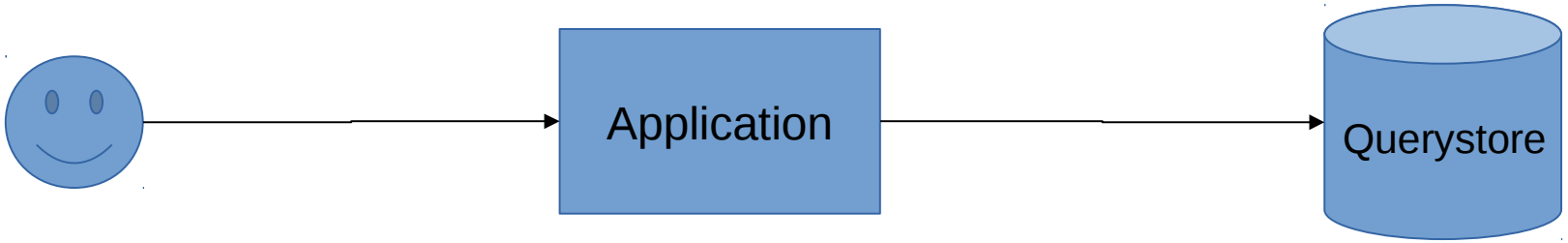
Results from vald VAMDC node

Unselect all	Spec Ion X	Wavelength (A) X	Wavelength reference X	Log10 Weighted Oscillator Strength X	Lower state description X	Lower energy(1/cm) X	Lower parity X	Lower total angular momentum X	Upper state description X	Upper energy(1/cm) X	Upper parity X	Upper total angular momentum X
	H 1	6564.60997919	Bvald-CDROM18	0.710	n=2	82259.1050	even	1.5	n=3	97492.3020	even	2.5

Line identification



Line identification



This ID is Hα :
em.line.Halpha

Persistent ID
=
em.line.Halpha

Line identification

- Users will have access to a web search interface
- Queries tagged with the searched ucd will be displayed
- Tagging is done by request author
- For well known lines, tagging could be performed by data providers
- In addition to a list of keywords, free text description could be used