

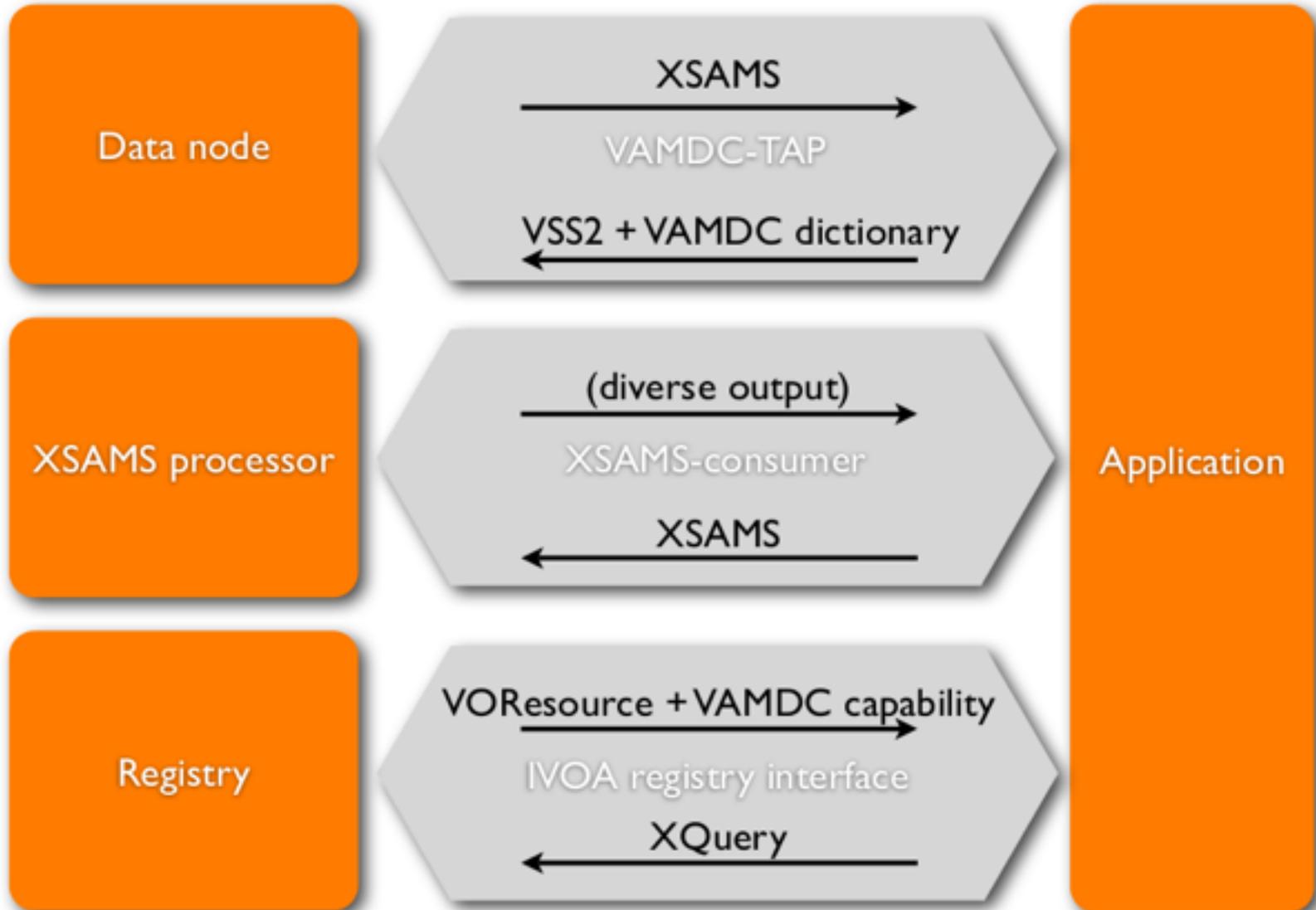
# La base de données des espèces chimiques et la sémantique dans VAMDC

N.Moreau

Paris Observatory, LERMA

- VAMDC : e-infrastructure pour l'accès aux données de physique A&M
- Utilisation d'un middleware placé au dessus des bases de données
- Définition de standards pour :
  - l'interrogation des bases (VAMDC-TAP)
  - le format d'échange des données (XSAMS)
  - la description des services (extension de VOResource)

# Infrastructure VAMDC



## La base de données des espèces

- <http://species.vamdc.eu>
- Un dépôt centralise les espèces de toutes les bases de données
- Découverte rapide du contenu de chaque base
- Ajout de fonctionnalités à d'autres éléments de l'infrastructure :
  - autocomplétion des noms
  - test dynamique de l'existence d'une espèce dans une base

- La représentation des noms des espèces chimiques n'est pas homogène
- Solution pour les atomes : utiliser le symbole ( + la charge pour les ions )
- Pour les molécules, l'écriture est ambiguë
- VAMDC a choisi d'utiliser le standard InChi (International Chemical Identifier)

- On décrit l'espèce au format InChi :

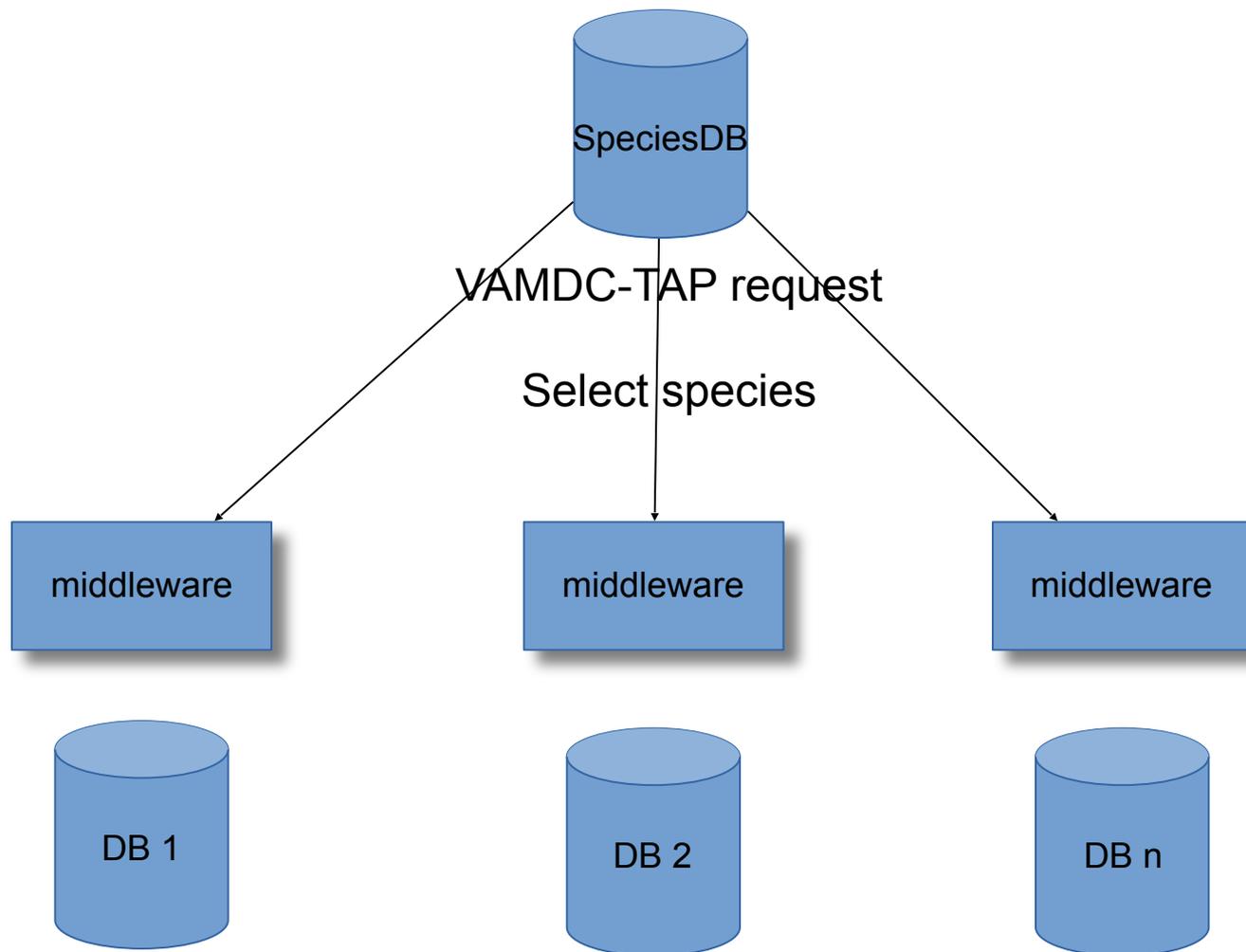
Ex : 1S/H, 1S/He, 1S/C/q+1

- On génère un hash de 27 caractères (SHA-256) :

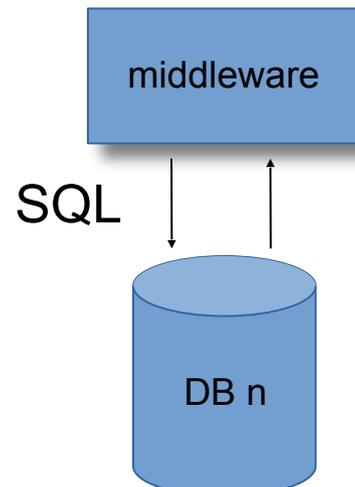
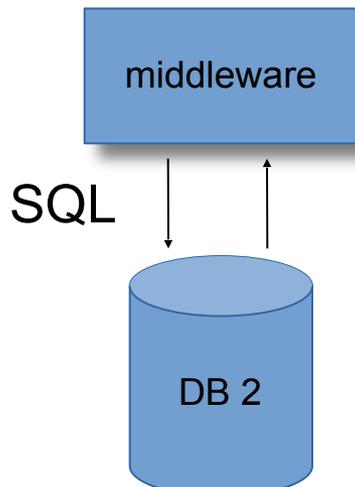
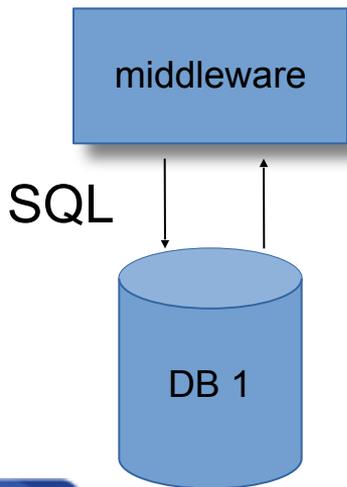
GKDCRJWYAGBLFY-UHFFFAOYSA-N

- Utilisation de l'identifiant pour lier les différentes "versions" d'une même espèce

# Processus d'alimentation

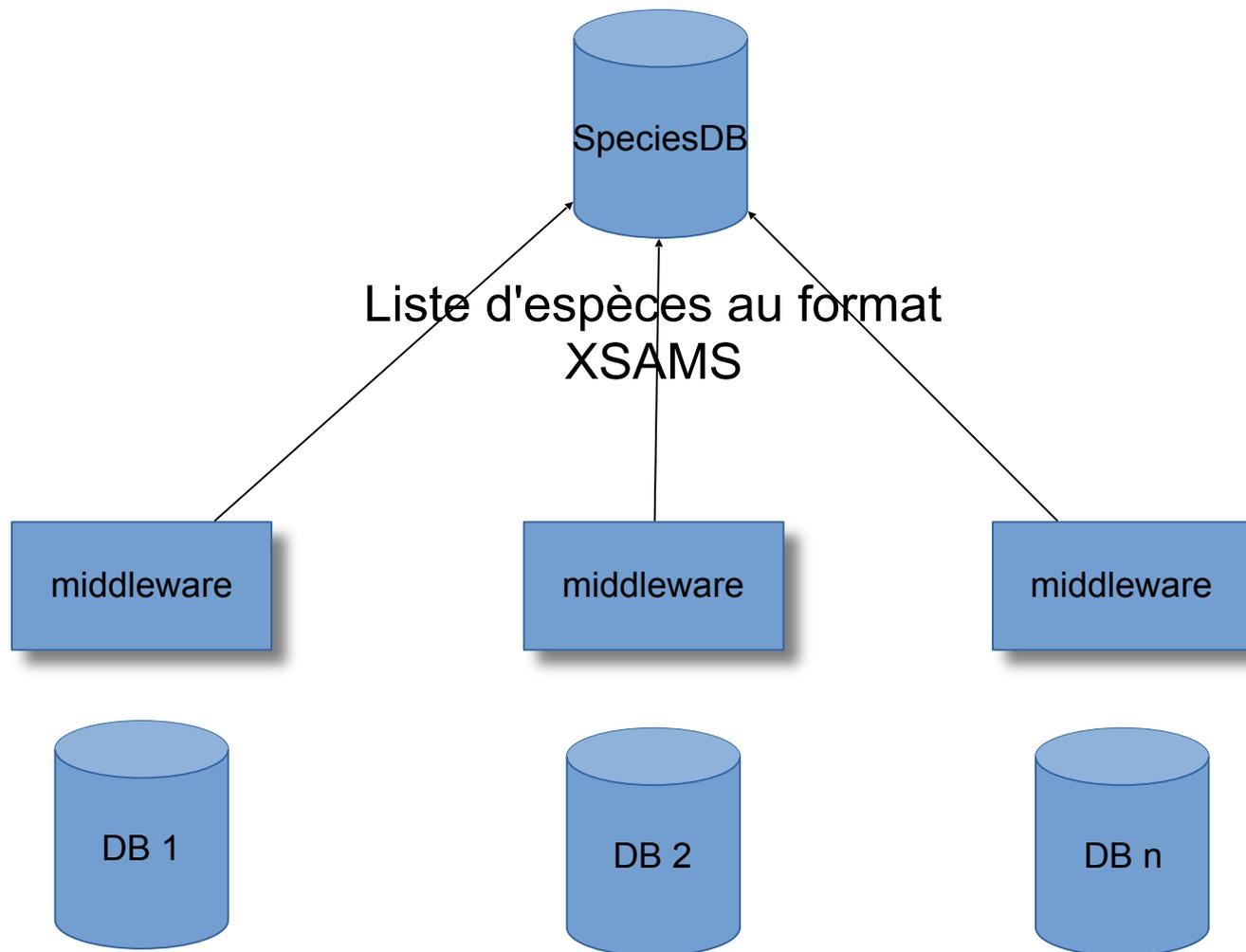


# Processus d'alimentation



ASOV 2017

# Processus d'alimentation



[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/c1-3-2	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/c1-3-2	36	NVLRFXKSQ QPKAD- UHFFFAOYS A-N	0

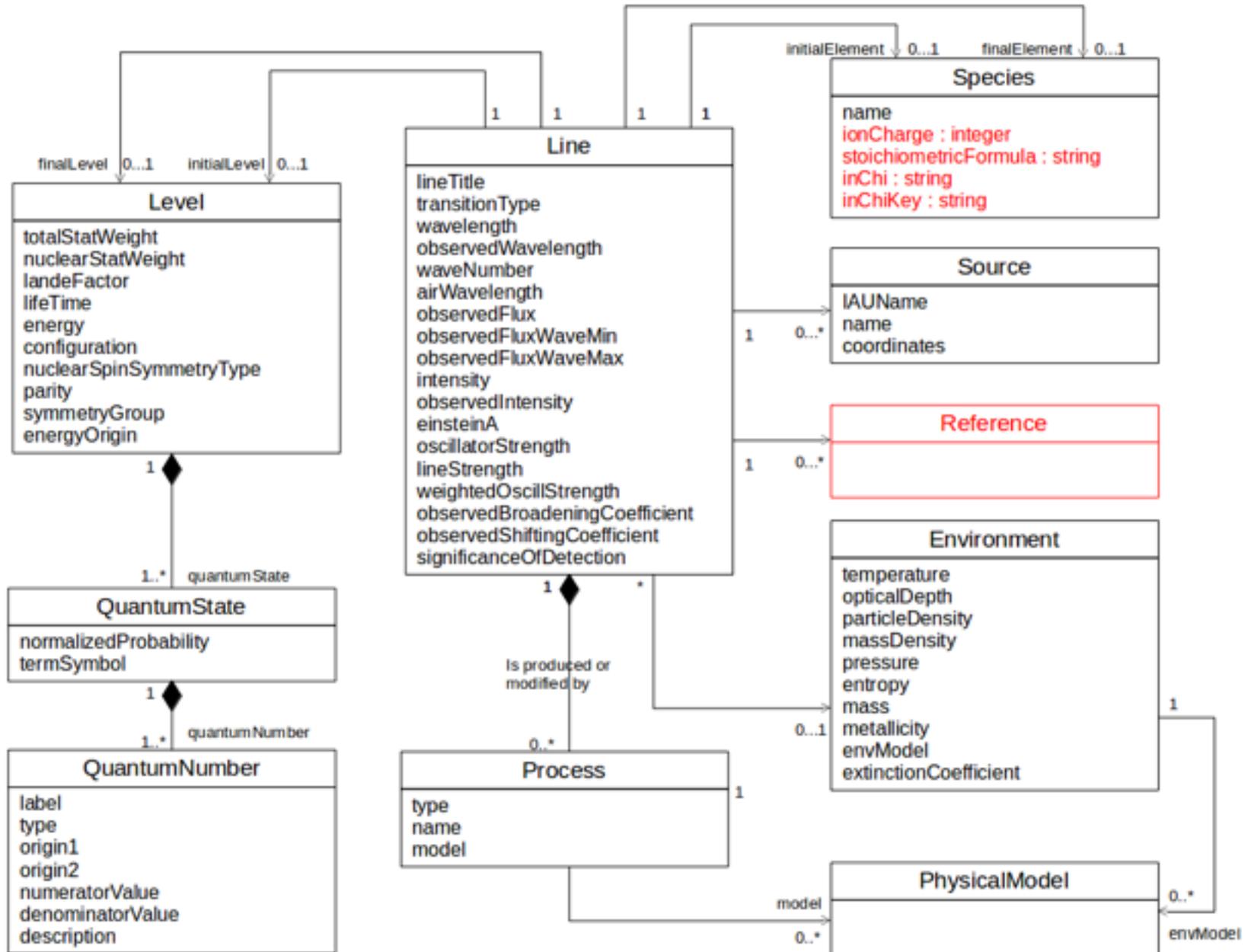
[LXcat](#) ( Atoms, Collisions, Molecules )

Name	Stoichiometric formula	Formula	InChI	Mass number	InChIKey	Charge
C3 radical	C3	C3	InChI=1S/C3/c1-3-2	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/c1-3-2	36	NVLRFXKSQ QPKAD- UHFFFAOYS A-N	0

- Objectif de convergence entre SLAP et VAMDC depuis Interop Trieste 2016
- Proposition d'ajouter l'inchkey comme attribut facultatif de Species dans le SSLDM
- Permet de comparer plus facilement l'identité de 2 espèces que l'analyse de chaînes de caractères



- Bases de données accessibles de façon standardisée, via un middleware
- Chaque base décrit ses propres capacités
  - type de données retournées
  - paramètres de requêtes acceptés
- Capabilities accessibles via une requête getCapabilities

Ex : [CDMS](#)

- Capabilities déclarées à partir de mots clés d'un vocabulaire

<http://dictionary.vamdc.eu>

- Mots clés créés à partir de leur position dans le schéma XSAMS

Ex :

AtomMassNumber (Atom → MassNumber)

AtomStateTotalAngMom (AtomState → TotalAngularMomentum)

RadTransProbabilityA (RadiativeTransition → ProbabilityA)

- Ils permettent de créer les requêtes TAP
- Les champs sont comme les colonnes d'une unique table
- Requêtes faites sans jointure :

```
select *  
where (RadTransWavelength >= 500.0  
AND RadTransWavelength <= 600.0)  
AND ((AtomSymbol = 'He'))
```

# Description des bases de données

- Descriptions des nodes analysées lors de l'ajout dans la base
- Mots clés utiliser pour décrire le contenu

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

- Possibilité d'ajouter un système de recherche par type de données
- Une grand partie de la procédure de recherche des données déjà faite par l'utilisateur