

# VAMDC : Le Centre Virtuel de Données Atomiques et Moléculaires

L'infrastructure et l'intégration des bases de données CaSDa

Cyril Richard

Laboratoire Interdisciplinaire Carnot de Bourgogne  
UMR 6303 CNRS – Université Bourgogne Europe  
Dijon, France

22 avril 2026

# Plan de la présentation

1 VAMDC : présentation et objectifs

2 Les bases CaSDa dans VAMDC

- 1 VAMDC : présentation et objectifs
- 2 Les bases CaSDa dans VAMDC

# Contexte : pourquoi des bases de données spectroscopiques ?

## Les données atomiques et moléculaires sont indispensables pour...

- La modélisation des **atmosphères planétaires** (Terre, exoplanètes, naines brunes)
- L'analyse des **spectres astrophysiques** (ISM, régions de formation stellaire)
- Les **calculs de transfert radiatif**
- La surveillance de **gaz à effet de serre** industriels
- Les applications en **plasmas** et fusion nucléaire

## Le problème

- **Dizaines** de bases de données hétérogènes
- Formats, unités, numérotations quantiques **différents**
- Difficile d'accès pour les non-spécialistes
- Données dynamiques : difficile à **citer**

⇒ Besoin d'une **infrastructure unifiée et interopérable**

# Qu'est-ce que VAMDC ?

## Virtual Atomic and Molecular Data Centre

Infrastructure européenne interconnectant des bases de données atomiques et moléculaires, offrant **un point d'accès unique** à l'ensemble de ces ressources.

### Historique

- 2009–2012 : Projet EU Framework 7 (FP7)
- 2012–2014 : Extension via SUP@VAMDC
- Depuis 2014 : **Consortium indépendant**  
~35 groupes de recherche
- Portail public : <https://portal.vamdc.eu>

### Chiffres clés (2020)

- **39 bases de données** connectées
- Couvrant la spectroscopie atomique et moléculaire, les processus collisionnels, la photodissociation...
- Principalement orienté **astrophysique**
- Impliqué dans l'EOSC (*European Open Science Cloud*)

## Le principe

VAMDC **ne stocke pas** les données : c'est un **enrobage** (*wrapper*) qui expose de façon unifiée des bases hétérogènes.

## Composants clés :

- **Node software** : transforme une BDD autonome en *data-node* VAMDC
- **Registre** : répertoire des services disponibles
- **VAMDC-TAP** : protocole d'accès aux données (dérivé de l'IVOA TAP)
- **VSS2** : langage de requête (sous-ensemble SQL)
- **XSAMS** : format de sortie unifié (XML)

## Exemple de requête VSS2

```
select * where  
(AtomSymbol = 'He'  
AND IonCharge = 0)
```

## Résultat

Toutes les bases connectées retournent leurs données sur He I dans le même format XSAMS.

## XML Schema for Atoms, Molecules and Solids

- Format de sortie **universel** pour toutes les bases VAMDC
- Décrit : niveaux d'énergie, transitions, processus collisionnels, références bibliographiques. . .
- Hautement flexible
- Compatible *case-by-case* pour les nombres quantiques moléculaires

## Outils de conversion disponibles

- XSAMS → format HITRAN 2012
- XSAMS → HTML avec export
- XSAMS → BibTeX

## Avantages

- Interopérabilité entre bases
- Comparaison directe de données
- Intégration dans des codes de modélisation (CASSIS, XCLASS, Cloudy. . .)
- Traçabilité des données avec DOI

Limitation : verbeux pour les très grandes bases (> millions de transitions)

# Le portail VAMDC : point d'entrée unique

## Interfaces de recherche

- Mode **guidé** (pas à pas) : pour les nouveaux utilisateurs
- Mode **avancé** : requêtes par espèce, processus, environnement
- Autocomplétion via la **Species Database** (InChIKey)
- Requête simultanée de **toutes les bases** connectées

## Affichage des résultats

- Tableau récapitulatif par base
- Téléchargement XSAMS
- Visualisation graphique des données
- Lien vers le **Query Store**

## Species Database

Référentiel centralisé mis à jour quotidiennement :

- Toutes les espèces de toutes les bases
- Identification unique par **InChIKey**
- Noms, formules, isotopologues. . .

## Cas concret

Chercher  $\text{Ca}^+$  : VAMDC garantit la *même* espèce dans NIST et VALD, évitant les confusions fréquentes (unités air/vide, stage d'ionisation. . .)

## Problème posé

Les données VAMDC sont **dynamiques** : les bases évoluent. Comment citer de façon reproductible ?

## Solution : le Query Store

- Implémente les recommandations de la **Research Data Alliance** (RDA)
- Chaque requête reçoit un **identifiant persistant unique**
- Possibilité d'assigner un **DOI** via Zenodo
- Les auteurs des données obtiennent un **crédit automatique** via Scholix/OpenAIRE
- Résultats interrogeables à <https://cite.vamdc.eu>

## Impact pour les producteurs de données

- Actuellement : seule la BDD est citée
- Avec VAMDC+DOI : les auteurs originaux des mesures/calculs sont cités
- Amélioration de l'impact bibliométrique des spectroscopistes

## Bases connectées au QS (2020) :

SHeCaSDa, MeCaSDa, GeCaSDa, TFMCaSDa, RuCaSDa, VALD, Stark-B, CDMS, JPL, BeamDB, MoID, SESAM...

## Spectroscopie moléculaire

- **HITRAN** – atmosphères terrestres et planétaires
- **CDMS / JPL** – spectrométrie micro-ondes/submm
- **ExoMolOP** – opacités pour exoplanètes
- **S&MPO** – ozone
- **CaSDa** (Dijon) – molécules de haute symétrie

## Spectroscopie atomique

- **VALD** – physique stellaire
- **NIST ASD** – données critiques
- **CHIANTI** – plasmas coronaux
- **Stark-B** – élargissement Stark

## Processus collisionnels & autres

- **BASECOL** – collisions rotationnelles/vibrationnelles
- **KIDA / UDfA** – cinétique astrochimique
- **BeamDB / MoID** – sections efficaces électroniques
- **GhoSST / SSHADE** – spectroscopie de solides
- **PAH** – hydrocarbures aromatiques polycycliques

## Principe FAIR

VAMDC implémente les principes **Findable, Accessible, Interoperable, Reusable** à une granularité fine pour toutes ses données.

## CASSIS (IRAP, Toulouse)

- Logiciel d'analyse de spectres astronomiques (ALMA, Herschel...)
- Interroge VAMDC via le protocole TAP
- Identification de raies et modélisation ETL/non-ETL
- Accès à CDMS, JPL, MeCaSDa... via une interface unique

## XCLASS (Université de Cologne)

- Plugin CASA pour l'analyse de spectres ALMA
- Base SQLite locale peuplée via `vamdcLib` (Python)
- Modélisation ETL et non-ETL, ajustement de cubes spectraux
- Identification automatique de molécules

## Message clé

VAMDC est bien plus qu'un catalogue : c'est une **infrastructure active** qui alimente des codes scientifiques de premier plan.

## Limitations actuelles

- **Grandes bases** : XSAMS verbeux, délais d'attente, troncature  
→ requêtes asynchrones en cours
- **Identification des états atomiques** entre bases différentes
- Mise à jour du *node software* (Django v2 → v3)
- Couverture incomplète du Query Store

## Développements futurs

- Interface utilisateur plus intuitive (NLP)
- Format SQLite local en cache
- Nouveau service bibliographique
- Recherche sémantique (ontologies)
- Visualisation dynamique  
Python/Bokeh/Jupyter
- Évaluation qualité et *rating reviews* des données

- 1 VAMDC : présentation et objectifs
- 2 Les bases CaSDa dans VAMDC

## CaSDa : Calculated Spectroscopic Database

Ensemble de bases de données de spectroscopie moléculaire haute résolution, développées au **Laboratoire ICB** (UMR 6303 CNRS – Université Bourgogne Europe), Dijon.

**Spécificité** : listes de raies *calculées* à partir d'analyses expérimentales

- Hamiltoniens effectifs et moments de transition ajustés sur spectres expérimentaux
- **Précision expérimentale** pour les positions
- Données non extrapolées au-delà du domaine analysé
- Description complète des états propres (vecteurs propres)

## Outils de calcul

- **XTDS** : interface Java pour simulation et analyse des polyades ( $T_d$ ,  $O_h$ ,  $C_{3v}$ ,  $D_{2h}$ )
- **SPVIEW** : affichage et assignation graphique des spectres haute résolution
- Disponibles sur <https://icb.cnrs.fr/software>

## Accès

Libre et sans inscription  
<https://vamdc.icb.cnrs.fr>

## Principe général

- Les niveaux vibrationnels sont regroupés en **polyades** d'après les résonances entre modes normaux
- Polyade  $P_n$  définie par :

$$n = \sum_{k=1}^N i_k \nu_k$$

- Les **Hamiltoniens effectifs** sont décomposés polyade par polyade
- Calculs dans une base rovibromotionnelle symétrisée
- Adapté aux molécules de **haute symétrie** :  $T_d$ ,  $O_h$ ,  $C_{3v}$ ,  $D_{2h}$

## Exemple : CH<sub>4</sub> (MeCaSDa)

$$\nu_1 \approx \nu_3 \approx 2\nu_2 \approx 2\nu_4$$

$$\text{Schéma : } (i_1, i_2, i_3, i_4) = (2, 1, 2, 1)$$

- $P_1$  : Dyade  $\nu_2/\nu_4$
- $P_2$  : Pentade  $\nu_1/\nu_3/2\nu_2 \dots$
- Etc.

Les bases contiennent également les **coefficients de décomposition** des vecteurs propres sur la base initiale, utiles pour les effets de mélange de raies.

10 molécules – accès via <https://vamdc.icb.cnrs.fr>

## Non mis à jour

- **MeCaSDa** ( $\text{CH}_4$ ) – >12 M transitions
- **RuCaSDa** ( $\text{RuO}_4$ ) – industrie nucléaire

## Mis à jour

- **ECaSDa** ( $\text{C}_2\text{H}_4$ )
- **SHeCaSDa** ( $\text{SF}_6$ )
- **TFMeCaSDa** ( $\text{CF}_4$ )
- **GeCaSDa** ( $\text{GeH}_4$ )
- **TFSiCaSDa** ( $\text{SiF}_4$ )

## Nouvelles (2024)

- **UHeCaSDa** ( $\text{UF}_6$ )
- **ChMeCaSDa** ( $\text{CH}_3\text{Cl}$ )
- **SiCaSDa** ( $\text{SiH}_4$ )

## Format de sortie

- Format HITRAN 2012 (160 caractères)
- Sections efficaces (2 colonnes)
- Tracé interactif intégré
- Compatible XSAMS (VAMDC)

*Note* : pas d'abondances isotopiques incluses (contrairement à HITRAN)

## Gaz à effet de serre

- **SF<sub>6</sub>** (SHeCaSDa) : GES d'origine anthropique, durée de vie ~3200 ans
- **CF<sub>4</sub>** (TFMeCaSDa) : PFC d'origine anthropique, durée de vie 50 000 ans ; suivi Protocole de Kyoto
- Nouvelles transitions hot-band pour modélisation atmosphérique

## Planétologie

- **GeH<sub>4</sub>** (GeCaSDa) : détecté dans Jupiter et Saturne ; mission Juno
- **C<sub>2</sub>H<sub>4</sub>** (ECaSDa) : Jupiter, Saturne, Titan, Neptune
- **CH<sub>4</sub>** (MeCaSDa) : omniprésent dans le système solaire, naines brunes, exoplanètes

## Volcanologie & astrophysique

- **SiF<sub>4</sub>** (TFSiCaSDa) : traceur de magma dans les panaches volcaniques ; potentiellement présent sur Io (lune de Jupiter)
- **CH<sub>3</sub>Cl** (ChMeCaSDa) : détecté en 2017 autour d'une protoétoile et dans la comète 67P ; destruction de l'ozone
- **SiH<sub>4</sub>** (SiCaSDa) : enveloppes circumstellaires (IRC+10216) ; nouvelles transitions rotationnelles SOLEIL

## Industrie nucléaire

- **RuO<sub>4</sub>** (RuCaSDa) : surveillance d'accidents nucléaires
- **UF<sub>6</sub>** (UHeCaSDa) : séparation isotopique

## Structure commune des bases CaSDa

- Modèle **relationnel SQL** – 13 tables identiques pour toutes les bases
- Tables : transitions, niveaux d'énergie, symétries, polyades, vecteurs propres. . .
- Portail web : `vamdc.icb.cnrs.fr`
- Requêtes paramétrables : isotopologue, plage en nombre d'onde, type de données
- Tracé interactif intégré (statique si >500 000 transitions)

## Format de sortie

- **HITRAN 2012** (160 car.) : format communauté
- **XSAMS** : format VAMDC pour interopérabilité
- Outil de conversion XSAMS → HITRAN disponible sur le portail VAMDC

## Particularité des nombres quantiques

Le formalisme de Dijon utilise les nombres quantiques des **toupies sphériques** (Groupe 3 HITRAN) pour *toutes* les molécules, y compris les toupies asymétriques ( $C_2H_4$ ) ou symétriques ( $CH_3Cl$ ) → différences de notation par rapport à HITRAN.

## Hébergement & archivage

- Serveurs : Direction du numérique, Université de Bourgogne
- Métadonnées : **DataUBFC** (référéncé par DataCite et <https://recherche.data.gouv.fr>)

## Mises à jour prévues

- Analyse de nouvelles bandes de **SiH<sub>4</sub>** → mise à jour de SiCaSDa
- Bandes  $\nu_1/\nu_4$  de **CH<sub>3</sub>Cl** (observation haute résolution à LISA)
- Développement d'une version de XTDS capable de fournir les **valeurs de  $K$**  pour les toupies symétriques
- Téléchargement des **vecteurs propres complets** depuis le portail web

## Nouvelles molécules potentielles

- Autres toupies sphériques/octaédriques (SeF<sub>6</sub>...)
- Molécules de symétrie C<sub>3v</sub> supplémentaires (Trioxane, CH<sub>3</sub>F, NF<sub>3</sub>...)
- Connexion complète au **Query Store VAMDC** pour toutes les bases
- Requêtes asynchrones pour les très grandes bases (MeCaSDa, SHeCaSDa)

## VAMDC

- Infrastructure interopérable fédérant **39 bases** de données A&M
- Standard XSAMS + protocole VAMDC-TAP
- Portail unique pour toutes les disciplines
- Query Store + DOI : **citation des données** résolue
- Pilier de l'Open Science européen (EOSC)
- Intégré dans CASSIS, XCLASS, Cloudy...

<https://portal.vamdc.eu>

## CaSDa dans VAMDC

- **10 molécules** de haute symétrie ( $T_d$ ,  $O_h$ ,  $C_{3v}$ ,  $D_{2h}$ )
- Précision expérimentale des positions et intensités
- Données utiles pour atmosphères, planétologie, astrophysique, nucléaire
- Intégré et librement accessible via VAMDC
- Données référencées dans DataUBFC / DataCite

<https://vamdc.icb.cnrs.fr>