Online database services for atomic and molecular data

Christian Hill
University College London

IAEA Meeting on Developments in Data Exchange
28-29 July 2016
Atomic and molecular data resources

- **VAMDC** — Virtual Atomic and Molecular Data Centre
- **ExoMol** — High-temperature molecular line lists for exoplanet and cool star atmospheres
- **Aladdin** — IAEA database of electron / photon / heavy particle / surface collisional cross sections
- **QDB** — Quantemol database of chemistries for low-temperature plasma research
- **HITRANonline** — Transition line lists for atmospheric molecular spectroscopy
VAMDC — Virtual Atomic and Molecular Data Centre

http://vamdc.eu/

• Standards and services for interchange of AM data
• “Portal” service for online access to multiple databases with a single query
• XSAMS data format for standardised representation of AM data
• “Node” software for hosting AM data
Applications capable of combining, extracting and processing data from all VAMDC member databases

Client Applications

Registry

VAMDC Core Infrastructure

Legacy Applications

Format converters

Monitoring

XSAMS

DB

VOSI
VAMDC — Queries with VSS2

• VSS2 is an “SQL-like” query language for requesting data from VAMDC-compliant databases

• Example 1: Data on Mn\(^+\), Mn, Mn\(^-\), Mn\(^-2\), …

  SELECT Species WHERE AtomNuclearCharge = 25 AND AtomIonCharge < 2

• Example 2: All data on He\(^z\) + O\(^z'\) → Products

  SELECT * WHERE reactant0.atomsymbol="He" and reactant1.atomsymbol="O"
VAMDC — The VSS2 API

• Submitting VSS2 queries to an online database as a “GET” request.

• Example: All data on He$^z$ + O$^{z'}$ → Products

http://aladdin2.xnx.webfactional.com/node/tap/sync//
node/tap/sync/?
REQUEST=doQuery&LANG=VSS2&FORMAT=XSAMS&QUERY=SELECT *
WHERE reactant0.atomsymbol="He" AND
reactant1.atomsymbol="O"

Returns an XSAMS document to the user
VAMDC — The VAMDC Portal


- Building VSS2 queries through a user-friendly interface in a browser
- Recent developments: “Guided Query”
  [http://portal.vamdc.org/vamdc_portal/queryTree.seam](http://portal.vamdc.org/vamdc_portal/queryTree.seam)
- Example: CO$_2$ absorption in 15 µm band
  - Select “For radiative process”
  - Select Wavelength=“14 — 16 um”
  - Select “Number of molecules”=1
  - Search for “Stoichiometric Formula”=CO2
  - Click “Submit query and find data”
VAMDC — The XSAMS format

http://standards.vamdc.org/

• An XML-based format for the transmission of AM data
• Originally conceived at the IAEA
• Developed and refined over 5 years of the VAMDC project
• “A consistent, self-describing format designed to promote good practice”:
  ‣ units
  ‣ uncertainties
  ‣ data source attribution
  ‣ time-stamping
  ‣ …
VAMDC — The XSAMS format

- Atomic States
- Molecular States
- Radiative Processes
- Collisional Processes
- “Environments”
- Functions with parameters and arguments
- Sources
- “Methods” (experimental, theoretical, …)
- Data tables, references to external files
VAMDC — The XSAMS format

- Example XSAMS snippet:

```xml
<AtomicState stateID="Schianti-1001006">
  <Description>2s2 2p</Description>
  <AtomicNumericalData>
    <StateEnergy methodRef="Mchianti-EXP">
      <Value units="1/cm">63.4</Value>
      <Evaluation recommended="true">
        <Quality>A+</Quality>
      </Evaluation>
      <Accuracy type="statistical" relative="False">0.2</Accuracy>
    </StateEnergy>
  </AtomicNumericalData>
  <AtomicQuantumNumbers>
    <TotalAngularMomentum>1.5</TotalAngularMomentum>
  </AtomicQuantumNumbers>
</AtomicState>
```
VAMDC — The XSAMS format

• Example XSAMS snippet (molecular state):

```xml
<MolecularState stateID="S14-H2O-1">
  <Description>A state of H2(16O)</Description>
  <MolecularStateCharacterisation>
    <StateEnergy>
      <Value units="1/cm">1813.223400</Value>
    </StateEnergy>
    <TotalStatisticalWeight>69</TotalStatisticalWeight>
  </MolecularStateCharacterisation>
  <nltcs:QNs>
    <nltcs:ElecStateLabel>X</nltcs:ElecStateLabel>
    <nltcs:J>11</nltcs:J>
    <nltcs:Ka>3</nltcs:Ka>
    <nltcs:Kc>8</nltcs:Kc>
    <nltcs:v1>0</nltcs:v1>
    <nltcs:v2>0</nltcs:v2>
    <nltcs:v3>0</nltcs:v3>
  </nltcs:QNs>
</MolecularState>
```
VAMDC — Databases producing XSAMS output

- Atomic Spectroscopy:
  - NIST
  - VALD
- Molecular Spectroscopy
  - CDMS
  - HITRAN
- Scattering and Reactions
  - KIDA
  - BASECOL
  - IDEADB
  - LXcat
  - UMIST
VAMDC — The XSAMS format

• Advantages
  ‣ Self-validating (to some extent)
  ‣ Text-based: no encoding issues
  ‣ XML-parsing tools exist

• Disadvantages
  ‣Verbose, large files
  ‣ Complex to write and read
  ‣ Conversion tools?
  ‣ Adoption by AM community?
VAMDC — More recent developments

- **29 databases** connected to the portal, including: BASECOL, CDMS, HITRAN, Chianti, LXCat, NIST Atomic Spectra Database, VALD
- **New output formats** from XSAMS:
  - BibTeX citations
  - HTML table view
  - XSAMS2SME (“Spectroscopy Made Easy”)
- **Time-stamping** (Last database update)
- **Direct editing** of VSS2 query
HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE 14 COMPETING STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.

YEAH!

[Soon:] SITUATION:
THERE ARE 15 COMPETING STANDARDS.
ExoMol — High-temperature molecular line lists

http://exomol.com/

- Ab initio calculations of energy levels and Einstein A coefficients
- For high-$T$, one has to consider many transitions, even for small molecules:
  - $^{12}\text{C}^{32}\text{S}$: $1.99 \times 10^5$ transitions
  - $^{51}\text{V}^{16}\text{O}$: $2.77 \times 10^6$ transitions
  - $^{12}\text{C}^1\text{H}_4$: $9.82 \times 10^9$ transitions
  - $^{31}\text{P}^1\text{H}_3$: $1.68 \times 10^{10}$ transitions
  - ... and more to come ($\text{SO}_3$, $\text{H}_2\text{CO}$, ...)
ExoMol — Data Format

- XSAMS is a non-starter with $10^{10}$ transitions
- Data provided in compressed, plain-text files
- Separate out the states from the transitions, e.g. CH$_4$:

**States**

$[8.19 \times 10^6]$

<table>
<thead>
<tr>
<th>ID</th>
<th>E</th>
<th>J</th>
<th>g</th>
<th>QNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Transitions**

$[9.82 \times 10^{10}]$

<table>
<thead>
<tr>
<th>ID’</th>
<th>ID”</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>...</td>
</tr>
</tbody>
</table>
ExoMol — API

- All ExoMol-calculated data is available at URLs with a defined format:

  http://www.exomol.com/db/<molec>/<iso>/
  <dataset>/<iso>__<dataset>..<ext>

  - `<molec>`: Molecular formula
  - `<iso>`: Isotopologue formula in “slug” form
  - `<dataset>`: Name of the calculation dataset
  - `<ext>`: File extension indicating data type

  e.g. Partition function for $^{12}\text{C}^{1}\text{H}_4$ (“YT10to10” data set)

  http://www.exomol.com/db/CH4/12C-1H4/
  YT10to10/12C-1H4__YT10to10.pf
ExoMol — API

• Machine-readable definitions files:
  • `exomol.all`:
    ‣ always available at http://www.exomol.com/db/exomol.all
    ‣ summarises current status of data availability
    ‣ can be easily checked for updates, new line lists, etc.
  • “Definitions” files (`<iso>__<dataset>.def`) provide definitions for fields in dataset files
    ‣ Quantum numbers used
    ‣ Maximum temperature
    ‣ Wavenumber range
    ‣ Availability of cooling functions, broadening coefficients, …
ExoMol — Derived Data

- Astronomers often can’t use very large line lists, so we provide:
  - Calculated absorption cross sections at 0.01 cm\(^{-1}\) grid spacing …
  - … and an online service for binning down to lower resolutions
  - \(k\)-coefficient tables
ExoMol — High-temperature molecular line lists

- Uncertainties are only estimates…
- … except where energy levels can be replaced by experimental values (e.g. MARVEL) …
- … and in many cases the intensities are expected to be more accurate than experiment
- As yet, no attempt has been made to quantify uncertainties in cross sections or $k$-coefficients.
QDB — Quantemol database of “chemistries” for plasma research

- Under development as a commercial product...
- … with a free data-access component and paid-for “value-added” content
- Calculated and experimental data on:
  - Electron collision cross sections
  - Heavy-particle collisional processes
  - Radiative processes in plasmas (to come)
QDB — Quantemol database of “chemistries” for plasma research

- Validation of data sets by:
  - Experimental benchmarking from open sources
  - Calculations performed for a range of models (HPEM, Global_Kin, ChemKin, …)
  - User-contributed assessment and comments
  - Validation of entire chemistries by comparison with experiment
QDB — Quantemol database of “chemistries” for plasma research

• Data formats
  ‣ Where possible, collision cross sections are provided as a function of electron energy (text table format)
  ‣ Heavy-particle processes described by a set of Arrhenius-like parameters:

\[
k(T) = A \left( \frac{T}{T_{\text{ref}}} \right)^n \exp \left( -\frac{E}{k_B T} \right)
\]

  ‣ Individual data sets are not large…
  ‣ … but there may be a great many of them (especially for state-resolved processes involving molecules)
HITRANonline

http://hitran.org/

• Molecular transition line lists for modelling planetary atmospheres (especially Earth’s)
• Fixed-width 160 character format since 2004
• Data moved to relational database format during VAMDC project
• New format allows an arbitrary number of parameters, quantum assignments, etc.
• Original references / sources
HITRANonline
HITRANonline

http://hitran.org/

• New interface allows users to create, edit and save their own output formats

• HITRANonline also functions as a VAMDC node, and produces valid XSAMS documents

• HAPI (HITRAN Application Programming Interface) provides a Python library for manipulating HITRAN data, creating plots, etc. programmatically
Conclusions

• You can’t force users to adopt new standards
• Sustainability is crucial
• Provide software tools for interconversion between data formats that are actually used
• Visualisation is nice
• Data should have provenance and be easy to cite / attribute