A Proposal for a Relational Model for Spectral Lines

Margarida Castro Neves\textsuperscript{1,2}, Markus Demleitner\textsuperscript{1,2}, Jean-Michel Glorian\textsuperscript{4}, Nicolas Moreau\textsuperscript{3}

\textsuperscript{1}ARI, ZAH, Universität Heidelberg
\textsuperscript{2}German Astrophysical Virtual Observatory
\textsuperscript{3}Paris Astronomical Data Centre
\textsuperscript{4}Institut de Recherche en Astrophysique et Planétologie

IVOA Interop November 2020
LineTAP proposal

- Motivation
- Goal
- Main use case
- Current state
- Some open questions
Motivation

What we have now:
- VO SLAP/SSLDM, very few services, and few clients.
- VAMDC/XSAMS, very detailed data model, very complex.

New version of SLAP/SSLDM: would it make things better?
Why not create a relational data model, that is simple and can be queried by the TAP protocol?
- TAP/ADQL is already well known, no need to learn a new protocol
- existing libraries, already implemented in clients
- good for users, data providers and programmers :)

Margarida Castro Neves, Markus Demleitner, Jean-Michel Glorian, Nicolas Moreau
Our goal is to have a simple data model, by
- collecting use cases from different areas
- selecting data quantities that cover most use cases (from the very detailed VAMDC data model)
- keep only parameters needed by most common use cases
- if possible, keep an one-table data model
Main use case that drove the development so far:

”A user with a rough idea of the kind of physics in an emitting or absorbing region wants to identify a line in a spectrum.”

- the first idea was to deal only with atoms and not to think about molecules
- but it seems that there are also some use cases using molecules, which brings more complexity
- more use cases are needed
Current state

What has been done up to now:

- Some use cases collected (need more!!!)
- Preliminary list of relevant quantities selected from the VAMDC data model
- Mapping from VAMDC Data model to LineTAP
- Draft of IVOA Note being written
<table>
<thead>
<tr>
<th>column name</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>name of the species originating the line.</td>
</tr>
<tr>
<td>spectral_location</td>
<td>energy equivalent to wavelength in vacuum</td>
</tr>
<tr>
<td>spectral_error</td>
<td>integrated error for the spectral location</td>
</tr>
<tr>
<td>method</td>
<td>method the wavelength was obtained with</td>
</tr>
<tr>
<td>stoichiometric_formula</td>
<td>the symbol of the chemical element or the molecule formula</td>
</tr>
<tr>
<td>ion_charge</td>
<td>ionisation level</td>
</tr>
<tr>
<td>mass_number</td>
<td>mass number</td>
</tr>
<tr>
<td>upper_state_configuration</td>
<td>upper state configuration</td>
</tr>
<tr>
<td>lower_state_configuration</td>
<td>lower state configuration</td>
</tr>
<tr>
<td>upper_energy</td>
<td>energy of the upper state</td>
</tr>
<tr>
<td>lower_energy</td>
<td>energy of the lower state</td>
</tr>
<tr>
<td>inchi</td>
<td>chemical species inchi</td>
</tr>
<tr>
<td>inchikey</td>
<td>chemical species inchikey</td>
</tr>
<tr>
<td>einstein_a</td>
<td>Einstein A coefficient</td>
</tr>
<tr>
<td>oscillator_strength</td>
<td>oscillator strength of radiative transition</td>
</tr>
<tr>
<td>weighted_oscillator_strength</td>
<td>Weighted oscillator strength of radiative transition</td>
</tr>
<tr>
<td>line_strength</td>
<td>Total absorption by a spectra line</td>
</tr>
<tr>
<td>line_reference_doi</td>
<td>Digital Object Identifier of bibliography source</td>
</tr>
<tr>
<td>line_reference_uri</td>
<td>Web link to the publication</td>
</tr>
</tbody>
</table>
Some questions: Units and errors for numerical quantities

- VAMDC: wavelength type provides vacuum/air information, a numerical data type containing units and an accuracy data type.
- We would prefer to have wavelength value (always in vacuum with a fixed unit) and wavelength error in different columns of a table.
  → ADQL User Defined Functions will be provided for conversion.
- Which unit to use as default? One idea is to use an Energy unit (J), but right now the majority wants to keep length (Angstrom or m).
- Errors: statistical, systematic, estimated, etc
  → for simplicity the best would be to provide a combined error (use cases?)
Some questions: Search for molecules

- Including molecules to the data model makes it more complex
- How to refer to a molecule in a query?
  - By name: (Water, Carbon Monoxide ...) - not unique
  - By formula: (H2O, CO) - not unique
  - by InChIKey: unique, not human readable
- Add new tables in the LineTAP Data Model to map InChi to names and formulas, and ...
- Leave this to external service?
The IVOA Note draft

- Please take a look: https://github.com/mmpcn/slapvamdc
- Document is still in an early phase, full of ”TO DO”s
- Comments and contributions are welcome
- We need more use-cases

Thank you!