

# New model for datasets citation and extraction reproducibility in VAMDC

Carlo Maria Zwölf\*, Nicolas Moreau\*, and Marie-Lise Dubernet\*

June 2, 2016

## Abstract

In this paper we present a new paradigm for the identification of datasets extracted from the Virtual Atomic and Molecular Data Centre (VAMDC) e-science infrastructure. Such identification includes information on the origin and version of the datasets, references associated to individual data in the datasets, as well as timestamps linked to the extraction procedure. This paradigm is described through the modifications of the language used to exchange data within the VAMDC and through the services that will implement those modifications. This new paradigm should enforce traceability of datasets, favour reproducibility of datasets extraction, and facilitate the systematic citation of the authors having originally measured and/or calculated the extracted atomic and molecular data.

**Keywords:** database, data citation, atomic data, molecular data

## 1 Introduction

Atomic and Molecular Data Providers are often at the forefront in defining new paradigms for the dissemination of their data to the scientific community:

Since the second half of 1990s, with the spread of the Internet, some historical databases such as VALD [19], HITRAN [23], [24], the Submillimeter, millimeter and microwave spectral line catalog (so-called JPL Catalogue) [18], CDMS [16], [15], GEISA [13], [12] were precursor of a free online sharing of data, thus anticipating of some decades trends later generalised by governments and institutions with the *open data* and *open knowledge initiatives* [11], [14], [20].

---

\*LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne University, UPMC Univ Paris 06, 5 Place Janssen, 92190 Meudon, France

During the last ten years, Atomic and Molecular resources were highly fragmented and available through a variety of interfaces. This situation negatively impacted the full exploitation of the scientific data and has been a major bottleneck between the data providers and the wide body of users. From 2009 to 2014, the goal of the European FP7 projects *VAMDC* [5] and its successor *SUP@VAMDC* [27] was to provide the international community with a unique technical and collaborative framework for Atomic and Molecular data sharing: the VAMDC International Consortium [8]. The VAMDC technical e-infrastructure federates, in an interoperable way, a wide range of heterogeneous databases containing data coming from different physical communities, provides a single entry-point both for discovering resources and for accessing the available data in a unified way (<http://portal.vamdc.eu>). For building and maintaining this e-infrastructure, VAMDC has developed specific procedures, protocols, data formats and query languages (<http://standards.vamdc.eu>). The Atomic and Molecular data providers community, organised around the VAMDC Consortium, has thus anticipated the identification and resolution of problems that are currently discussed in international groups belonging to the Research Data Alliance (RDA). The goal of RDA is indeed to build the social and technical bridges that enable open sharing of data in the current highly fragmented global research data landscape.

Using the VAMDC facilities, scientists can now easily discover the atomic and molecular resources, and can access in a unique and practical way their contained data. However, the adoption of VAMDC by a large community has revealed a new set of challenging issues. The VAMDC infrastructure data are dynamic. A database exposed through the VAMDC infrastructure may evolve over time: the most recent and precise version of given data may replace old ones as in VALD [19, 25], or new datasets may be added to the already existing ones as in BASECOL [6, 4]. It may happen that some of these database evolutions are not systematically reported through new publications.

Assume then that a scientist extracts from VAMDC at a given time a “dataset”, composed of an ensemble of “data”, and wishes to use this “dataset” in order to produce some science that will be published into a scientific paper: how can he/she cite this “dataset” and the individual “data”? Since the database content may evolve, for the consistency of the scientific publication, the citation should refer to “datasets” well defined in space (where the “dataset” physically comes from) and time (at what time the “dataset” was produced and extracted). In addition the citations should contain pointers to the authors who originally measured, calculated and/or fitted the individual data. Moreover, for the reproducibility of the scientific process described into the paper referencing the “dataset”, everybody wishing to verify step-by step the procedures described into the paper, should be

able to easily recover the original “dataset” and replay the data-production workflow.

In spring 2014 we joined the RDA Data Citation Working Group in order to study these open questions. The goal of this Working Group [22] is to create the identification mechanisms that

- allows us to identify and cite arbitrary views of a “dataset” in a precise, machine actionable manner;
- allows us to cite and retrieve that “dataset” as it existed at a certain point in time, whenever the database is static or highly dynamic;
- is stable across different technologies and technological changes.

The very new model we will discuss through this paper is the mechanisms we may apply on the data coming from the VAMDC e-infrastructure in order to meet the recommendations of the RDA Working Group on Data Citation with the goal of obtaining

- a rigorous and reproducible data citation model,
- the reproducibility of data production (or data processing) workflows,

and in order to facilitate the citation of the original authors of the atomic and molecular data.

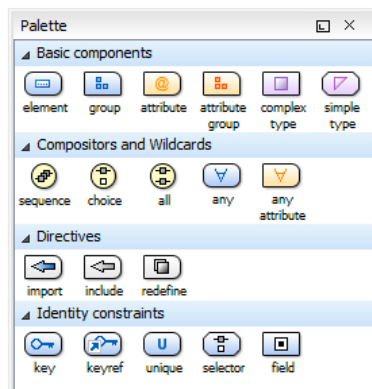


Figure 1: Graphical convention adopted for building graphical representations, starting from the XML schema.

## 2 Existing VAMDC solution for data citation

All the output of the VAMDC infrastructure are built upon an abstract object model: each specific output is an instance of this model. For historical

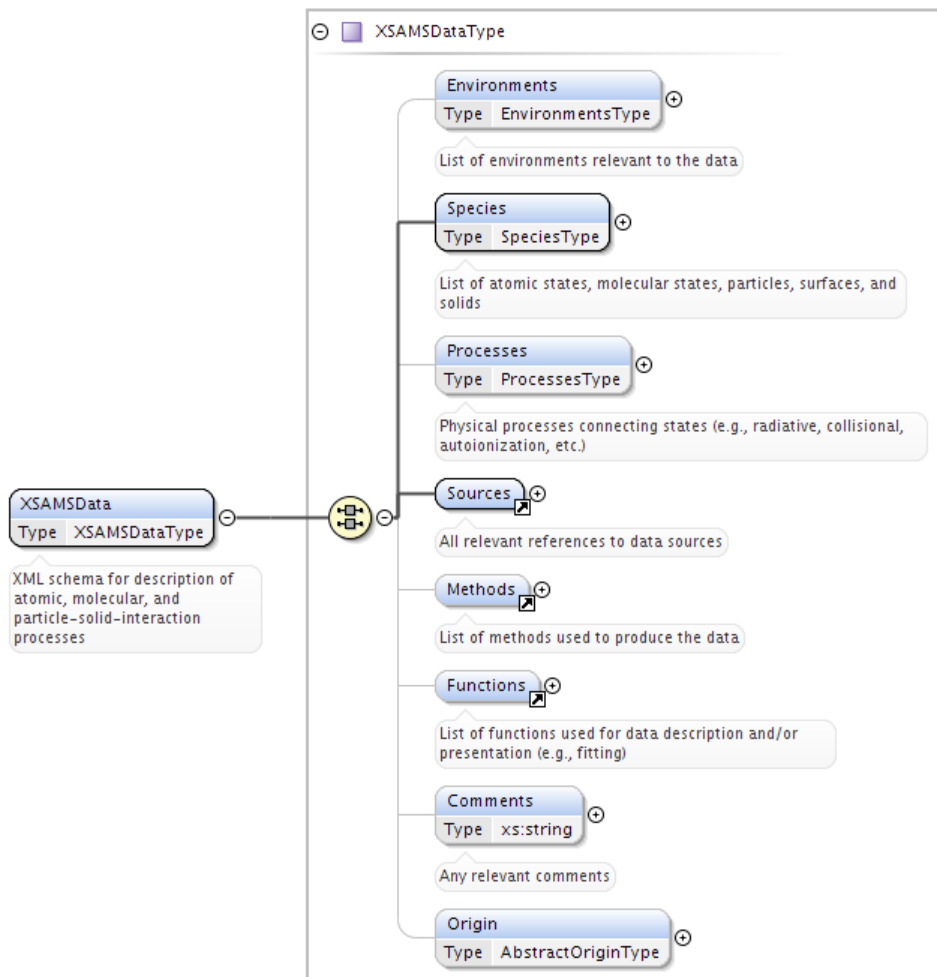


Figure 2: Representation of the root element of an XSAMS file.

reasons, without loss of generality, the model has been fixed into an XML format, the XSAMS XSD schema<sup>1</sup>.

## 2.1 Graphical convention for representing XML elements

The graphical diagrams displayed in this document are a simple rendering of every XML element contained in the XSAMS, and they are obtained following the graphical conventions of Fig. 1. In this graphical representation (see Fig. 2 for example) every complex element is linked with segments to the contained sub-elements. A bold segment indicates that the sub-element is required and a thin segment indicates that the sub-element is optional. Moreover, the cardinality of the contained sub-elements can be expressed on the segments. The detailed explanation of the XSAMS components can be found in the official VAMDC XSAMS documentation at <http://www.vamdc.org/standards>.

## 2.2 Example of data-citation need for scientific authors

Surveys of interstellar regions requires the use of spectroscopic information within the observed range of wavelengths/frequencies. As an example, the survey by [10] covers frequencies from 83302 MHz to 262404 MHz and detect emission from about 36 species. For that survey, [10] indicate that they used catalogues from two public databases [18], [15] and one private database of J. Cernicharo (private communication). We note that there is no knowledge of the exact dataset used in the analysis, and therefore the analysis may not be reproducible if the database contents evolve over the years. Secondly, we note there is no citation of the authors who produced the spectroscopic data. Obviously for such large surveys with so many species there is a large contribution from many experimental/theoretical spectroscopic papers. On the contrary, that for the non-local thermodynamical equilibrium analysis of spectra (that includes the use of collisional rate coefficients) about 12 publications related to collisional data are cited. This dichotomy of treatment could be first explained by the complexity of citing/finding many spectroscopy authors, while it is easy to cite a few collisional papers. Similarly, another survey [2] cites many spectroscopic databases without citing either the original authors or the version of data used in the survey's analysis. The study of Punanova et al. [21] cites the authors of transitions that are not part of a database, such as the hyperfine transitions of  $\text{N}_2\text{H}^+$  [17] and such as the  $1 \rightarrow 0$  transition of  $\text{C}^{17}\text{O}$  [9], but they cite the splatalogue catalog (<http://www.cv.nrao.edu/php/splat/>) for the  $1 \rightarrow 0$  transition

---

<sup>1</sup>XSAMS is the acronym for: XML Schema for Atoms, Molecules and Solids

of C<sup>17</sup>O. Those examples confirm that the origin and versioning of spectroscopic data might be lost along the chain of usage as soon as spectroscopic data are taken from a database.

Through the VAMDC e-infrastructure, the protocols presently allow to retrieve the references associated to the spectroscopic and collisional data in a straightforward and quick manner. For example, from the result page of the VAMDC portal (cf. figure 3) the “bibtex” visualisation processor allows one to get references attached to the data in a usable form for publication. This processor produces XSAMS output file (produced by the VAMDC infrastructure) by extracting the bibliographic information contained into the *Sources* field, which is embedded into each data entry (cf. figure 2 and 4). Thus, the use of the current version of VAMDC allows full citation of any data and such citation relies on the good will of the users. Even so, the citation process may become cumbersome when the extracted datasets come from many sources.

This latter issue and the former issue of reproducibility of the extraction process are addressed in the new paradigm of datasets citation that we describe below.

### 3 Introducing the evolutions of the XSAMS format

In this paragraph we discuss new mechanisms allowing one to identify through time and location, in a sustainable and unique way, the data sets extracted by users from the VAMDC e-infrastructure. The proposed mechanisms comes with an evolution of the XSAMS standard.

The practical usage of the described features is presented in section 5

#### 3.1 Introduction of the new schema elements for handling data versioning and origin

The root element of a XSAMS file is the *XSAMSData* object (cf. figure 2), which is an instance of the *XSAMSDataType* class. The innovation compared with the current official version of the XSAMS format (version 12.07) is the introduction of the *Origin* field, that contains the new mechanisms for handling the datasets citations.

The *Origin* field is an instance of a class extending the *AbstractOriginType* (cf. figure 5). This abstract type must contain:

- a unique Timestamp. This field defines the exact time when the data have been extracted by the infrastructure;

The screenshot shows the VAMDC consortium website interface. At the top, there is a navigation bar with links: Home, VAMDC databases, Guided query, Advanced query, Saved queries, Disclaimer, Citation policy, Info, and Feedback. Below this, the 'Query Execution' section shows a 'Done' status with buttons for 'Modify query', 'Stop waiting', and 'Save query'. A 'Comments' box is also present. The 'Your request' section contains a SQL query: `select * where ((target.InchiKey = 'UGFAIRIUMAVXCW-UHFFFAOYSA-N')) AND ((collider.AtomSymbol = 'He')) AND (CollisionCode = 'inel')`. The main part of the page is a table titled 'Results by node' with columns: Name, View data, Response, Last database update, Download, Species, States, and Processes. A dropdown menu is open over the 'View data' column, listing options: '-- Choose display --', '-- Choose display --', '\*\* Collisional data XSAMS to HTML', '\*\* BibTeX from XSAMS', '\*\* XSAMS multiplexor', 'Table views of XSAMS', 'Xsams2SME', 'Atomic spectroscopy XSAMS to HTML', and 'Molecular spectroscopy XSAMS to HTML'. The 'BibTeX from XSAMS' option is highlighted.

Name	View data	Response	Last database update	Download	Species	States	Processes
BASECOL: VAMDC-TAP interface	-- Choose display --	OK	01/07/2005 00:00	<a href="#">XSAMS</a>	6	0	2
KIDA: VAMDC-TAP interface	** BibTeX from XSAMS		19/09/2011 10:44	<a href="#">XSAMS</a>	3	0	0
RADAM - Ion Interactions	** XSAMS multiplexor		Not available		0	0	0
IDEADB - Innsbruck Dissociative Electron Attachment Database	Table views of XSAMS		Not available		0	0	0
	Xsams2SME						
	Atomic spectroscopy XSAMS to HTML						
	Molecular spectroscopy XSAMS to HTML						

Figure 3: Detail of the result page of the VAMDC portal highlighting the “bibtex” processor.

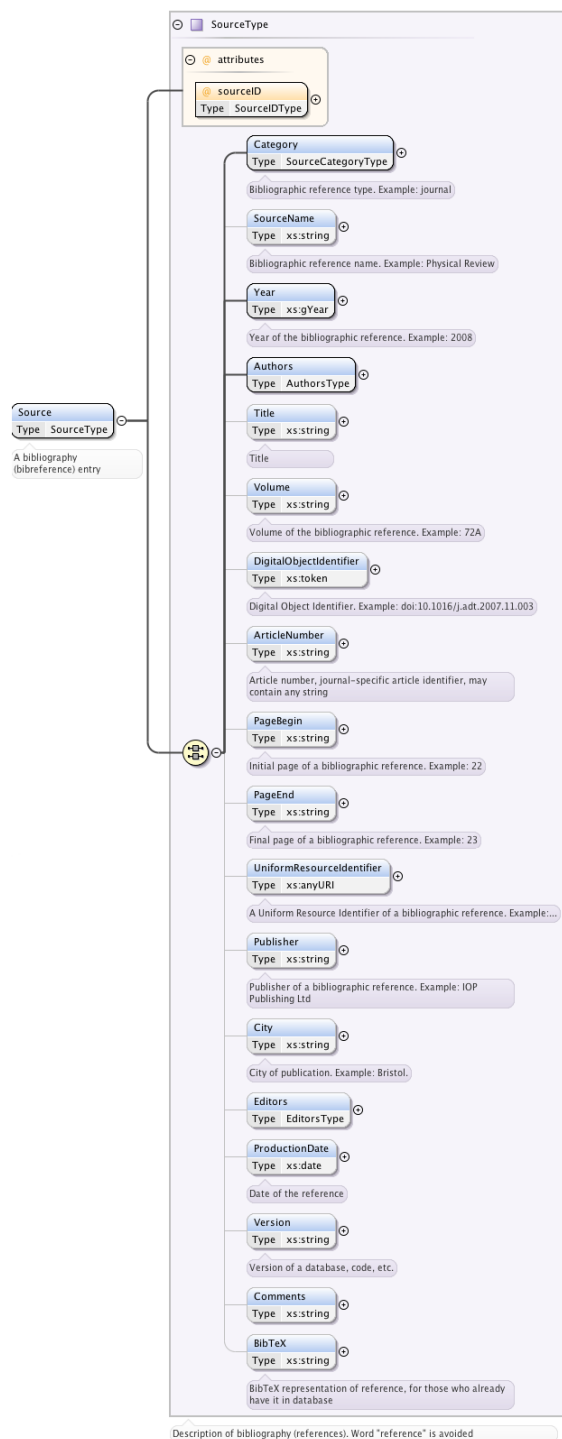


Figure 4: Details of the *Source* field contained in XSAMS files.



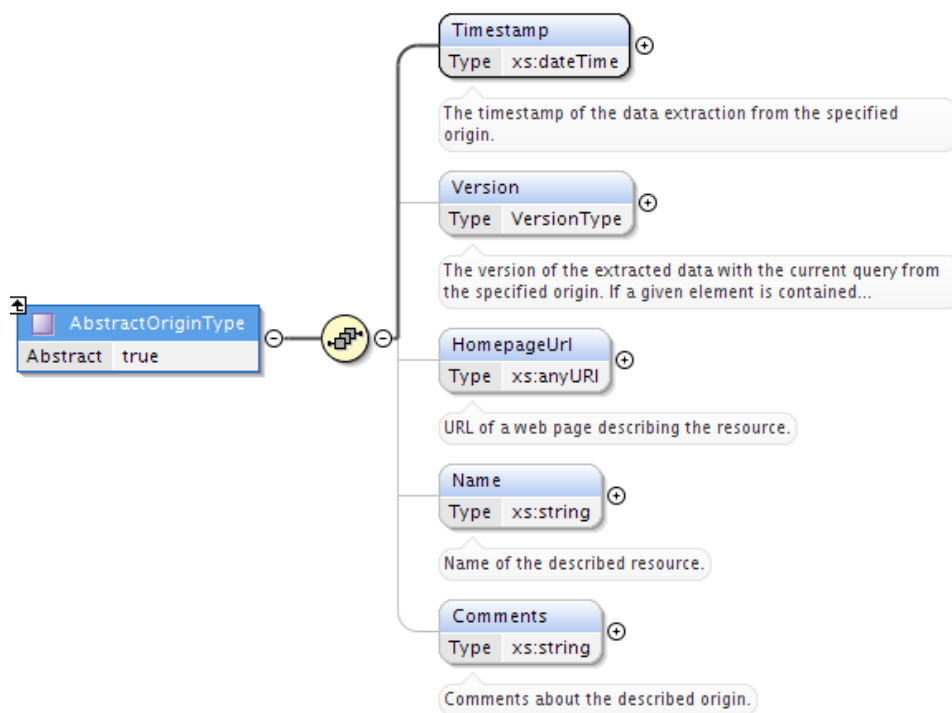


Figure 5: Representation of the abstract element *AbstractOriginType*.

- a *Version* field, instance of *VersionType* class (cf. paragraph 3.1.1);
- a *HomepageURL* field, which is the link to the website describing the resource that produced the data;
- a *Name* field, which is the name of the resource producing the data;
- a *Comments* field, which (as its name indicates), contains the comments related with the resources producing the data.

The three classes extending the abstract type *AbstractOriginType* are: *VamdcNodeOriginType*, *VamdcProcessorOriginType*, and *OtherOriginType*.

The *VamdcNodeOriginType* (cf. figure 6), as its name indicates, is used when the resource producing the data is a VAMDC node. It extends the *AbstractOriginType* by mandatorily adding:

- a *Query* field, which is a text string defining the query used for extracting the data from the described node.
- an *OriginIdentifier* field, which is the identifier of the node into the VAMDC infrastructure, as it appears into the VAMDC registry.

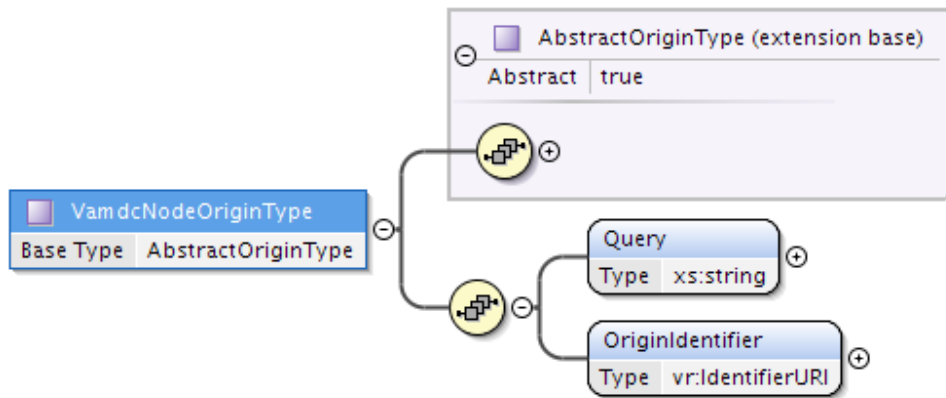


Figure 6: Representation of the element *VamdcNodeOriginType*.

The *VamdcProcessorOriginType* (cf. figure 7), as its name indicates, is used when the resource producing the data is a VAMDC processor. It extends the *AbstractOriginType* by mandatorily adding:

- an *OriginIdentifier* field, which is the identifier of the processor into the VAMDC infrastructure, as it appears into the VAMDC registry.

- the list (which must contain at least one element) of the *Origins* that produced the data used as inputs by the current processor.

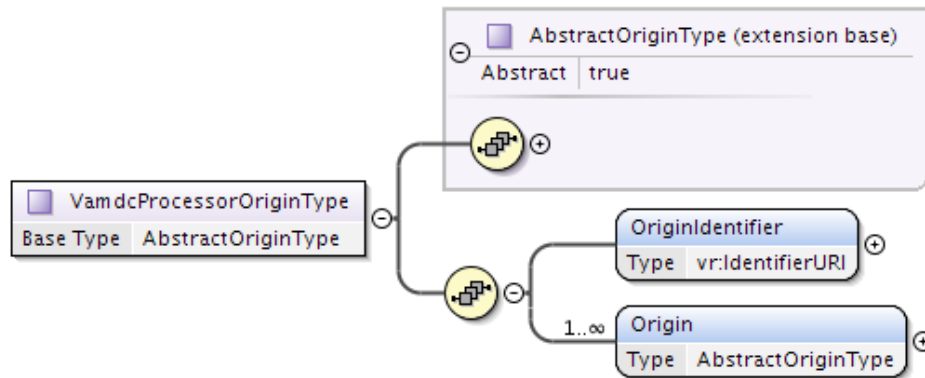


Figure 7: Representation of the element *VamdcProcessorOriginType*.

The *OtherOriginType* (cf. figure 8), as its name indicates, is used when the resource producing the data is not a VAMDC node and neither a VAMDC processor. It extends the *AbstractOriginType* by adding:

- an *OriginIdentifier* field, which is the identifier of the resource into the VAMDC infrastructure, as it appears in the VAMDC registry.
- the list of the *Origins* that produced the data used as inputs by the current resource.

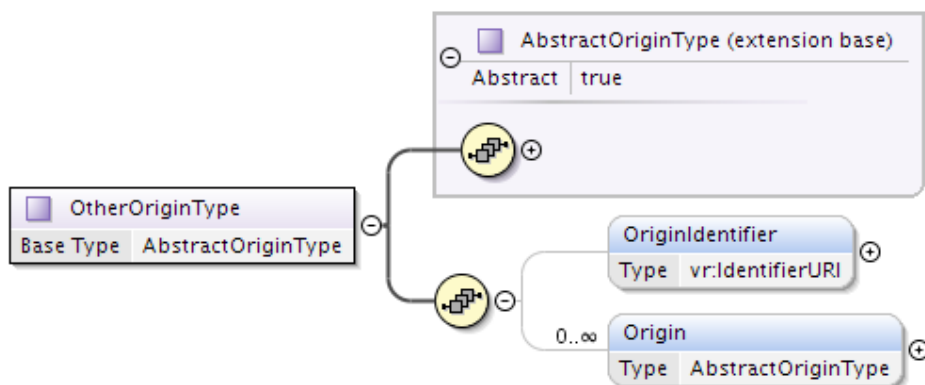


Figure 8: Representation of the element *OtherOriginType*.

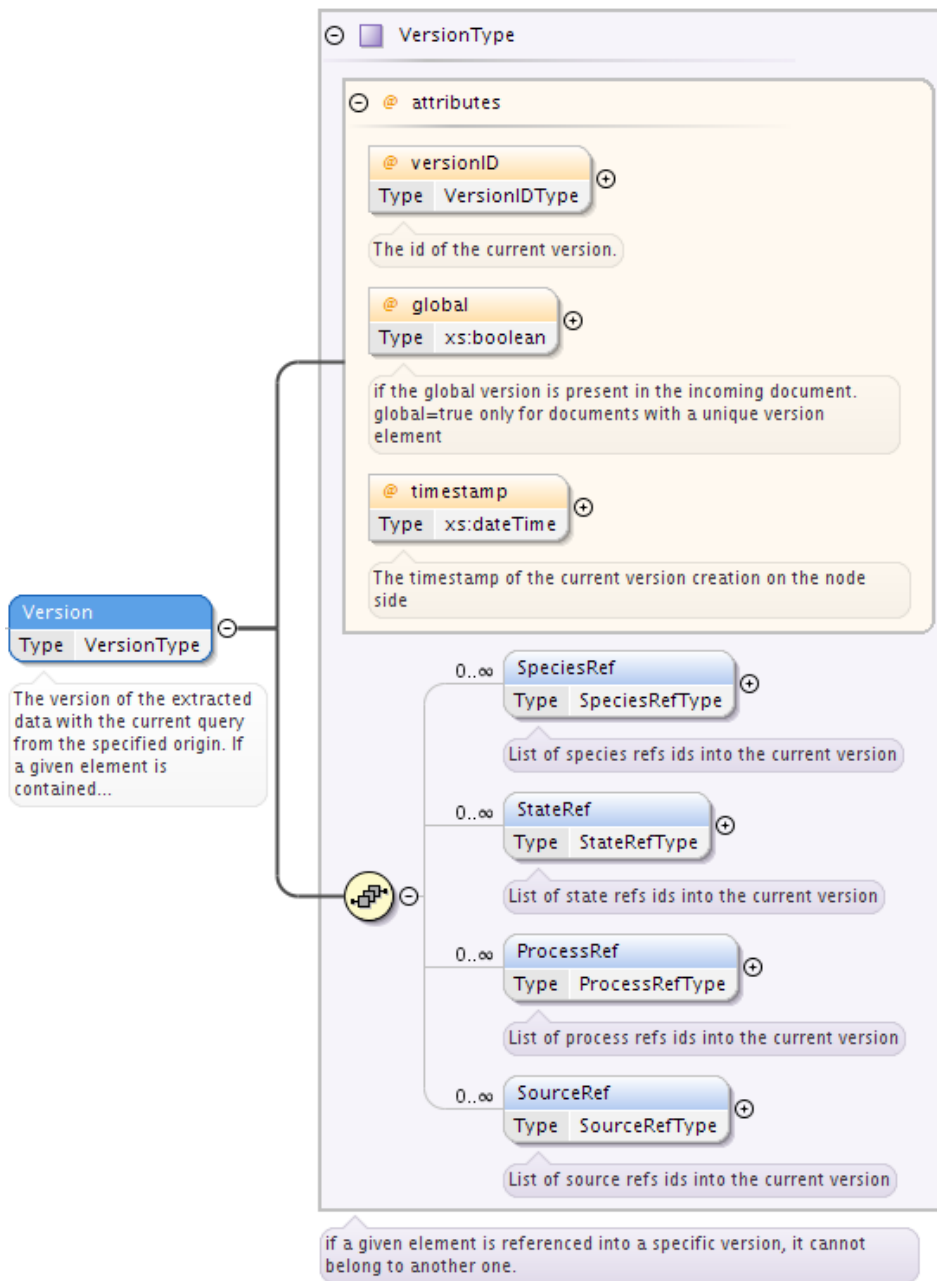


Figure 9: Representation of the element *VersionType*.

These three classes make sure that the users know what is the latest treatment of the data, and what are the original sources of the data, thus making possible to trace the original public exposure of the data.

### 3.1.1 The Version element

With the *VersionType* class the data contained in a XSAMS file are managed with fine grained granularity (cf. figure 9). This class contains three attributes:

- a *versionID* which is the unique identifier of the current version instance;
- a *global* boolean field. If *global* is true, then XSAMS file should consist of a unique instance of *VersionType* and all the data contained into the file will automatically belong to this Version. This implies that the attribute *global* equal to true is legal for XSAMS document with a unique *Origin* element (since *Origin* automatically contains a *Version*).
- a unique *timestamp* field which define the timestamp of the current version creation on the resources providing the data.

Moreover, this class contains three lists of elements:

- a list of *SpeciesRef* objects of type *SpeciesRefType*. In other words, this list contains the references to all the species (defined elsewhere into the XSAMS file) which belong to the current Version;
- a list of *StateRef* object of type *StateRefType*. This list contains the references to all the states (defined elsewhere into the XSAMS file) which belong to the current Version;
- a list of *ProcessRef* object of type *ProcessRefType*. This list contains the references to all the process (defined elsewhere into the XSAMS file) which belong to the current Version.
- a list of *SourceRef* object of type *SourceRefType*. This list contains the references to all the bibliography sources (defined elsewhere into the XSAMS file) which belong to the current version.

If a given element is referenced into a specific *Version* object, it can not belong to another one.

**Remark:** It is important to underline the difference between the *timestamp* field contained into an *Origin* object and the *timestamp* contained into a *Version* object. The first corresponds to the time when the data were extracted by the user (i.e. the data when the file was created) while the second corresponds to the time when that specific version of data was made available to community by the resource providing the data.

## 4 Some examples of data extracted from VAMDC, formatted with the proposed schema

In this paragraph we will provide examples of data extracted from the VAMDC infrastructure and formatted with the schema described in the previous section. For reason of clarity, we have lightened those examples, by manually selecting only a subset of the data really available into VAMDC. This will not affect the methods and the accuracy of the described approach.

The VAMDC internal convention is to use wavelength expressed in angstrom when querying radiative data from databases. This is imposed by the need of interoperability within VAMDC. Nevertheless from our portal ([8], <http://portal.vamdc.eu>) the users may query the radiative databases using whatever relevant quantities (frequency, energy, wavenumber, wavelength) associated to a choice of relevant units. No change in the original data is performed in the extracted output. For example B, line 108 shows that CDMS provides frequencies in MHz.

### 4.1 Example of an extraction from the BASECOL node

Let us consider an elastic collisional process, where the target is *CO* and the collider is *He*. Let us try to find some data on this collision into the VAMDC infrastructure. The associated query (generated by the portal from the Graphical User Interface and passed to all the VAMDC nodes) is:

```
select * where
((target.MoleculeStoichiometricFormula = 'CO')) AND
((collider.AtomSymbol = 'he'))
```

The file reported in its entirety in A is the output that the BASECOL node [4] generates when answering the previous query (with the reservations made at the beginning of paragraph 4). Let us browse the content of this file and focus on the *Origin* element:

```
1 <Origin xsi:type="VamdcNodeOriginType">
2 <Timestamp>2015-12-03T14:40:21+01:00</Timestamp>
3 <Version versionID="VER001" global="false"
4 timestamp="2015-09-01T08:10:12+01:00">
5 <SpeciesRef>XBAS2</SpeciesRef>
6 <SpeciesRef>XBAS52</SpeciesRef>
7 <StateRef>SBASET54-1</StateRef>
8 <StateRef>SBASET52-1</StateRef>
9 <StateRef>SBASET52-2</StateRef>
10 <ProcessRef>PBASC50t2T1c1C1</ProcessRef>
11 <SourceRef>BBAS0</SourceRef>
12 <SourceRef>BBAS849</SourceRef>
```

```

12     </Version>
13     <HomepageUrl>http://basecol.vamdc.org</HomepageUrl>
14     <Name>Basecol</Name>
15     <Query>select * where
16         ((target.MoleculeStoichiometricFormula = 'CO')) AND
17         ((collider.AtomSymbol = 'he'))</Query>
18     <OriginIdentifier>ivo://vamdc/basecol/vamdc-tap_12.07</OriginIdentifier>
19 </Origin>

```

The first line indicates that the resource producing the file is a VAMDC node and the second line indicates when the extraction has been performed (this time also corresponds to the creation of the file). On lines 3 to line 12 we find the information about the versioning of the contained data: it is specified that

- the species referenced by the identifiers *XBAS2*, *XBAS52* (defined respectively at lines 56 and 89 of A);
- the states referenced by the identifiers *SBASET54-1*, *SBASET52-1*, *SBASET52-2* (defined respectively at lines 58, 119 and 134 of A);
- the processes referenced by the identifier *PBASC50t2T1c1C1* (defined at line 154 of A)
- the sources referenced by the identifiers *BBAS0*, *BBAS849* (defined respectively at line 194 and 207 of A).

belong to the version *VER001* of the database. This version was made available on the date 2015-09-01T08:10:12+01:00.

Starting from line 13 we have the information about the database that produced the data: the link to its homepage (line 13), its Name (line 14), the query used for generating extracting the data and generating the file (lines 15 to 16), the unique identifier of the resource as it appears into the VAMDC infrastructure registry (line 17).

We recall that the two bibliographic references *BBAS0* and *BBAS849* included into this *VER001* version (cf. the last bullet of the previous item list) are respectively defined at line 194 of A:

```

1 <Source sourceID="BBAS0">
2     <Category>database</Category>
3     <SourceName>BASECOL database</SourceName>
4     <Year>2015</Year>
5     <Authors>
6         <Author>
7             <Name>M.-L. Dubernet</Name>
8         </Author>
9     </Authors>

```

```

10     <UniformResourceIdentifier>http://basecol.obspm.fr</UniformResourceIdentifier>
11     <ProductionDate>2015-12-03</ProductionDate>
12     <Comments>select * where
        ((target.MoleculeStoichiometricFormula = 'CO')) AND
        ((collider.AtomSymbol = 'he'))</Comments>
13     </Source>

```

and line 207 of A:

```

1  <Source sourceID="BBAS849">
2    <Category>journal</Category>
3    <SourceName>apj</SourceName>
4    <Year>2002</Year>
5    <Authors>
6      <Author>
7        <Name>N. Balakrishnan</Name>
8      </Author>
9      <Author>
10       <Name>A. Dalgarno</Name>
11     </Author>
12     <Author>
13       <Name>C. Cecchi-Pestellini</Name>
14     </Author>
15     <Author>
16       <Name>E. Bodo</Name>
17     </Author>
18   </Authors>
19   <Title>Rotational and Vibrational Excitation of CO
        Molecules by Collisions with  $^4\text{He}$  Atoms</Title>
20   <Volume>571</Volume>
21   <PageBegin>1015</PageBegin>
22   <PageEnd>1020</PageEnd>
23   <UniformResourceIdentifier>
24     http://adsabs.harvard.edu/cgi-bin/nph-bib_query?
25     bibcode=2002JChPh.116.4517K&db_key=PHY
26   </UniformResourceIdentifier>
27 </Source>

```

This last reference is [1]

## 4.2 Example of an extraction from the CDMS node

Let us now try to extract from the VAMDC infrastructure some spectroscopic data for the carbon monoxide in a range of wavelength between  $2.6006 \cdot 10^7 \text{ \AA}$  and  $2.6008 \cdot 10^7 \text{ \AA}$ . The associates query (typically generated by the portal from the Graphical User Web Interface and passed to all the VAMDC nodes) is

```

select * where
(RadTransWavelength >= 2.6006E7 AND

```



```
RadTransWavelength <= 2.6008E7) AND
((MoleculeStoichiometricFormula = 'CO'))
```

The file reported in its entirety in B is the output that the CDMS node [15] generates when answering the previous query (with the reservations made at the beginning of paragraph 4). Let us browse the content of this file and focus on the *Origin* element:

```
1 <Origin xsi:type="VamdcNodeOriginType">
2 <Timestamp>2015-12-03T15:50:21+01:00</Timestamp>
3 <Version versionID="VERCDMS1" global="false"
  timestamp="2000-10-01T12:00:00+01:00">
4 <SpeciesRef>XCDMS-83</SpeciesRef>
5 <StateRef>SCDMS-83-1</StateRef>
6 <StateRef>SCDMS-83-2</StateRef>
7 <StateRef>SCDMS-origin-83</StateRef>
8 <ProcessRef>PCDMS-R15140649</ProcessRef>
9 <SourceRef>BCDMS0</SourceRef>
10 <SourceRef>BCDMS-1921</SourceRef>
11 <SourceRef>BCDMS-1681</SourceRef>
12 </Version>
13 <HomepageUrl>http://cdms.ph1.uni-koeln.de/</HomepageUrl>
14 <Name>CDMS database</Name>
15 <Query>select * where (RadTransWavelength &gt;= 2.6006E7
  AND RadTransWavelength &lt;= 2.6008E7) AND
  ((MoleculeStoichiometricFormula = 'CO'))</Query>
16 <OriginIdentifier>ivo://vamdc/cdms/vamdc-tap_12.07</OriginIdentifier>
17 </Origin>
```

As in the example from BASECOL, the first line indicates that the resource producing the file is a VAMDC node and the second line indicates when the file was generated. From line 3 to 12 we find the information about the versioning of the contained data: it is specified that

- the species referenced by the identifier *XCDMS-83* (defined at line 24 of B);
- the states referenced by the identifiers *SCDMS-83-1*, *SCDMS-83-2*, *SCDMS-origin-83* (respectively defined at lines 67, 83, 51 of B);
- the process referenced by the identifier *PCDMS-R15140649* (defined at line 105 of B)
- the sources referenced by the identifier *BCDMS0*, *BCDMS-1921*, *BCDMS-1681* (defined respectively at line 131, 150 and 170 of B)

belong to the version *VER1* of the database. This version was made available on the date 2000-10-01T12:00:00+01:00.

Starting from line 13 we have the information about the database that produced the data: the link to its homepage (line 13), its name (line 14), the

query used for extracting the data and generating the file (line 13), the unique identifier of the resource, as it appears into the VAMDC infrastructure registry (line 15).

We recall that the three bibliographic references *BCDMS0*, *BCDMS-1921* and *BCDMS-1681* (cf. the last bullet of the previous item list) are respectively defined at line 131 of B

```
1 <Source sourceID="BCDMS0">
2   <Comments>This Source is a self-reference.
3   It represents the database and the query that produced the
4   xml document.
5   The sourceID contains a timestamp.
6   The full URL is given in the tag UniformResourceIdentifier
7   but you need
8   to unescape ampersands and angle brackets to re-use it.
9   Query was: select * where
10  ((target.MoleculeStoichiometricFormula = 'CO')) AND
11  ((collider.AtomSymbol = 'he'))</Comments>
12  <Year>2015</Year>
13  <Category>database</Category>
14  <UniformResourceIdentifier>http://cdms.ph1.uni-koeln.de/cdms/tap/sync?
15  LANG=VSS2&
16  REQUEST=doQuery&FORMAT=XSAMS&QUERY= select *
17  where ((target.MoleculeStoichiometricFormula = 'CO'))
18  AND ((collider.AtomSymbol = 'he'))</
19  UniformResourceIdentifier>
20  <ProductionDate>2015-12-03</ProductionDate>
21  <Authors>
22    <Author>
23      <Name>N.N.</Name>
24    </Author>
25  </Authors>
26</Source>
```

at line 150 of B

```
1 <Source sourceID="BCDMS-1921">
2   <Authors>
3     <Author>
4       <Name>Mller, H. S. P.</Name>
5     </Author>
6     <Author>
7       <Name>Endres, C. P.</Name>
8     </Author>
9     <Author>
10      <Name>Schlemmer, S.</Name>
11    </Author>
12    <Author>
13      <Name>Stutzki, J.</Name>
```

```

14         </Author>
15     </Authors>
16     <Title />
17     <Category>database</Category>
18     <Year>2012</Year>
19     <SourceName>CDMS database</SourceName>
20 </Source>

```

and line 170 of B

```

1 <Source sourceID="BCDMS-1681">
2     <Authors>
3         <Author>
4             <Name>Winnewisser, G.</Name>
5         </Author>
6         <Author>
7             <Name>Belov, S. P.</Name>
8         </Author>
9         <Author>
10            <Name>Klaus, T.</Name>
11        </Author>
12        <Author>
13            <Name>Schieder, R.</Name>
14        </Author>
15    </Authors>
16    <Title />
17    <Category>journal</Category>
18    <Year>1997</Year>
19    <SourceName>J. Mol. Spectrosc.</SourceName>
20    <Volume>184</Volume>
21    <PageBegin>468</PageBegin>
22    <DigitalObjectIdentifier>
23        10.1006/jmsp.1997.7341
24    </DigitalObjectIdentifier>
25 </Source>

```

This last reference is [26].

### 4.3 Documenting data processing workflow: merging of two files with SpectCol

We used the SpectCol tool [3],[7]<sup>2</sup> for cross-matching the data from the two files described into the previous subparagraphs. The SpectCol tool allows to create combined spectroscopic and collisional data files where the collisional rate coefficients come from a collisional database (BASECOL in this example) and the spectroscopic data (energy levels and Einstein coefficients) come from a spectroscopic database (CDMS in this example). This process requires to match the spectroscopic energy levels of CDMS with those of

<sup>2</sup><http://www.vamdc.org/software/spectcol>

BASECOL on the basis of identical quantum numbers. We asked Spectcol to merge the two files by identifying the states of the Carbon Monoxide characterised by the same quantum number values  $J$ : the state referenced by the identifier *SBASET52-1* ( $J = 0$ , cf. line 117 of A) has been identified and assimilated with *CDMS-83-1* (cf. line 64 of B) while the state referenced by the identifier *SBASET52-2* ( $J = 1$ , cf. line 132 of A) has been identified and assimilated with *CDMS-83-1* (cf. line 80 of B).

The resulting merged file is reported in C. Let us browse the content of this file and focus on the root *Origin* element:

```

1  <Origin xsi:type="OtherOriginType">
2  <Timestamp>2015-12-07T15:50:21+01:00</Timestamp>
3  <Version versionID="VERMER1" global="false"
   timestamp="2000-10-01T12:00:00+01:00">
4  <ProcessRef>PBASC50t2T1c1C1</ProcessRef>
5  <SourceRef>BBAS0</SourceRef>
6  <SourceRef>BBAS849</SourceRef>
7  <SourceRef>BCDMS0</SourceRef>
8  <SourceRef>BCDMS-1921</SourceRef>
9  <SourceRef>BCDMS-1681</SourceRef>
10 </Version>
11 <HomepageUrl>
12   http://www.vamdc.org/activities/research/software/spectcol/
13 </HomepageUrl>
14 <Name>Spectcol</Name>
15 <Comments>Merging between collisional data from Basecol and
   spectroscopic data from CDMS by cross matching state with
   same J values in CO</Comments>
16
17 <Origin xsi:type="VamdcNodeOriginType">
18 <Timestamp>2015-12-03T15:50:21+01:00</Timestamp>
19 <Version versionID="VERCDMS1" global="false"
   timestamp="2000-10-01T12:00:00+01:00">
20 <SpeciesRef>XCDMS-83</SpeciesRef>
21 <StateRef>SCDMS-83-1</StateRef>
22 <StateRef>SCDMS-83-2</StateRef>
23 <StateRef>SCDMS-origin-83</StateRef>
24 <ProcessRef>PCDMS-R15140649</ProcessRef>
25 </Version>
26 <HomepageUrl>http://cdms.ph1.uni-koeln.de/</HomepageUrl>
27 <Name>CDMS database</Name>
28 <Comments>Spectroscopic data extracted from CDMS VAMDC
   node.</Comments>
29 <Query>select * where (RadTransWavelength >= 2.6006E7
   AND RadTransWavelength <= 2.6008E7) AND
   ((MoleculeStoichiometricFormula = 'CO'))</Query>
30 <OriginIdentifier>ivo://vamdc/cdms/vamdc-tap_12.07</OriginIdentifier>
31 </Origin>
32

```

```

33 <Origin xsi:type="VamdcNodeOriginType">
34   <Timestamp>2015-12-03T14:40:21+01:00</Timestamp>
35   <Version versionID="VER001" global="false"
      timestamp="2015-09-01T08:10:12+01:00">
36     <SpeciesRef>XBAS2</SpeciesRef>
37     <StateRef>SBASET54-1</StateRef>
38   </Version>
39   <HomepageUrl>http://basecol.vamdc.org</HomepageUrl>
40   <Name>Basecol</Name>
41   <Comments>Collisional data extracted from Basecol VAMDC
      node.</Comments>
42   <Query>select * where
      ((target.MoleculeStoichiometricFormula = 'CO')) AND
      ((collider.AtomSymbol = 'he'))</Query>
43   <OriginIdentifier>ivo://vamdc/basecol/vamdc-tap_12.07</OriginIdentifier>
44 </Origin>
45
46 </Origin>

```

We see (line 1) that the file was not produced by a VAMDC node, but by the Spectcol tool (line 11 to 15). The current file has been produced on 2015-12-07T15:50:21+01:00 and results from the merging of data coming from BASECOL and from CDMS, where the states of *CO* characterised by the same values of the quantum number *J* are crossmatched (line 15). The data coming from BASECOL and CDMS are detailed into the two contained *Origin* sub-elements:

- the first one is defined from line 17 to line 31 and holds the data coming from the CDMS node. This includes the species referenced by the identifier *XCDMS-83*, the states referenced by the identifier *SCDMS-83-1*, *SCDMS-83-2*, *SCDMS-origin-83* (respectively defined at line 137, 151, 165 of C), the process referenced by the identifier *PCDMS-R15140649* (defined at line 186 of C). The extraction of those data was performed on 2015-12-03T15:50:21+01:00 from the CDMS Node by submitting the query

```

select * where
(RadTransWavelength >= 2.6006E7 AND
RadTransWavelength <= 2.6008E7) AND
((MoleculeStoichiometricFormula = 'CO'))

```

The extracted data correspond to the version *VERCDMS1* that was available to the community since 2000-10-01T12:00:00+01:00

- the second one is defined from line 31 to line 42 and holds the data coming from the BASECOL node. This includes the species referenced by the identifier *XBAS2* (defined at line 75 of C) and the state referenced by the identifier *SBASET54-1* (defined at line 77 of C). The extraction of those data was performed on 2015-12-03T14:40:21+01:00

from the BASECOL node by submitting the query

```
select * where
((target.MoleculeStoichiometricFormula = 'CO')) AND
((collider.AtomSymbol = 'he'))
```

The extracted data correspond to the version *VER001* of the database.

This was available to the community since 2015-09-01T08:10:12+01:00.

Let us focus again on the root *Origin* element:

- As we can see from line 5 to 9, all the sources references from the merged files are now contained into the *Version* object of the root *Origin*. Indeed during the merging procedure, the sources from both the BASECOL and the CDMS files are compiled together into the root element since the data coming from all the sources are used. Moreover for data citation, it is very convenient to store at a same location all the references linked to the current file.
- The process referenced by *PBASC50t2T1c1C1* belongs to the *Version* field of the root element (line 4). If we compare with A, line 154, we discover that this is a collisional process coming from BASECOL. It is important to understand why it is not contained into the second sub-origin element which, as we previously shown, contains the data coming from BASECOL: in its original extraction from BASECOL (cf. A, line 154) this process is a collision between *CO* (identified with the label *XBAS52*) and *He* (identified with the label *XBAS2*). During the collision, the *CO* pass from the state referenced by *SBASET52-2* to the state referenced by *SBASET52-1*. All these identifiers and references have consistency in the BASECOL perimeter. Let us look at the same collisional process as it is described on the merged file (cf. line 212 of C): the collision is between *CO* (now identified with the label *XCDMS-83*, since during the crossmatching procedure we have replaced the *CO* from BASECOL with the one coming from CDMS) and *He*. During the collision *CO* pass from the state referenced by *SCDMS-83-2* to the state referenced by *SCDMS-83-1* (indeed we remember that during the crossmatching we have identified the state *SBASET52-2* from BASECOL with the state *SCDMS-83-2* from CDMS and the states *SBASET52-1* from BASECOL with the state *SCDMS-83-1* from CDMS). Due to these changes, the collisional process refer now to data coming both from BASECOL and CDMS and the data is no more the original one freshly extracted from BASECOL (as it was the case in A, line 154). As a direct consequence BASECOL is no more the source origin of this data. This combined collision is created by SpectCol and it is natural that it appears into the root *Origin* element which holds the data produced by SpectCol during the crossmatching procedure.

## 5 Practical use of the new XSAMS features

In section 3 we introduced new changes and additions to the XSAMS format which allows one to identify unequivocally through time and location the datasets extracted from VAMDC: all the relevant metadata (the extraction query, the date when the query was submitted to the infrastructure, the responding node, the version of the responding node, etc...) are automatically embedded into the output file. But from the practical point of view of the final user, how to use this information ? Our aim is to simplify the “data citation” landscape and not to complicate the actual one described in section 2. In this paragraph we are going to sketch the future VAMDC services that will allow a simple and fluid use of these new features:

When a user queries data from the VAMDC e-infrastructure by submitting a given query, he/she will receive an output file formatted as described in section 3. This file will be associated with a unique digital identifier. This identifier (as it is the case for the DOI), will be resolvable and the landing page<sup>3</sup> will summarize the information contained in the *Origin* field of the output file, the data-file itself (via query re-execution, if the node supports this feature)<sup>4</sup>, the bibtex references of all the papers used for compiling the output file. This sketched service is a part of a *Query Store* as it is understood in the Data Citation RDA context and described in [22].

This query store may be used in collaboration with editors. We imagine the following data-citation workflow: when a scientist wishes to cite a dataset extracted from VAMDC, he/she has just to put into the references the *unique identifier* returned by the VAMDC infrastructure during the extraction procedure. Since this unique identifier is resolvable, the editors may resolve it and, by accessing the bibtex information of the landing page, give credits to all the listed authors. This automatic credit delegation method may also work with current systems (e.g. the SAO/NASA Astrophysics Data System<sup>5</sup>).

---

<sup>3</sup>The landing page will be human readable, but a machine actionable version may exist for automatic treatments

<sup>4</sup>The evolved XSAMS standard described in section 3 allows a data provider to univocally tag version of their data. Even if we strongly suggest that the VAMDC nodes give access (through the VAMC infrastructure) to all the subsequent versions of their data, we cannot enforce the data policy on the federated nodes. A given node may delete an old version when a new one is available. In that case the metadata policy will always allow the unambiguous identification of the extracted data, but data extraction re-execution will not be possible

<sup>5</sup><http://adswww.harvard.edu/>

## 6 Concluding remarks

Through this work we have discussed an evolution of the VAMDC standard output format which provides a mechanisms allowing to identify through time and location, in a sustainable and unique way, data sets extracted by users from the VAMDC e-infrastructure. With our proposed solution the information about the origin and version of data are embedded into (and indissoluble from) the file containing the atomic and molecular data. The same mechanisms define an unambiguous method for tracing and documenting the data processing workflows.

The methods and technical solutions introduced in this paper are designed for scientist wishing to efficiently cite the datasets and the data-producers of the Atomic and Molecular data extracted from the VAMDC infrastructure:

let us consider an XSAMS data file containing  $N$  references ( $N \gg 1$ ) to the sources used for the data file compilation.

Before the introduction of the solution discussed into this paper, a scientist wishing to publish a work based on that data file would need to include the  $N$  publications in order to properly cite the data.

With the introduction of the proposed solution, the scientist will only cite the Digital Unique Identifier (DUI) returned by the VAMDC infrastructure during the extraction procedure. This DUI is resolvable and the landing page will contain the original query used for the generation of the data file and the bibliographic references to the  $N$  publications.

## 7 Acknowledgements

Support for VAMDC has been provided through the VAMDC and the SUP@VAMDC projects funded under the “Combination of Collaborative Projects and Coordination and Support Actions” Funding Scheme of The Seventh Framework Program. Call topic: INFRA-2008-1.2.2 and INFRA-2012 Scientific Data Infrastructure. Grant Agreement numbers: 239108 and 313284. We acknowledge support from Paris Astronomical Data Center. **We thank Thomas Marquart, Guy Rixon and Mickail Doronin from the VAMDC Consortium for fruitful discussions.**

## A Sample output from BASECOL

```
1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <XSAMSData xmlns="http://vamdc.org/xml/xsams/1.0"
3   xmlns:cml="http://www.xml-cml.org/schema"
4   xmlns:asymcs="http://vamdc.org/xml/xsams/1.0/cases/asymcs"
```



```

5  xmlns:asymos="http://vamdc.org/xml/xsams/1.0/cases/asymos"
6  xmlns:dcs="http://vamdc.org/xml/xsams/1.0/cases/dcs"
7  xmlns:gen="http://vamdc.org/xml/xsams/1.0/cases/gen"
8  xmlns:hunda="http://vamdc.org/xml/xsams/1.0/cases/hunda"
9  xmlns:hundb="http://vamdc.org/xml/xsams/1.0/cases/hundb"
10 xmlns:lpcs="http://vamdc.org/xml/xsams/1.0/cases/lpcs"
11 xmlns:lpos="http://vamdc.org/xml/xsams/1.0/cases/lpos"
12 xmlns:ltcs="http://vamdc.org/xml/xsams/1.0/cases/ltcs"
13 xmlns:ltos="http://vamdc.org/xml/xsams/1.0/cases/ltos"
14 xmlns:nltcs="http://vamdc.org/xml/xsams/1.0/cases/nltcs"
15 xmlns:nltos="http://vamdc.org/xml/xsams/1.0/cases/nltos"
16 xmlns:sphcs="http://vamdc.org/xml/xsams/1.0/cases/sphcs"
17 xmlns:sphos="http://vamdc.org/xml/xsams/1.0/cases/sphos"
18 xmlns:stcs="http://vamdc.org/xml/xsams/1.0/cases/stcs"
19 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
20 xsi:schemaLocation="http://vamdc.org/xml/xsams/1.0
21     https://raw.githubusercontent.com/nicolasmoreau/
22     VAMDC-XSAMS/abstract_origin/xsams.xsd">
23
24 <Origin xsi:type="VamdcNodeOriginType">
25   <Timestamp>2015-12-03T14:40:21+01:00</Timestamp>
26   <Version versionID="VER001" global="false"
27     timestamp="2015-09-01T08:10:12+01:00">
28     <SpeciesRef>XBAS2</SpeciesRef>
29     <SpeciesRef>XBAS52</SpeciesRef>
30     <StateRef>SBASET54-1</StateRef>
31     <StateRef>SBASET52-1</StateRef>
32     <StateRef>SBASET52-2</StateRef>
33     <ProcessRef>PBASC50t2T1c1C1</ProcessRef>
34     <SourceRef>BBAS0</SourceRef>
35     <SourceRef>BBAS849</SourceRef>
36   </Version>
37   <HomepageUrl>http://basecol.vamdc.org</HomepageUrl>
38   <Name>Basecol</Name>
39   <Query>select * where
40     ((target.MoleculeStoichiometricFormula = 'CO')) AND
41     ((collider.AtomSymbol = 'he'))</Query>
42   <OriginIdentifier>ivo://vamdc/basecol/vamdc-tap_12.07</OriginIdentifier>
43 </Origin>
44 <Species>
45   <Atoms>
46     <Atom>
47       <ChemicalElement>
48         <NuclearCharge>2</NuclearCharge>
49         <ElementSymbol>He</ElementSymbol>
50       </ChemicalElement>
51       <Isotope>
52         <IsotopeParameters>
53           <MassNumber>4</MassNumber>

```

```

52     <Mass>
53         <Value units="amu">4.0026</Value>
54     </Mass>
55 </IsotopeParameters>
56 <Ion speciesID="XBAS2">
57     <IonCharge>0</IonCharge>
58     <AtomicState stateID="SBASET54-1">
59         <Comments>Energy level of He (no
        structure)</Comments>
60         <SourceRef>BBAS0</SourceRef>
61         <AtomicNumericalData>
62             <StateEnergy>
63                 <Value units="1/cm">0.0</Value>
64             </StateEnergy>
65         </AtomicNumericalData>
66         <AtomicQuantumNumbers>
67             <TotalAngularMomentum>0.0</TotalAngularMomentum>
68         </AtomicQuantumNumbers>
69         <AtomicComposition>
70             <Component>
71                 <Term>
72                     <LS>
73                         <L>
74                             <Value>0</Value>
75                             <Symbol>L</Symbol>
76                         </L>
77                     <S>0.0</S>
78                 </LS>
79             </Term>
80         </Component>
81     </AtomicComposition>
82 </AtomicState>
83 <InChIKey>SWQJXJQGLNCZEY-UHFFFAOYSA-N</InChIKey>
84 </Ion>
85 </Isotope>
86 </Atom>
87 </Atoms>
88 <Molecules>
89     <Molecule speciesID="XBAS52">
90         <MolecularChemicalSpecies>
91             <OrdinaryStructuralFormula>
92                 <Value>CO</Value>
93             </OrdinaryStructuralFormula>
94             <StoichiometricFormula>CO</StoichiometricFormula>
95             <ChemicalName>
96                 <Value>CO</Value>
97             </ChemicalName>
98             <InChI>InChI=1S/CO/c1-2</InChI>
99             <InChIKey>UGFAIRIUMAVXCW-UHFFFAOYSA-N</InChIKey>

```

```

100 <VAMDCSpeciesID>UGFAIRIUMAVXCW-UHFFFAOYSA-N</VAMDCSpeciesID>
101 <MoleculeStructure>
102 <cml:atomArray>
103 <cml:atom hydrogenCount="0" isotopeNumber="16"
count="1.0"
104 formalCharge="1" elementType="O" id="RBAS43N357"/>
105 <cml:atom hydrogenCount="0" isotopeNumber="12"
count="1.0"
106 formalCharge="-1" elementType="C" id="RBAS43N358"/>
107 </cml:atomArray>
108 <cml:bondArray>
109 <cml:bond order="T" atomRefs2="RBAS43N357
RBAS43N358"/>
110 </cml:bondArray>
111 </MoleculeStructure>
112 <StableMolecularProperties>
113 <MolecularWeight>
114 <Value units="amu">27.99</Value>
115 </MolecularWeight>
116 </StableMolecularProperties>
117 <Comment>Theoretical rotational energy levels of CO
(Cecchi-Pestellini, 2002)</Comment>
118 </MolecularChemicalSpecies>
119 <MolecularState stateID="SBASET52-1">
120 <SourceRef>BBAS0</SourceRef>
121 <Description>Theoretical rotational energy levels of
CO (Cecchi-Pestellini, 2002)</Description>
122 <MolecularStateCharacterisation>
123 <StateEnergy energyOrigin="SBASET52-1">
124 <Value units="1/cm">0.0</Value>
125 </StateEnergy>
126 </MolecularStateCharacterisation>
127 <Case
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:type="dcs:Case" caseID="dcs">
128 <dcs:QNs>
129 <dcs:ElecStateLabel>X</dcs:ElecStateLabel>
130 <dcs:J>0</dcs:J>
131 </dcs:QNs>
132 </Case>
133 </MolecularState>
134 <MolecularState stateID="SBASET52-2">
135 <SourceRef>BBAS0</SourceRef>
136 <Description>Theoretical rotational energy levels of
CO (Cecchi-Pestellini, 2002)</Description>
137 <MolecularStateCharacterisation>
138 <StateEnergy energyOrigin="SBASET52-1">
139 <Value units="1/cm">3.85</Value>
140 </StateEnergy>

```

```

141     </MolecularStateCharacterisation>
142     <Case
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="dcs:Case" caseID="dcs">
143         <dcs:QNs>
144             <dcs:ElecStateLabel>X</dcs:ElecStateLabel>
145             <dcs:J>1</dcs:J>
146         </dcs:QNs>
147     </Case>
148 </MolecularState>
149 </Molecule>
150 </Molecules>
151 </Species>
152 <Processes>
153     <Collisions>
154         <CollisionalTransition id="PBASC50t2T1c1C1">
155             <Comments> Rotational de-excitation of CO (v=0) by He
                (Cecchi-Pestellini & al, 2002)</Comments>
156             <SourceRef>BBAS0</SourceRef>
157             <SourceRef>BBAS849</SourceRef>
158             <ProcessClass>
159                 <Code>inel</Code>
160             </ProcessClass>
161             <Reactant>
162                 <SpeciesRef>XBAS52</SpeciesRef>
163                 <StateRef>SBASET52-2</StateRef>
164             </Reactant>
165             <Reactant>
166                 <SpeciesRef>XBAS2</SpeciesRef>
167                 <StateRef>SBASET54-1</StateRef>
168             </Reactant>
169             <Product>
170                 <SpeciesRef>XBAS52</SpeciesRef>
171                 <StateRef>SBASET52-1</StateRef>
172             </Product>
173             <Product>
174                 <SpeciesRef>XBAS2</SpeciesRef>
175                 <StateRef>SBASET54-1</StateRef>
176             </Product>
177             <DataSets>
178                 <DataSet dataDescription="rateCoefficient">
179                     <TabulatedData>
180                         <Comments>Rate coefficients</Comments>
181                         <X units="K">
182                             <DataList>5.0 10.0 20.0 40.0 60.0 80.0 100.0
                                    200.0 300.0 500.0</DataList>
183                         </X>
184                         <Y units="cm3/s">

```

```

185         <DataList>3.4E-11 3.2E-11 3.0E-11 2.8E-11
           2.7E-11 2.6E-11 2.6E-11 2.5E-11 2.5E-11
           2.6E-11</DataList>
186     </Y>
187 </TabulatedData>
188 </DataSet>
189 </DataSets>
190 </CollisionalTransition>
191 </Collisions>
192 </Processes>
193 <Sources>
194   <Source sourceID="BBAS0">
195     <Category>database</Category>
196     <SourceName>BASECOL database</SourceName>
197     <Year>2015</Year>
198     <Authors>
199       <Author>
200         <Name>M.-L. Dubernet</Name>
201       </Author>
202     </Authors>
203     <UniformResourceIdentifier>http://basecol.obspm.fr</UniformResourceIdentifier>
204     <ProductionDate>2015-12-03</ProductionDate>
205     <Comments>select * where
           ((target.MoleculeStoichiometricFormula = 'CO')) AND
           ((collider.AtomSymbol = 'he'))</Comments>
206   </Source>
207   <Source sourceID="BBAS849">
208     <Category>journal</Category>
209     <SourceName>apj</SourceName>
210     <Year>2002</Year>
211     <Authors>
212       <Author>
213         <Name>N. Balakrishnan</Name>
214       </Author>
215       <Author>
216         <Name>A. Dalgarno</Name>
217       </Author>
218       <Author>
219         <Name>C. Cecchi-Pestellini</Name>
220       </Author>
221       <Author>
222         <Name>E. Bodo</Name>
223       </Author>
224     </Authors>
225     <Title>Rotational and Vibrational Excitation of CO
           Molecules by Collisions with  ${}^4\text{He}$  Atoms</Title>
226     <Volume>571</Volume>
227     <PageBegin>1015</PageBegin>
228     <PageEnd>1020</PageEnd>

```

```

229     <UniformResourceIdentifier>
230     http://adsabs.harvard.edu/cgi-bin/nph-bib_query?
231     bibcode=2002JChPh.116.4517K&db_key=PHY
232     </UniformResourceIdentifier>
233 </Source>
234 </Sources>
235 </XSAMSData>

```

## B Sample output from CDMS

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <XSAMSData xmlns="http://vamdc.org/xml/xsams/1.0"
3   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4   xmlns:cml="http://www.xml-cml.org/schema"
5   xsi:schemaLocation="http://vamdc.org/xml/xsams/1.0
6     https://raw.githubusercontent.com/nicolasmoreau/VAMDC-XSAMS/abstract_origin/xsams.xsd">
7   <Origin xsi:type="VamdcNodeOriginType">
8     <Timestamp>2015-12-03T15:50:21+01:00</Timestamp>
9     <Version versionID="VERCDMS1" global="false"
10    timestamp="2000-10-01T12:00:00+01:00">
11     <SpeciesRef>XCDMS-83</SpeciesRef>
12     <StateRef>SCDMS-83-1</StateRef>
13     <StateRef>SCDMS-83-2</StateRef>
14     <StateRef>SCDMS-origin-83</StateRef>
15     <ProcessRef>PCDMS-R15140649</ProcessRef>
16     <SourceRef>BCDMS0</SourceRef>
17     <SourceRef>BCDMS-1921</SourceRef>
18     <SourceRef>BCDMS-1681</SourceRef>
19   </Version>
20   <HomepageUrl>http://cdms.ph1.uni-koeln.de/</HomepageUrl>
21   <Name>CDMS database</Name>
22   <Query>select * where (RadTransWavelength >= 2.6006E7
23     AND RadTransWavelength <= 2.6008E7) AND
24     ((MoleculeStoichiometricFormula = 'CO'))</Query>
25   <OriginIdentifier>ivo://vamdc/cdms/vamdc-tap_12.07</OriginIdentifier>
26 </Origin>
27 <Species>
28   <Molecules>
29     <Molecule speciesID="XCDMS-83">
30     <MolecularChemicalSpecies>
31     <OrdinaryStructuralFormula>
32     <Value>CO</Value>
33     </OrdinaryStructuralFormula>
34     <StoichiometricFormula>CO</StoichiometricFormula>
35     <ChemicalName>
36     <Value>Carbon Monoxide</Value>
37     </ChemicalName>
38     <InChI>1S/CO/c1-2</InChI>
39     <InChIKey>UGFAIRIUMAVXCW-UHFFFAOYSA-N</InChIKey>

```

```

35 <VAMDCSpeciesID>UGFAIRIUMAVXCW-UHFFFAOYSA-N</VAMDCSpeciesID>
36 <PartitionFunction>
37   <T units="K">
38     <DataList>1.072 1.148 1.23 1.318 1.413 1.514 1.622
      1.738 1.862 1.995 2.138 2.291 2.455 2.63 2.725
      2.818 3.02 3.236 3.467 3.715 3.981 4.266 4.571
      4.898 5.0 5.248 5.623 6.026 6.457 6.918 7.413
      7.943 8.511 9.12 9.375 9.772 10.471 11.22 12.023
      12.882 13.804 14.791 15.849 16.982 18.197 18.75
      19.498 20.893 22.387 23.988 25.704 27.542 29.512
      31.623 33.884 36.308 37.5 38.905 41.687 44.668
      47.863 51.286 54.954 58.884 63.096 67.608 72.444
      75.0 77.625 83.176 89.125 95.499 102.329 109.648
      117.49 125.893 134.896 144.544 150.0 154.882
      165.959 177.828 190.546 204.174 218.776 225.0
      234.423 251.189 269.153 288.403 300.0 309.03
      331.131 354.813 380.189 407.38 436.516 467.735
      500.0 501.187 537.032 575.44 616.595 660.693
      707.946 758.578 812.831 870.964 933.254
      1000.0</DataList>
39   </T>
40   <Q>
41     <DataList>1.01721581835 1.02422837677
      1.03341102522 1.04512141245 1.05985020123
      1.07774884294 1.09923417933 1.12473229226
      1.15441650589 1.18864057708 1.22774305695
      1.27174864212 1.32092124733 1.37520246158
      1.40530654907 1.43514349042 1.50100121324
      1.57270208471 1.65050158216 1.73501239088
      1.8265276783 1.92535019369 2.03179140659
      2.14652192999 2.18241666936 2.26987115366
      2.40252640741 2.54553752733 2.69889427601
      2.8632973732 3.04016627563 3.22985295285
      3.4334258678 3.65195638949 3.74352703447
      3.88615863094 4.13746641418 4.40695607053
      4.69606494313 5.00551115749 5.33781506844
      5.69369660756 6.0753182363 6.48412210856
      6.9226335574 7.12225433182 7.3922960322
      7.89599749116 8.4355434648 9.0138233311
      9.633727213 10.2977845092 11.0096090065
      11.7724538098 12.5895724935 13.4656648032
      13.8965040828 14.4043468795 15.4099582553
      16.4875620501 17.6425836137 18.8800875183
      20.2062236736 21.6271428763 23.1500815655
      24.7815541073 26.5302455525 27.4545173427
      28.40375754 30.4111396059 32.5625279126
      34.8676989471 37.3378779761 39.9850161394
      42.8214291179 45.8608826253 49.1175078254
      52.6076105394 54.5813579159 56.3475015965

```

```

60.3549445917 64.6491571735 69.2508130864
74.1820459087 79.4660945815 81.7185083304
85.1287658147 91.1970158766 97.6997287418
104.669260558 108.868854329 112.139428854
120.146852376 128.732781329 137.942190312
147.825233889 158.438549717 169.84604324
181.685035016 182.121688241 195.35029188
209.62906021 225.070653126 241.803934912
259.976802963 279.755720344 301.328756805
324.90561697 350.71803967 379.021079637</DataList>
42 </Q>
43 </PartitionFunction>
44 <StableMolecularProperties>
45 <MolecularWeight>
46 <Value units="unitless">28</Value>
47 </MolecularWeight>
48 </StableMolecularProperties>
49 <Comment> 28503- v 1:CO; $v=0$</Comment>
50 </MolecularChemicalSpecies>
51 <MolecularState auxillary="true"
stateID="SCDMS-origin-83">
52 <MolecularStateCharacterisation>
53 <StateEnergy energyOrigin="SCDMS-origin-83">
54 <Value units="1/cm">0.0</Value>
55 </StateEnergy>
56 <TotalStatisticalWeight>1</TotalStatisticalWeight>
57 <NuclearStatisticalWeight>1</NuclearStatisticalWeight>
58 </MolecularStateCharacterisation>
59 <Case
xmlns:dcs="http://vamdc.org/xml/xsams/1.0/cases/dcs"
xsi:type="dcs:Case" caseID="dcs"
xsi:schemaLocation="http://vamdc.org/xml/xsams/1.0/cases/dcs
../../cases/dcs.xsd">
60 <dcs:QNs>
61 <dcs:ElecStateLabel>X</dcs:ElecStateLabel>
62 <dcs:v>0</dcs:v>
63 <dcs:J>0</dcs:J>
64 </dcs:QNs>
65 </Case>
66 </MolecularState>
67 <MolecularState stateID="SCDMS-83-1">
68 <MolecularStateCharacterisation>
69 <StateEnergy energyOrigin="SCDMS-origin-83">
70 <Value units="1/cm">0.0</Value>
71 </StateEnergy>
72 <TotalStatisticalWeight>1</TotalStatisticalWeight>
73 <NuclearStatisticalWeight>1</NuclearStatisticalWeight>
74 </MolecularStateCharacterisation>

```



```

75     <Case
       xmlns:dcs="http://vamdc.org/xml/xsams/1.0/cases/dcs"
       xsi:type="dcs:Case" caseID="dcs"
       xsi:schemaLocation="http://vamdc.org/xml/xsams/1.0/cases/dcs
       ../../cases/dcs.xsd">
76         <dcs:QNs>
77             <dcs:ElecStateLabel>X</dcs:ElecStateLabel>
78             <dcs:v>0</dcs:v>
79             <dcs:J>0</dcs:J>
80         </dcs:QNs>
81     </Case>
82 </MolecularState>
83 <MolecularState stateID="SCDMS-83-2">
84     <MolecularStateCharacterisation>
85         <StateEnergy energyOrigin="SCDMS-origin-83">
86             <Value units="1/cm">3.845033</Value>
87         </StateEnergy>
88         <TotalStatisticalWeight>3</TotalStatisticalWeight>
89         <NuclearStatisticalWeight>1</NuclearStatisticalWeight>
90     </MolecularStateCharacterisation>
91     <Case
       xmlns:dcs="http://vamdc.org/xml/xsams/1.0/cases/dcs"
       xsi:type="dcs:Case" caseID="dcs"
       xsi:schemaLocation="http://vamdc.org/xml/xsams/1.0/cases/dcs
       ../../cases/dcs.xsd">
92         <dcs:QNs>
93             <dcs:ElecStateLabel>X</dcs:ElecStateLabel>
94             <dcs:v>0</dcs:v>
95             <dcs:J>1</dcs:J>
96         </dcs:QNs>
97     </Case>
98 </MolecularState>
99 </Molecule>
100 </Molecules>
101 </Species>
102 <Processes>
103     <Radiative>
104         <RadiativeTransition id="PCDMS-R15140649"
           process="excitation">
105             <SourceRef>BCDMS0</SourceRef>
106             <EnergyWavelength>
107                 <Frequency>
108                     <Value units="MHz">115271.2021</Value>
109                     <Accuracy>0.0001</Accuracy>
110                 </Frequency>
111             </EnergyWavelength>
112             <UpperStateRef>SCDMS-83-2</UpperStateRef>
113             <LowerStateRef>SCDMS-83-1</LowerStateRef>
114             <SpeciesRef>XCDMS-83</SpeciesRef>

```

```

115     <Probability>
116     <TransitionProbabilityA>
117     <Value units="1/cm">7.20360334988e-08</Value>
118     </TransitionProbabilityA>
119     <IdealisedIntensity>
120     <Value units="unitless">-5.0105</Value>
121     </IdealisedIntensity>
122     <Multipole>E2</Multipole>
123     </Probability>
124     <ProcessClass>
125     <Code>rota</Code>
126     </ProcessClass>
127     </RadiativeTransition>
128 </Radiative>
129 </Processes>
130 <Sources>
131     <Source sourceID="BCDMS0">
132     <Comments>This Source is a self-reference.
133     It represents the database and the query that produced the
134     xml document.
135     The sourceID contains a timestamp.
136     The full URL is given in the tag UniformResourceIdentifier
137     but you need
138     to unescape ampersands and angle brackets to re-use it.
139     Query was: select * where
140     ((target.MoleculeStoichiometricFormula = 'CO')) AND
141     ((collider.AtomSymbol = 'he'))</Comments>
142     <Year>2015</Year>
143     <Category>database</Category>
144     <UniformResourceIdentifier>http://cdms.ph1.uni-koeln.de/cdms/tap/sync?
145     LANG=VSS2&
146     REQUEST=doQuery&FORMAT=XSAMS&QUERY= select *
147     where ((target.MoleculeStoichiometricFormula = 'CO'))
148     AND ((collider.AtomSymbol = 'he'))</
149     UniformResourceIdentifier>
150     <ProductionDate>2015-12-03</ProductionDate>
151     <Authors>
152     <Author>
153     <Name>N.N.</Name>
154     </Author>
155     </Authors>
156 </Source>
157 <Source sourceID="BCDMS-1921">
158 <Authors>
159 <Author>
160 <Name>Mller, H. S. P.</Name>
161 </Author>
162 <Author>
163 <Name>Endres, C. P.</Name>

```

```

157         </Author>
158     <Author>
159         <Name>Schlemmer, S.</Name>
160     </Author>
161     <Author>
162         <Name>Stutzki, J.</Name>
163     </Author>
164 </Authors>
165 <Title />
166 <Category>database</Category>
167 <Year>2012</Year>
168 <SourceName>CDMS database</SourceName>
169 </Source>
170 <Source sourceID="BCDMS-1681">
171     <Authors>
172         <Author>
173             <Name>Winnewisser, G.</Name>
174         </Author>
175         <Author>
176             <Name>Belov, S. P.</Name>
177         </Author>
178         <Author>
179             <Name>Klaus, T.</Name>
180         </Author>
181         <Author>
182             <Name>Schieder, R.</Name>
183         </Author>
184     </Authors>
185     <Title />
186     <Category>journal</Category>
187     <Year>1997</Year>
188     <SourceName>J. Mol. Spectrosc.</SourceName>
189     <Volume>184</Volume>
190     <PageBegin>468</PageBegin>
191     <DigitalObjectIdentifier>10.1006/jmsp.1997.7341</DigitalObjectIdentifier>
192 </Source>
193 </Sources>
194 </XSAMSData>

```

## C Merged output from SpectCol

```

1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <XSAMSData xmlns="http://vamdc.org/xml/xsams/1.0"
3   xmlns:cml="http://www.xml-cml.org/schema"
4   xmlns:asymcs="http://vamdc.org/xml/xsams/1.0/cases/asymcs"
5   xmlns:asymos="http://vamdc.org/xml/xsams/1.0/cases/asymos"
6   xmlns:dcs="http://vamdc.org/xml/xsams/1.0/cases/dcs"
7   xmlns:gen="http://vamdc.org/xml/xsams/1.0/cases/gen"
8   xmlns:hunda="http://vamdc.org/xml/xsams/1.0/cases/hunda"

```

```

9   xmlns:hundb="http://vamdc.org/xml/xsams/1.0/cases/hundb"
10  xmlns:lpcs="http://vamdc.org/xml/xsams/1.0/cases/lpcs"
11  xmlns:lpos="http://vamdc.org/xml/xsams/1.0/cases/lpos"
12  xmlns:ltcs="http://vamdc.org/xml/xsams/1.0/cases/ltcs"
13  xmlns:ltos="http://vamdc.org/xml/xsams/1.0/cases/ltos"
14  xmlns:nltcs="http://vamdc.org/xml/xsams/1.0/cases/nltcs"
15  xmlns:nltos="http://vamdc.org/xml/xsams/1.0/cases/nltos"
16  xmlns:sphcs="http://vamdc.org/xml/xsams/1.0/cases/sphcs"
17  xmlns:sphos="http://vamdc.org/xml/xsams/1.0/cases/sphos"
18  xmlns:stcs="http://vamdc.org/xml/xsams/1.0/cases/stcs"
19  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
20  xsi:schemaLocation="http://vamdc.org/xml/xsams/1.0
https://raw.githubusercontent.com/nicolasmoreau/VAMDC-XSAMS/abstract_origin/xsams.xsd"
21  >
22  <Origin xsi:type="OtherOriginType">
23    <Timestamp>2015-12-07T15:50:21+01:00</Timestamp>
24    <Version versionID="VERMER1" global="false"
timestamp="2000-10-01T12:00:00+01:00">
25      <ProcessRef>PBASC50t2T1c1C1</ProcessRef>
26      <SourceRef>BBAS0</SourceRef>
27      <SourceRef>BBAS849</SourceRef>
28      <SourceRef>BCDMS0</SourceRef>
29      <SourceRef>BCDMS-1921</SourceRef>
30      <SourceRef>BCDMS-1681</SourceRef> </Version>
31    <HomepageUrl>http://www.vamdc.org/activities/research/software/spectcol/</HomepageUr
32    <Name>Spectcol</Name>
33    <Comments>Merging between collisional data from Basecol and
spectroscopic data from CDMS by cross matching state with
same J values in CO</Comments>
34    <Origin xsi:type="VamdcNodeOriginType">
35      <Timestamp>2015-12-03T15:50:21+01:00</Timestamp>
36      <Version versionID="VERCDMS1" global="false"
timestamp="2000-10-01T12:00:00+01:00">
37        <SpeciesRef>XCDMS-83</SpeciesRef>
38        <StateRef>SCDMS-83-1</StateRef>
39        <StateRef>SCDMS-83-2</StateRef>
40        <StateRef>SCDMS-origin-83</StateRef>
41        <ProcessRef>PCDMS-R15140649</ProcessRef>
42      </Version>
43      <HomepageUrl>http://cdms.ph1.uni-koeln.de/</HomepageUrl>
44      <Name>CDMS database</Name>
45      <Comments>Spectroscopic data extracted from CDMS VAMDC
node.</Comments>
46      <Query>select * where (RadTransWavelength >= 2.6006E7
AND RadTransWavelength <= 2.6008E7) AND
((MoleculeStoichiometricFormula = 'CO'))</Query>
47      <OriginIdentifier>ivo://vamdc/cdms/vamdc-tap_12.07</OriginIdentifier>
48    </Origin>
49    <Origin xsi:type="VamdcNodeOriginType">

```

```

50     <Timestamp>2015-12-03T14:40:21+01:00</Timestamp>
51     <Version versionID="VER001" global="false"
      timestamp="2015-09-01T08:10:12+01:00">
52         <SpeciesRef>XBAS2</SpeciesRef>
53         <StateRef>SBASET54-1</StateRef>
54     </Version>
55     <HomepageUrl>http://basecol.vamdc.org</HomepageUrl>
56     <Name>Basecol</Name>
57     <Comments>Collisional data extracted from Basecol VAMDC
      node.</Comments>
58     <Query>select * where
      ((target.MoleculeStoichiometricFormula = 'CO')) AND
      ((collider.AtomSymbol = 'he'))</Query>
59     <OriginIdentifier>ivo://vamdc/basecol/vamdc-tap_12.07</OriginIdentifier>
60 </Origin>
61 </Origin>
62 <Species>
63     <Atoms>
64         <Atom>
65             <ChemicalElement>
66                 <NuclearCharge>2</NuclearCharge>
67                 <ElementSymbol>He</ElementSymbol>
68             </ChemicalElement>
69             <Isotope>
70                 <IsotopeParameters>
71                     <MassNumber>4</MassNumber>
72                     <Mass>
73                         <Value units="amu">4.0026</Value>
74                     </Mass>
75                 </IsotopeParameters>
76                 <Ion speciesID="XBAS2">
77                     <IonCharge>0</IonCharge>
78                     <AtomicState stateID="SBASET54-1">
79                         <Comments>Energy level of He (no
                          structure)</Comments>
80                         <SourceRef>BBAS0</SourceRef>
81                         <AtomicNumericalData>
82                             <StateEnergy>
83                                 <Value units="1/cm">0.0</Value>
84                             </StateEnergy>
85                         </AtomicNumericalData>
86                         <AtomicQuantumNumbers>
87                             <TotalAngularMomentum>0.0</TotalAngularMomentum>
88                         </AtomicQuantumNumbers>
89                         <AtomicComposition>
90                             <Component>
91                                 <Term>
92                                     <LS>
93                                         <L>

```

```

94         <Value>0</Value>
95         <Symbol>L</Symbol>
96     </L>
97     <S>0.0</S>
98 </LS>
99 </Term>
100 </Component>
101 </AtomicComposition>
102 </AtomicState>
103 <InChIKey>SWQJXJQGLNCZEY-UHFFFAOYSA-N</InChIKey>
104 </Ion>
105 </Isotope>
106 </Atom>
107 </Atoms>
108 <Molecules>
109     <Molecule speciesID="XCDMS-83">
110         <MolecularChemicalSpecies>
111             <OrdinaryStructuralFormula>
112                 <Value>CO</Value>
113             </OrdinaryStructuralFormula>
114             <StoichiometricFormula>CO</StoichiometricFormula>
115             <ChemicalName>
116                 <Value>Carbon Monoxide</Value>
117             </ChemicalName>
118             <InChI>1S/CO/c1-2</InChI>
119             <InChIKey>UGFAIRIUMAVXCW-UHFFFAOYSA-N</InChIKey>
120             <VAMDCSpeciesID>UGFAIRIUMAVXCW-UHFFFAOYSA-N</VAMDCSpeciesID>
121             <PartitionFunction>
122                 <T units="K">
123                     <DataList>1.072 1.148 1.23 1.318 1.413 1.514 1.622
1.738 1.862 1.995 2.138 2.291 2.455 2.63 2.725
2.818 3.02 3.236 3.467 3.715 3.981 4.266 4.571
4.898 5.0 5.248 5.623 6.026 6.457 6.918 7.413
7.943 8.511 9.12 9.375 9.772 10.471 11.22 12.023
12.882 13.804 14.791 15.849 16.982 18.197 18.75
19.498 20.893 22.387 23.988 25.704 27.542 29.512
31.623 33.884 36.308 37.5 38.905 41.687 44.668
47.863 51.286 54.954 58.884 63.096 67.608 72.444
75.0 77.625 83.176 89.125 95.499 102.329 109.648
117.49 125.893 134.896 144.544 150.0 154.882
165.959 177.828 190.546 204.174 218.776 225.0
234.423 251.189 269.153 288.403 300.0 309.03
331.131 354.813 380.189 407.38 436.516 467.735
500.0 501.187 537.032 575.44 616.595 660.693
707.946 758.578 812.831 870.964 933.254
1000.0</DataList>
124                 </T>
125             <Q>

```

```

126      <DataList>1.01721581835 1.02422837677
1.03341102522 1.04512141245 1.05985020123
1.07774884294 1.09923417933 1.12473229226
1.15441650589 1.18864057708 1.22774305695
1.27174864212 1.32092124733 1.37520246158
1.40530654907 1.43514349042 1.50100121324
1.57270208471 1.65050158216 1.73501239088
1.8265276783 1.92535019369 2.03179140659
2.14652192999 2.18241666936 2.26987115366
2.40252640741 2.54553752733 2.69889427601
2.8632973732 3.04016627563 3.22985295285
3.4334258678 3.65195638949 3.74352703447
3.88615863094 4.13746641418 4.40695607053
4.69606494313 5.00551115749 5.33781506844
5.69369660756 6.0753182363 6.48412210856
6.9226335574 7.12225433182 7.3922960322
7.89599749116 8.4355434648 9.0138233311
9.633727213 10.2977845092 11.0096090065
11.7724538098 12.5895724935 13.4656648032
13.8965040828 14.4043468795 15.4099582553
16.4875620501 17.6425836137 18.8800875183
20.2062236736 21.6271428763 23.1500815655
24.7815541073 26.5302455525 27.4545173427
28.40375754 30.4111396059 32.5625279126
34.8676989471 37.3378779761 39.9850161394
42.8214291179 45.8608826253 49.1175078254
52.6076105394 54.5813579159 56.3475015965
60.3549445917 64.6491571735 69.2508130864
74.1820459087 79.4660945815 81.7185083304
85.1287658147 91.1970158766 97.6997287418
104.669260558 108.868854329 112.139428854
120.146852376 128.732781329 137.942190312
147.825233889 158.438549717 169.84604324
181.685035016 182.121688241 195.35029188
209.62906021 225.070653126 241.803934912
259.976802963 279.755720344 301.328756805
324.90561697 350.71803967 379.021079637</DataList>
127      </Q>
128      </PartitionFunction>
129      <StableMolecularProperties>
130      <MolecularWeight>
131      <Value units="unitless">28.0</Value>
132      </MolecularWeight>
133      </StableMolecularProperties>
134      <Comment> 28503- v 1:CO; $v=0$</Comment>
135      </MolecularChemicalSpecies>
136      <MolecularState stateID="SCDMS-83-1">
137      <MolecularStateCharacterisation>
138      <StateEnergy energyOrigin="SCDMS-origin-83">

```

```

139         <Value units="1/cm">0.0</Value>
140     </StateEnergy>
141     <TotalStatisticalWeight>1</TotalStatisticalWeight>
142     <NuclearStatisticalWeight>1</NuclearStatisticalWeight>
143 </MolecularStateCharacterisation>
144 <Case
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:type="dcs:Case" caseID="dcs">
145     <dcs:QNs>
146         <dcs:J>0</dcs:J>
147     </dcs:QNs>
148 </Case>
149 </MolecularState>
150 <MolecularState stateID="SCDMS-83-2">
151     <MolecularStateCharacterisation>
152         <StateEnergy energyOrigin="SCDMS-origin-83">
153             <Value units="1/cm">3.845033</Value>
154         </StateEnergy>
155         <TotalStatisticalWeight>3</TotalStatisticalWeight>
156         <NuclearStatisticalWeight>1</NuclearStatisticalWeight>
157     </MolecularStateCharacterisation>
158     <Case
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="dcs:Case" caseID="dcs">
159         <dcs:QNs>
160             <dcs:J>1</dcs:J>
161         </dcs:QNs>
162     </Case>
163 </MolecularState>
164 <MolecularState auxillary="true"
    stateID="SCDMS-origin-83">
165     <MolecularStateCharacterisation>
166         <StateEnergy energyOrigin="SCDMS-origin-83">
167             <Value units="1/cm">0.0</Value>
168         </StateEnergy>
169         <TotalStatisticalWeight>1</TotalStatisticalWeight>
170         <NuclearStatisticalWeight>1</NuclearStatisticalWeight>
171     </MolecularStateCharacterisation>
172     <Case
        xmlns:dcs="http://vamdc.org/xml/xsams/1.0/cases/dcs"
        xsi:type="dcs:Case" caseID="dcs"
        xsi:schemaLocation="http://vamdc.org/xml/xsams/1.0/cases/dcs
        ../../cases/dcs.xsd">
173         <dcs:QNs>
174             <dcs:ElecStateLabel>X</dcs:ElecStateLabel>
175             <dcs:v>0</dcs:v>
176             <dcs:J>0</dcs:J>
177         </dcs:QNs>
178     </Case>

```



```

179     </MolecularState>
180 </Molecule>
181 </Molecules>
182 </Species>
183 <Processes>
184   <Radiative>
185     <RadiativeTransition id="PCDMS-R15140649"
186     process="excitation">
187       <SourceRef>BCDMS0</SourceRef>
188       <EnergyWavelength>
189         <Frequency>
190           <Value units="MHz">115271.2021</Value>
191           <Accuracy>1.0E-4</Accuracy>
192         </Frequency>
193       </EnergyWavelength>
194       <UpperStateRef>SCDMS-83-2</UpperStateRef>
195       <LowerStateRef>SCDMS-83-1</LowerStateRef>
196       <SpeciesRef>XCDMS-83</SpeciesRef>
197       <Probability>
198         <TransitionProbabilityA>
199           <Value units="1/cm">7.20360334988E-8</Value>
200         </TransitionProbabilityA>
201         <IdealisedIntensity>
202           <Value units="unitless">-5.0105</Value>
203         </IdealisedIntensity>
204         <Multipole>E2</Multipole>
205       </Probability>
206       <ProcessClass>
207         <Code>rota</Code>
208       </ProcessClass>
209     </RadiativeTransition>
210   </Radiative>
211   <Collisions>
212     <CollisionalTransition id="PBASC50t2T1c1C1">
213       <Comments> Rotational de-excitation of CO (v=0) by He
214       (Cecchi-Pestellini & al, 2002)</Comments>
215       <SourceRef>BBAS0</SourceRef>
216       <ProcessClass>
217         <Code>inel</Code>
218       </ProcessClass>
219       <Reactant>
220         <SpeciesRef>XCDMS-83</SpeciesRef>
221         <StateRef>SCDMS-83-2</StateRef>
222       </Reactant>
223       <Reactant>
224         <SpeciesRef>XBAS2</SpeciesRef>
225         <StateRef>SBASET54-1</StateRef>
226       </Reactant>
227     </CollisionalTransition>
228   </Collisions>
229 </Processes>

```

```

226     <SpeciesRef>XCDMS-83</SpeciesRef>
227     <StateRef>SCDMS-83-1</StateRef>
228 </Product>
229 <Product>
230     <SpeciesRef>XBAS2</SpeciesRef>
231     <StateRef>SBASET54-1</StateRef>
232 </Product>
233 <DataSets>
234     <DataSet dataDescription="rateCoefficient">
235         <TabulatedData>
236             <Comments>Rate coefficients</Comments>
237             <X units="K">
238                 <DataList>5.0 10.0 20.0 40.0 60.0 80.0 100.0
239                     200.0 300.0 500.0</DataList>
240             </X>
241             <Y units="cm3/s">
242                 <DataList>3.4E-11 3.2E-11 3.0E-11 2.8E-11
243                     2.7E-11 2.6E-11 2.6E-11 2.5E-11 2.5E-11
244                     2.6E-11</DataList>
245             </Y>
246         </TabulatedData>
247     </DataSet>
248 </DataSets>
249 </CollisionalTransition>
250 </Collisions>
251 </Processes>
252 <Sources>
253     <Source sourceID="BBAS0">
254         <Category>database</Category>
255         <SourceName>BASECOL database</SourceName>
256         <Year>2015</Year>
257         <Authors>
258             <Author>
259                 <Name>M.-L. Dubernet</Name>
260             </Author>
261         </Authors>
262         <UniformResourceIdentifier>http://basecol.obspm.fr</UniformResourceIdentifier>
263         <ProductionDate>2015-12-03</ProductionDate>
264         <Comments>select * where
265             ((target.MoleculeStoichiometricFormula = 'CO')) AND
266             ((collider.AtomSymbol = 'he'))</Comments>
267     </Source>
268     <Source sourceID="BBAS849">
269         <Category>journal</Category>
270         <SourceName>apj</SourceName>
271         <Year>2002</Year>
272         <Authors>
273             <Author>
274                 <Name>N. Balakrishnan</Name>

```

```

270     </Author>
271     <Author>
272         <Name>A. Dalgarno</Name>
273     </Author>
274     <Author>
275         <Name>C. Cecchi-Pestellini</Name>
276     </Author>
277     <Author>
278         <Name>E. Bodo</Name>
279     </Author>
280 </Authors>
281 <Title>Rotational and Vibrational Excitation of CO
Molecules by Collisions with  $^4\text{He}$  Atoms</Title>
282 <Volume>571</Volume>
283 <PageBegin>1015</PageBegin>
284 <PageEnd>1020</PageEnd>
285 <UniformResourceIdentifier>
286 http://adsabs.harvard.edu/cgi-bin/nph-bib_query?
287     bibcode=2002JChPh.116.4517K&db_key=PHY
288 </UniformResourceIdentifier>
289 </Source>
290 <Source sourceID="BCDMS0">
291 <Comments>This Source is a self-reference.
292 It represents the database and the query that produced the
xml document.
293 The sourceID contains a timestamp.
294 The full URL is given in the tag UniformResourceIdentifier
but you need
295 to unescape ampersands and angle brackets to re-use it.
296 Query was: select * where
((target.MoleculeStoichiometricFormula = 'CO')) AND
((collider.AtomSymbol = 'he'))</Comments>
297     <Year>2015</Year>
298     <Category>database</Category>
299     <UniformResourceIdentifier>http://cdms.ph1.uni-koeln.de/cdms/tap/sync?
300     LANG=VSS2&
301     REQUEST=doQuery&FORMAT=XSAMS&QUERY= select *
where ((target.MoleculeStoichiometricFormula = 'CO'))
AND ((collider.AtomSymbol = 'he'))</
UniformResourceIdentifier>
302     <ProductionDate>2015-12-03</ProductionDate>
303     <Authors>
304         <Author>
305             <Name>N.N.</Name>
306         </Author>
307     </Authors>
308 </Source>
309 <Source sourceID="BCDMS-1921">
310     <Authors>

```

```

311         <Author>
312             <Name>Mller, H. S. P.</Name>
313         </Author>
314         <Author>
315             <Name>Endres, C. P.</Name>
316         </Author>
317         <Author>
318             <Name>Schlemmer, S.</Name>
319         </Author>
320         <Author>
321             <Name>Stutzki, J.</Name>
322         </Author>
323     </Authors>
324     <Title />
325     <Category>database</Category>
326     <Year>2012</Year>
327     <SourceName>CDMS database</SourceName>
328 </Source>
329 <Source sourceID="BCDMS-1681">
330     <Authors>
331         <Author>
332             <Name>Winnewisser, G.</Name>
333         </Author>
334         <Author>
335             <Name>Belov, S. P.</Name>
336         </Author>
337         <Author>
338             <Name>Klaus, T.</Name>
339         </Author>
340         <Author>
341             <Name>Schieder, R.</Name>
342         </Author>
343     </Authors>
344     <Title />
345     <Category>journal</Category>
346     <Year>1997</Year>
347     <SourceName>J. Mol. Spectrosc.</SourceName>
348     <Volume>184</Volume>
349     <PageBegin>468</PageBegin>
350     <DigitalObjectIdentifier>10.1006/jmsp.1997.7341</DigitalObjectIdentifier>
351 </Source>
352 </Sources>
353 <Comments>Data merged by SPECTCOL.</Comments>
354 </XSAMSData>

```

## References

- [1] Cesare Cecchi-Pestellini, Enrico Bodo, N. Balakrishnan, and Alexander Dalgarno. Rotational and vibrational excitation of co molecules

- by collisions with 4He atoms. *The Astrophysical Journal*, 571(2):1015, 2002.
- [2] J. F. Corby, P. A. Jones, M. R. Cunningham, K. M. Menten, A. Belloche, F. R. Schwab, A. J. Walsh, E. Balnozan, L. Bronfman, N. Lo, and A. J. Remijan. An ATCA survey of Sagittarius B2 at 7 mm: chemical complexity meets broad-band interferometry. *M.N.R.A.S.*, 452:3969–3993, October 2015.
- [3] M. Dubernet, L. Nenadovic, and N. Doronin. SPECTCOL: A New Software to Combine Spectroscopic Data and Collisional Data within VAMDC. In P. Ballester, D. Egret, and N. P. F. Lorente, editors, *Astronomical Data Analysis Software and Systems XXI*, volume 461 of *Astronomical Society of the Pacific Conference Series*, page 335, September 2012.
- [4] M.-L. Dubernet, M. H. Alexander, Y. A. Ba, N. Balakrishnan, C. Balança, C. Ceccarelli, J. Cernicharo, F. Daniel, F. Dayou, M. Doronin, F. Dumouchel, A. Faure, N. Feautrier, D. R. Flower, A. Grosjean, P. Halvick, J. Klos, F. Lique, G. C. McBane, S. Marinakis, N. Moreau, R. Moszynski, D. A. Neufeld, E. Roueff, P. Schilke, A. Spielfiedel, P. C. Stancil, T. Stoecklin, J. Tennyson, B. Yang, A.-M. Vasserot, and L. Wiesenfeld. BASECOL2012: A collisional database repository and web service within the Virtual Atomic and Molecular Data Centre (VAMDC). *Astron. & Astrophys.*, 553:A50, May 2013.
- [5] M. L. Dubernet, V. Boudon, J. L. Culhane, M. S. Dimitrijevic, A. Z. Fazliev, C. Joblin, F. Kupka, G. Leto, P. Le Sidaner, P. A. Loboda, H. E. Mason, N. J. Mason, C. Mendoza, G. Mulas, T. J. Millar, L. A. Nuñez, V. I. Perevalov, N. Piskunov, Y. Ralchenko, G. Rixon, L. S. Rothman, E. Roueff, T. A. Ryabchikova, A. Ryabtsev, S. Sahal-Bréchet, B. Schmitt, S. Schlemmer, J. Tennyson, V. G. Tyuterev, N. A. Walton, V. Wakelam, and C. J. Zeippen. Virtual atomic and molecular data centre. *J. Quant. Spectrosc. & Rad. Transfer*, 111:2151–2159, October 2010.
- [6] M.-L. Dubernet, A. Grosjean, D. Flower, E. Roueff, F. Daniel, N. Moreau, and B. Debray. Ro-Vibrational Collisional Excitation Database: BASECOL - <http://www.obspm.fr/basecol>. *Journal of Plasma Research SERIES*, 7:356–357, January 2006.
- [7] M.-L. Dubernet and L. Nenadovic. SPECTCOL: Spectroscopic and Collisional Data Retrieval. Astrophysics Source Code Library, November 2011.
- [8] Marie-Lise Dubernet, Bobby Antony, Yaye-Awa Ba, Yuri Babikov, Klaus Bartschat, Vincent Boudon, B Braams, Hyun-Kyung Chung,

Fabien Daniel, Franck Delahaye, Giulio Del Zanna, Jaime de Urquijo, Milan Dimitrijevic, Alicja Domaracka, Mikhail Doronin, Brian Drouin, Christian Endres, Alexandr Fazliev, Serguey Gagarin, Iouli Gordon, Pierre Gratier, Ulrike Heiter, Christian Hill, Darko Jevremovic, Christine Joblin, Aurlie Karsprzak, E. Krishnakumar, Giuseppe Leto, Peter A. Loboda, Thierry Louge, Sylvain Maclot, Bratislav Marinkovic, Andrew Markwick Kemper, Thomas Marquart, Helen Mason, Nigel Mason, Claudio Mendoza, Anatoly Mihajlov, Tom Millar, Nicolas Moreau, Giacomo Mulas, Yury Pakhomov, Patrick Palmeri, Sergey Pancheshnyi, Valery I. Perevalov, Nikolai Piskunov, Johannes Postler, Ernesto Luis Quinet, Pascaland Quintas Snchez, Yuri Ralchenko, Yong-Joo Rhee, Guy Rixon, Laurence Rothman, Evelyne Roueff, Tatiana Ryabchikova, Sylvie Sahal-Brechot, Paul Scheier, Stephan Schlemmer, Bernard Schmitt, Eric Stempels, Serguey Tashkun, Jonathan Tennyson, Vladimir Tyuterev, Veljko Vujcic, Valentine Wakelam, Nicolas Walton, Oleg Zatsarinny, Claude Zeppen, and Carlo Maria Zwölf. The Virtual Atomic and Molecular Data Centre (VAMDC) Consortium. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 49(7):074003, 2016.

- [9] M. A. Frerking and W. D. Langer. A measurement of the hyperfine structure of CO-17. *J. Chem. Phys.*, 74:6990, June 1981.
- [10] Ginard, D., Gonzlez-Garca, M., Fuente, A., Cernicharo, J., Alonso-Albi, T., Pilleri, P., Gerin, M., Garca-Burillo, S., Ossenkopf, V., Rizzo, J. R., Kramer, C., Goicoechea, J. R., Pety, J., Bern, O., and Joblin, C. Spectral line survey of the ultracompact hii region monoceros r2? *Astron. & Astrophys.*, 543:A27, 2012.
- [11] N. Huijboom and T. Van den Broek. Open data: an international comparison of strategies. *European Journal of ePractice*, N. 12, March-April 2011 - ISSN: 1988-625X, 2011.
- [12] N. Jacquinet-Husson, L. Crepeau, R. Armante, C. Boutammine, A. Chédin, N. A. Scott, C. Crevoisier, V. Capelle, C. Boone, N. Poulet-Crovisier, A. Barbe, A. Campargue, D. C. Benner, Y. Benilan, B. Bézard, V. Boudon, L. R. Brown, L. H. Coudert, A. Coustenis, V. Dana, V. M. Devi, S. Fally, A. Fayt, J.-M. Flaud, A. Goldman, M. Herman, G. J. Harris, D. Jacquemart, A. Jolly, I. Kleiner, A. Kleinböhl, F. Kwabia-Tchana, N. Lavrentieva, N. Lacome, L.-H. Xu, O. M. Lyulin, J.-Y. Mandin, A. Maki, S. Mikhailenko, C. E. Miller, T. Mishina, N. Moazzen-Ahmadi, H. S. P. Müller, A. Nikitin, J. Orphal, V. Perevalov, A. Perrin, D. T. Petkie, A. Predoi-Cross, C. P. Rinsland, J. J. Remedios, M. Rotger, M. A. H. Smith, K. Sung, S. Tashkun, J. Tennyson, R. A. Toth, A.-C. Vandaele, and J. Vander Auwera. The

- 2009 edition of the GEISA spectroscopic database. *J. Quant. Spectrosc. & Rad. Transfer*, 112:2395–2445, November 2011.
- [13] N. Jacquinet-Husson, N. A. Scott, A. Chedin, B. Bonnet, A. Barbe, V. G. Tyuterev, J. P. Champion, M. Winnewisser, L. R. Brown, R. Gamache, V. F. Golovko, and A. A. Chursin. The GEISA system in 1996: towards an operational tool for the second generation vertical sounders radiance simulation. *J. Quant. Spectrosc. & Rad. Transfer*, 59:511–527, May 1998.
- [14] JC. Molloy. The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biol* 9(12): e1001195, 2011.
- [15] H. S. P. Müller, F. Schlöder, J. Stutzki, and G. Winnewisser. The Cologne Database for Molecular Spectroscopy, CDMS: a useful tool for astronomers and spectroscopists. *Journal of Molecular Structure*, 742:215–227, May 2005.
- [16] H. S. P. Müller, S. Thorwirth, D. A. Roth, and G. Winnewisser. The Cologne Database for Molecular Spectroscopy, CDMS. *Astron. & Astrophys.*, 370:L49–L52, April 2001.
- [17] L. Pagani, F. Daniel, and M.-L. Dubernet. On the frequency of  $N\{_{-2}H^+$  and  $N\{_{-2}D\{^+}$ . *Astron. & Astrophys.*, 494:719–727, February 2009.
- [18] H. M. Pickett, R. L. Poynter, E. A. Cohen, M. L. Delitsky, J. C. Pearson, and H. S. P. Müller. Submillimeter, millimeter and microwave spectral line catalog. *J. Quant. Spectrosc. & Rad. Transfer*, 60:883–890, November 1998.
- [19] N. E. Piskunov, F. Kupka, T. A. Ryabchikova, W. W. Weiss, and C. S. Jeffery. VALD: The Vienna Atomic Line Data Base. *Astron. Astrophys. Suppl. Ser.*, 112:525, September 1995.
- [20] HA. Piwowar and TJ. Vision. Data reuse and the open data citation advantage. *ed. PeerJ*. 2013;1:e175, 2013.
- [21] A. Punanova, P. Caselli, A. Pon, A. Belloche, and P. André. Deuterium fractionation in the Ophiuchus molecular cloud. *Astron. & Astrophys.*, 587:A118, March 2016.
- [22] Andreas Rauber, Ari Asmi, Dieter van Uytvanck, and Stefan Proell. Data Citation of Evolving Data - Recommendations of the Working Group on Data Citation. <https://rd-alliance.org/rda-wgdc-recommendations-vers-sep-24-2015.html>, September 2015. Draft - Request for Comments.

- [23] L. S. Rothman. Atmospheric absorption-line parameters: the HITRAN data base (A). *Journal of the Optical Society of America (1917-1983)*, 73, 1983.
- [24] L. S. Rothman, I. E. Gordon, Y. Babikov, A. Barbe, D. Chris Benner, P. F. Bernath, M. Birk, L. Bizzocchi, V. Boudon, L. R. Brown, A. Campargue, K. Chance, E. A. Cohen, L. H. Coudert, V. M. Devi, B. J. Drouin, A. Fayt, J.-M. Flaud, R. R. Gamache, J. J. Harrison, J.-M. Hartmann, C. Hill, J. T. Hodges, D. Jacquemart, A. Jolly, J. Lamouroux, R. J. Le Roy, G. Li, D. A. Long, O. M. Lyulin, C. J. Mackie, S. T. Massie, S. Mikhailenko, H. S. P. Müller, O. V. Naumenko, A. V. Nikitin, J. Orphal, V. Perevalov, A. Perrin, E. R. Polovtseva, C. Richard, M. A. H. Smith, E. Starikova, K. Sung, S. Tashkun, J. Tennyson, G. C. Toon, V. G. Tyuterev, and G. Wagner. The HITRAN2012 molecular spectroscopic database. *J. Quant. Spectrosc. & Rad. Transfer*, 130:4–50, November 2013.
- [25] T. Ryabchikova, N. Piskunov, R. L. Kurucz, H. C. Stempels, U. Heiter, Y. Pakhomov, and P. S. Barklem. A major upgrade of the VALD database. *Physica Scripta*, 90(5):054005, May 2015.
- [26] G. Winnewisser, S. P. Belov, T. Klaus, and R. Schieder. Sub-Doppler Measurements on the Rotational Transitions of Carbon Monoxide. *Journal of Molecular Spectroscopy*, 184:468–472, August 1997.
- [27] C.M. Zwolf, M.-L. Dubernet, Y.-A. Ba, and N. Moreau. Experience and feedbacks from the sustainability for the virtual atomic and molecular data centre e-infrastructure. In *IST-Africa Conference Proceedings, 2014*, pages 1–9, May 2014.