BASECOL2020 New Technical Design

Yaye-Awa Ba [†], Marie-Lise Dubernet *,[†] [ID], Nicolas Moreau [†] [ID] and Carlo Maria Zwölf [†] [ID]

Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, LERMA, 92190 Meudon, France;
yaye-awa.ba@observatoiredeparis.psl.eu (Y.-A.B.); nicolas.moreau@observatoiredeparis.psl.eu (N.M.);
carlo-maria.zwolf@observatoiredeparis.psl.eu (C.M.Z.)

* Correspondence: marie-lise.dubernet@observatoiredeparis.psl.eu

† These authors contributed equally to this work.

Abstract: The BASECOL database has been created and scientifically enriched since 2004. It contains collisional excitation rate coefficients of molecules for application to the interstellar medium and to cometary atmospheres. Recently, major technical updates have been performed in order to be compliant with international standards for management of data and in order to provide a more friendly environment to query and to present the data. The current paper aims at presenting the key features of the technical updates and to underline the compatibility of BASECOL database with the Virtual Atomic and Molecular Data Center. This latter aims to interconnect atomic and molecular databases, thus providing a single location where users can access atomic and molecular data.

Keywords: molecular data; database; interoperability; FAIR data

1. Introduction to BASECOL

The current publication aims at providing a description of the latest technical developments performed on the BASECOL database. The intention is to provide a paper that conveys the technical quality of the BASECOL service in relation with the Virtual Atomic and Molecular Data Center (VAMDC) [1–3], which is supported by the VAMDC consortium[1]. The current VAMDC e-infrastructure interconnects about 38 atomic and molecular databases that cover atomic and molecular spectroscopy and processes. About 90% of the inter-connected databases handle data that are used for the interpretation of astronomical spectra and for the modeling in media of many fields of astrophysics. VAMDC offers a common entry point to all encorporated databases through the VAMDC portal[2] and it develops central services such as the species database[3] or the Query Store service[4]. VAMDC also develops standalone tools in order to retrieve and handle the data such as the SPECTCOL tool [4][5]. VAMDC provides software and support in order to include new databases within the VAMDC e-infrastructure[6]. One current feature of VAMDC e-infrastructure is the constrained environment for the description of data, in particular the XSAMS schema and other standardized protocols[7] that ensure a higher quality for the distribution of data.

---

1 http://www.vamdc.org/structure/how-to-join-us/
2 http://portal.vamdc.eu
3 http://species.vamdc.eu
4 http://cite.vamdc.eu
5 A new version will be released start of 2021 and a new publication will be published.
6 http://www.vamdc.org/activities/research/software/java-nodesoftware/, http://www.vamdc.eu/documents/nodesoftware/
7 http://www.vamdc.org/standards

The genesis of this database was prompted by the need of collisional inelastic rate coefficients for the analysis of astrophysical spectra obtained with the latest spatial and ground heterodyne instruments. The first publication [5] on BASECOL introduced the status of BASECOL in 2012 and had an exhaustive table about the 2012 scientific content. The database has been partially updated since 2013, and, once completed, further communication will provide the scientific status. An excellent review of the scientific needs and issues related to such collisional data can be found in [6].

The BASECOL database collects from the refereed literature the state-to-state rate coefficients for the inelastic excitation of rotational, vibrational and ro-vibrational levels of molecules by atoms, molecules, and electrons. Currently, the collisional rate coefficients come from theoretical calculations. The collisional processes are described in the temperature range relevant to the interstellar medium, to circumstellar atmospheres and to cometary atmospheres. In addition, the database includes the energy tables that contain identification of the levels involved in the collisional transitions, some fitted collisional data sets to the numerical data and the references attached to the collisional data sets. Using the services described below ("Search Collision" and "Search Article"), we provide here the following statistics[8]: there currently are 284 articles and 217 collisional sets. Among those 217 collisional data sets, we identify the excitation of 58 molecular species (among those 16 species are molecular ions) and of four atomic species. The molecular species span 26 diatomic molecules, 20 triatomic molecules, and 12 molecules with more than three atoms. Among the 217 collisional sets, 71, 97, 19, and 30 collisional sets have as a collider, respectively, He, H_2 (among those, 32 sets are with the ortho-H_2 collider), H, and electron.

BASECOL is primarily a scientific product that can be trusted and that is aimed at both producers of collisional rate coefficients and at users of those quantities. Indeed, one aspect of the BASECOL database is the quality, completeness, and safety of the provided information, as BASECOL is managed by scientists having a deep knowledge of the type of data ingested in the database. This means that the data are checked with respect to the methodologies used to calculate the data and with respect to the consistency of the data sets. The data policy of BASECOL is to keep a record of all available and published data sets. Other important scientific aspects of BASECOL are the versioning of the collisional data sets and the references attached to all sets of data. This allows traceability of the information.

From a technical point of view, BASECOL is a sophisticated product that follows international rules for management of FAIR data [7] (cf. Section 6) and that is compliant with the VAMDC infrastructure. In the following sections, we will describe the new multitier architecture of the BASECOL service, the steps taken towards improving data integrity and how the connection to the VAMDC is integrated into the BASECOL service. Finally, we will describe the public web interface that provides access to the BASECOL data.

2. The New Architecture of the BASECOL Service

In past years, major technical updates of BASECOL were prompted for the following reasons: (1) some of the technologies used in BASECOL2012 [5] were deprecated and difficult to maintain; (2) we needed to adapt our internal processes and procedures to the new international data-quality standards. Indeed, the data-feeding procedure of BASECOL2012 was manual with no strict trace of the actions performed by the scientific maintainer. In addition, it lacked an automatic ingestion of the metadata necessary to ensure the interoperability with the VAMDC ecosystem.

2.1. New Features

The new version of the BASECOL ecosystem is divided into three software components:

---

[8]　request from the referee.

- The core component is a relational database built upon the MariaDB RDBMS. A critical point associated with this component is how to preserve the scientific quality of the data, and how to guarantee data-traceability and reproducibility when numerical data, metadata, documentation, and bibliographic references are ingested in the database (see Section 3).
- The second software component is the website which is a web user interface (WebUI) embedded into a web browser: it provides a public and human visible access to the data (see Section 4).
- The last component is a REST-like API that provides a software interface to the database. It is a set of web services that perform the data-extraction from the central database. This API may be used not only by the website, but also by any other client. This API relies on the the Spring Java framework. The endpoint of each web service composing the API, reachable through the HTTP protocol, returns JSON[9] formatted data.

As shown in Figure 1, the BASECOL system is composed of a private "ingestion instance" and of a public "production instance". The latter is interfaced with VAMDC. As shown in Figure 2, a BASECOL "instance" is composed of its relational database, its API, and its web user interface. The "ingestion instance" allows the database maintainer to verify and to validate the imported data. The "ingestion database" is dumped into the "production database" at regular intervals, and a specific procedure is put into place in order to ensure that only the validated data are visible in the "production instance" website.
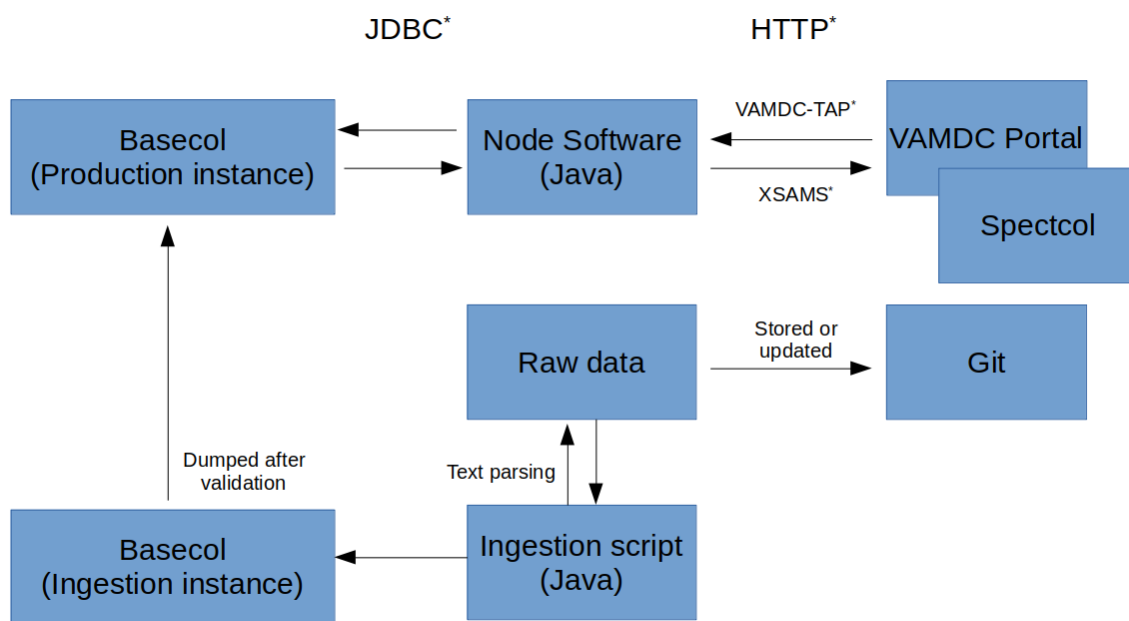

Figure 1. New structure of the BASECOL system. The term labeled with an asterisk is found in the Abbreviation section.

---

9   https://www.json.org/json-en.html

HTTP*   JDBC*

Web site (React*) ⇄ JSON* / Call endpoint ⇄ Rest API (Spring*) ⇄ DB (MariaDB*)
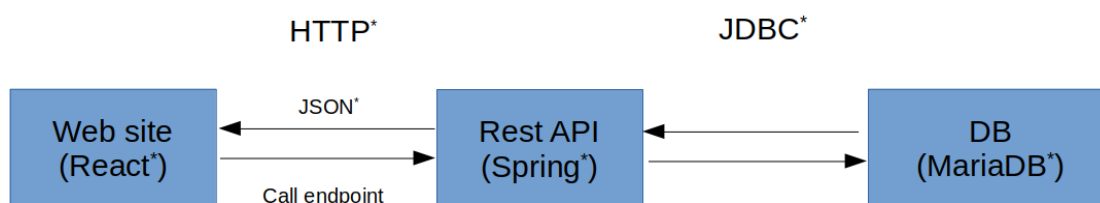
Figure 2. Organisation of a BASECOL instance. The terms labeled with an asterisk are found in the Abbreviation and Glossary sections.

Contrary to the previous version of the BASECOL environment, the new software organization follows the multitier architecture: the user interface and the database access are now completely decoupled, making future evolution of the system easier. The structure of the database can be modified without any impact on the user interface layer, as long as the REST API remains the same.

2.2. Removed Features

Now, the bibliographic section of the database only includes the publications that are referenced from the collisional data sets (cf. Section 4.2) and from the energy tables. This reduces the bibliographic entries to about 300 entries and this is justified by the overall coherence of the ingestion and versioning procedures.

The VAMDC portal and the SPECTCOL tool [4] implement the IVOA-SAMP protocol [8], created to connect client-side tools to improve productivity when working with multiple data types. VAMDC data can be directly piped into any tool implementing the SAMP protocol, e.g., TOPCAT[10]. Since the TOPCAT tool may be used to compare (or cross-match) different rate coefficients, we removed the cross-matching feature from the BASECOL web-interface.

We also removed the visualization of output files in a format compatible with the RADEX code[11](the format of those files is explained in Van der Tak et al. [6]) as users can build their own RADEX files using the SPECTCOL tool. In addition, at least two websites already provide this kind of information: the LAMDA database [6] and the CASSIS database [9][12]. The BASECOL database is rather a reference repository for collisional inelastic rate coefficients and the advice would be to take collisional data from the BASECOL database in order to build the RADEX files on those known websites [6,9] or in any other internal or public secondary display of collisional rate coefficients. This would ensure a proper versioning of the collisional data sets as a specific effort on versioning is made in BASECOL (see Section 3 and Appendix B), and a similar versioning effort is available in the molecular spectroscopic databases connected through VAMDC: CDMS, JPL, and HITRAN [10–13].

3. Technical Evolution for Improved Data Integrity and for VAMDC Content Requirements

We designed a new data import system that improves data integrity and authenticity (R7 of CoretrustSeal recommendation[13]), and that introduces rigorous procedures in managing archival long lasting storage of data. The data import system is composed of an import ASCII file, of a Java application that parses and loads the import file into the database and of a git repository. We first describe the import ASCII file and then the import procedure.

---

[10] http://www.star.bris.ac.uk/~mbt/topcat/
[11] Numerical code used to calculate molecular line profiles from the interstellar medium, https://personal.sron.nl/~vdtak/radex/index.shtml
[12] http://cassis.irap.omp.eu/
[13] https://www.coretrustseal.org/wp-content/uploads/2019/02/CoreTrustSeal-Draft_Requirements_2020-2022_v00_01.pdf

3.1. Description of the Import ASCII File

The import ASCII file is composed of several "Data" sections, and the concatenation of these sections constitutes what is called a "collisional data set". Each section of the import file corresponds to an "object" that follows the versioning rules (cf. Appendix B). In this paper, we give the same name both to the section of the import ASCII file and to the "object" which is the concept to which rules of versioning are applied. An "object" may include metadata and numerical data or just metadata. Some of the metadata are aimed to allow the compatibility of BASECOL with the VAMDC architecture, some metadata are either mandatory or optional, and some medatata are related to the uniqueness of an "object". The concept of uniqueness is attached to the numerical data as well, and is used in the versioning process handled by the import script (cf. Section 3.2 and Appendix B). The different data sections are the "Element" section, the "Energy table" section, the "Energy Origin" section, the "Rate Coefficients" section, the optional "Fitting coefficients" section and the "Publications" section. The "Element", "Energy table", and "Energy Origin" sections must be provided for the molecule called "target", i.e., the molecule whose collisional excitation is of interest for astrophysical users, and for the "collider", i.e., the perturbing atom, molecule or electron. These sections are described thereafter.

3.1.1. Element Section

The "Element" section is compulsory and corresponds to the object "Element" of Appendix B which is composed of a list of metadata given in Table 1. An example of an "Element" section input for N_2H^+ is given in Table A1. The metadata of the "Element" object allowing compatibility with VAMDC concern the identification of the species. The VAMDC standards provide standardization of the species through the metadata "stoichiometric formula" (the atoms are ordered in alphabetic order and the number of occurences of the atom is given after the atomic symbol), and through the "inchikey". When a species is already present in BASECOL it is possible to retrieve the species metadata from BASECOL through a web user interface (cf. Appendix C); otherwise, information can be retrieved from the species database[14] when the metadata are linked to VAMDC, with the exception of "mass". If neither BASECOL nor the species database contain the species information, the BASECOL data provider can contact the VAMDC team to get advice (support@vamdc.eu).

Table 1. List of properties/metadata for the object "Element". The symbols [M] and [O] indicate that the metadata are either mandatory or optional in the import file. In the column "Uniq.", we indicate whether the metadata defines the uniqueness of the object. In the column "Purpose", we indicate whether the information is meant for the BASECOL (B) public WebUI display and/or for VAMDC (V) interoperability. In "V(parameter)", "parameter" indicates the name of column in the species database. The mass unit is atomic mass unit (a.m.u) and the default value for mass is zero.

Metadata Name	Description	Type	Uniq.	Purpose
name	species name	[M]	yes [a]	V(name)/B
inchi	species structure description in inchi format	[M] [b]		V(InChi)
inchikey	species inchikey (27 characters hash of inchi) [c]	[M] [a]	yes	V(InChiKey)
stoichiometricFormula	species stoichiometric formula	[M]		V(Stoichiometric)/B
htmlName	name of species in HTML	[M]		B
latex	name of species in latex	[M]		V(Formula)/B
mass	species mass	[M]		V
molecularConstant	rotational constant (linear molecule)	[O]		B
charge	species charge	[M]		V (charge)
elementType	atom, molecule or particle	[M]		V

[a]: for particle only [b] [M]: only for atoms and molecules [c] https://www.inchi-trust.org/.

3.1.2. Energy Table Section

The "Energy Table" section is compulsory and corresponds to the object "Energy Table" of Appendix B, which is composed of a list of metadata given in Table 2 and of a numerical table. The imported numerical energy table identifies the labels of the collisional transition in the rate coefficients table. The administrator can import the energy table used in the rate coefficients calculations, or an energy table created with data from spectroscopic databases such as CDMS, JPL, and HITRAN [10–13]. A template of an "Energy table" section input is given in Table A2. The numerical energy table is composed of labels, values of the quantum numbers, and numerical values for the energy. The energy tables are currently being harmonized so that the default energy unit be "cm^{-1}". The quantum numbers' description follows the VAMDC standards: the atomic quantum numbers are standardized in the XSAMS document[15] and the molecular quantum numbers are described in the case-by-case document[16] where different situations are identified (for example: case ="dcs" means *diatomic close-shell molecules*, case = "hunda" means *Hund's case (a) diatomics*, etc....). For each "case", a set of quantum numbers is defined, and the import file must follow those standards. If a given pattern of quantum numbers characterization exists in BASECOL (similar type of molecule and coupling case), the choice of quantum numbers can be retrieved from the import file WebUI (see Appendix C).

Table 2. List of properties/metadata for the "Energy Table" object. "case" refers to the case-by-case document (http://www.vamdc.eu/documents/cbc-1.0/). The symbols [M] and [O] indicate that the metadata are either mandatory or optional in the import file. In the column "Uniq.", we indicate whether the metadata defines the uniqueness of the object. In the column "Purpose", we indicate whether the information is meant for the BASECOL (B) public WebUI display and/or for VAMDC (V) interoperability.

Property Name	Description	Type	Uniq.	Purpose
case	VAMDC classification for quantum numbers sets	[M] [a]	yes [a]	V
symmetry	species symmetry	[M]	yes [a]	V/B [a]
title	name of the energy table	[M]		V/B
comments	how the energy levels have been obtained	[O]		B
type	theoretical or experimental table	[M]		V/B
quantumNumbers	list of quantum numbers characterizing energy levels	[M]	yes	V/B

[a] [M]: only for molecules.

3.1.3. Energy Origin Section

The "Energy Origin" section is compulsory and corresponds to the object "Energy Origin" of Appendix B which is composed of a list of metadata given in Table 3 and of a numerical table with a single line. This single line gives the values of the quantum numbers associated with the energy origin of the corresponding "Energy Table" section (cf. Section 3.1.2). A template of an "Energy Origin" section input is given in Table A3.

15　https://standards.vamdc.eu/#data-model
16　http://www.vamdc.eu/documents/cbc-1.0/

Table 3. List of properties/metadata for "Energy Origin" object tables. "case" refers to the case-by-case document (http://www.vamdc.eu/documents/cbc-1.0/). The symbols [M] and [O] indicate that the metadata are either mandatory or optional in the import file. In the column "Uniq.", we indicate whether the metadata defines the uniqueness of the object. In the column "Purpose", we indicate whether the information is meant for the BASECOL (B) public WebUI display and/or for VAMDC (V) interoperability.

Property Name	Description	Type	Uniq.	Purpose
case	VAMDC classification for quantum numbers sets	[M] [a]	yes [a]	V
symmetry	species symmetry	[M] [a]	yes [a]	V/B
quantumNumbers	list of quantum numbers characterizing energy levels	[M]	yes	V/B

[a] [M]: only for molecules.

3.1.4. Rate Coefficients Section

The "Rate Coefficients" section comes next, it is compulsory, and corresponds to the object "Rate Coefficients" of Appendix B, which is composed of a list of metadata given in Table 4 and of a numerical table. A template of a "Rate Coefficients" section input is given in Table A4.

BASECOL allows the collider to undergo transitions, and therefore the initial and final levels of the collider must be provided. The BASECOL collisional rate coefficients (in units of cm^3s^{-1}) can be state-to-state rate coefficients, effective rate coefficients, and thermalized coefficients; those items are scientifically defined in Dubernet et al. [5] and must be included in the the metadata "rateType" of Table A4.

Table 4. List of properties/metadata for the object "Rate Coefficients". The symbols [M] and [O] indicate that the metadata are either mandatory or optional in the import file. In the column "Uniq.", we indicate whether the metadata defines the uniqueness of the object. In the column "Purpose", we indicate whether the information is meant for the BASECOL (B) public WebUI display and/or for VAMDC (V) interoperability. The unit of "temperatures" is Kelvin.

Property Name	Description	Type	Uniq.	Purpose
title	collision title	[M]		V/B
comment	description of the data set	[O]		V/B
pesComment	description of PES	[O]		V/B
methodComment	indicates method	[O]		V/B
precision	given by author or by administrator	[O]		V/B
process	collision process	[M]		B [a]
year	Publication year of article containing the data	[M]		V/B
rateType	see Table A4	[M]	yes	B [b]
temperatures	list of temperatures for which rates are available	[M]	yes	V/B

[a] The BASECOL "process" parameter will be returned in VAMDC query results in the next update of the BASECOL node software. For now, BASECOL has a returnable CollisionCode = "inel" (return value in VAMDC; see Section 8). [b] This parameter will be returned in VAMDC query results once the VAMDC standards have evolved.

3.1.5. Fitting Coefficients Section

The import script optionally processes the "Fitting Coefficients" section. This section corresponds to the object "Fitting Coefficients" of Appendix B, which is composed of a list of metadata given in Table 5 and of a numerical table. A template of a "Fitting Coefficients" section input is given in Table A5. The transitions of the numerical fitting coefficients table must be organized identically to the transitions of the related rate coefficients table.

Table 5. List of properties/metadata for the object "Fitting coefficients". The symbols [M] and [O] indicate that the metadata are either mandatory or optional in the import file. In the column "Uniq.", we indicate whether the metadata define the uniqueness of the object. In the column "Purpose", we indicate whether the information is meant for the BASECOL (B) public WebUI display and/or for VAMDC (V) interoperability.

Property Name	Description	Type	Uniq.	Purpose
title	title for the set of fitting coefficients	[M]		B
function	equation of function fitting the data	[M]	yes	B
comment	description of the function	[O]		B
designation	identifier of the function	[M]		B
coefficients	list of fit coefficients names	[M]	yes	B

Currently, 86 collisional rate coefficients sets have fitting tables, and those fits have been either provided by the authors or calculated by the BASECOL maintainers. The list of fitted collisional sets is indicated in Table A1 of our previous publication [5]. No additional sets have been fitted since 2013 because of lack of manpower[17]. The fitting functions are the so-called "common fit equation":

$$\log(R(T)) = \sum_{k=1}^{N-1} a_k \left[log\left(\frac{T}{\epsilon}\right) \right]^{k-1} + a_N \left(\frac{\epsilon}{T} - 1\right), \tag{1}$$

with T the temperature in Kelvin, R the rate coefficients in cm^3/s, a_k, a_N, and ϵ the fitted parameters, and specific fitting functions used in various publications for their data: "Faure et al. 2004" [14], "Faure et al. 2001" [15], "Sarpal et al. 1993" [16], "Lim et al. 1999" [17] as indicated in Table A5.

3.1.6. Publications Section

Finally, the import file imposes to process the "Publications" section. This section may include several "Publication" objects; each object is composed of a list of metadata given in Table 6 and follows the versioning defined in Appendix B. The "Publications" section input file must contain at least one "Publication" object, which is the reference to the paper where the collisional data set has been published. This main reference is indicated in the publication input file with \$\$mainArticle = yes; this reference appears in red in the "References" section of the public WebUI (cf. Section 4.2). If the collisional dataset is obtained through several papers, additional references, identified with \$\$mainArticle = no, can be included. The BASECOL team strongly recommends to include the references to the potential energy surface used in the collisional calculations. Additional references may be considered such as references linked to the collider and to the target energy levels, as well as to the method used in the calculations. We do not provide an input file for this entry as some minor format modifications are currently being made. In the case of old papers with no DOI, a random DOI is created with the string "tmp_doi" followed by a random unique identifier. This random internal DOI is not provided to VAMDC as DOI is not mandatory in VAMDC. In the import file, DOI is mandatory as we want to enforce the good practice of using unique identifiers. The references' entries of the BASECOL database are currently updated with their DOI. The keywords metadata are composed of a key metadata followed by its value. The key metadata are "Perturbing Element", "Target Element", "Possible systems", "Transitions", "Type of data", "Possible Method", and "Miscellaneous". To each reference, a set of "key metadata:value" is attached that characterizes the content of the corresponding paper. These keywords allow the search by keywords in the "Articles" part of the BASECOL website (cf. Section 4).

---

[17] We are considering re-offering this service through subscription.

Table 6. List of properties/metadata for the object "Publications". The symbols [M] and [O] indicate that the metadata are either mandatory or optional in the import file. In the column "Uniq.", we indicate whether the metadata define the uniqueness of the object. In the column "Purpose", we indicate whether the information is meant for the BASECOL (B) public WebUI display and/or for VAMDC (V) interoperability.

Property Name	Description	Type	Uniq.	Purpose
authors	authors of the article	[M]		V/B
title	title of the article	[O]		V/B
journal	acronyms [a]	[M]	yes	V/B
year	publication year of the article	[M]	yes	V/B
volume	publication volume of the article	[M]	yes	V/B
pages	publication pages of the article	[M]	yes	V/B
month	publication month of the article	[O]		V/B
issue	publication issue of the article	[O]		B
url	webpage linked to the article	[O]		V/B
DOI	DOI of the article	[M] [d]	note	V/B [b]
keywords	tag [c] to allow search by themes	[O]		B
mainArticle	flag	[M]		B

[a] the acronyms follow the convention of Astrophysics Data System (ADS); [b] the BASECOL node software is currently updated to provide the DOI. [c] those keywords are chosen on-the-fly through the import script. [d] if it does not exist, we create a DOI (see text). note: the DOI is not used for uniqueness because some old papers have an in-house random DOI.

It should be stressed that the publication part of the BASECOL database is extremely important because it ensures provenance of the data, and it ensures that BASECOL follows FAIR principles [7] (cf. Section 6), which is an intrinsic feature in VAMDC.

In addition, the BASECOL policies ask the users to cite the authors of the collisional datasets. Some improvement of that section could be made possible in the future by deploying a software similar to the one developed for the HITRAN and the AMBDAS databases [18] that allows for retrieving the information using a DOI.

3.2. Import Procedure

The import script performs integrity and format checks on the import ASCII file. The importation process starts when the import file is valid. The import script is designed either to import a whole new "collisional data set" or to update an existing "collisional data set". In each "Data" section of the "collisional data set" (Section 3.1), the import script checks the values of the metadata related to the uniqueness of an "object", and, in addition, for the "Energy Table", "Energy Origin", "Rate Coefficients", and "Fitting Coefficients" sections, it checks the values of the numerical tables. If the value of a single item that we call "uniqueness item", metadata, or numerical value does not correspond to what is already in the database, the "collisional data set" is considered as a new collisional data set; if all the uniqueness items have the same value as in the database, the " collisional data set" is considered as an update of an existing "collisional data set".

When a new "collisional data set" is imported in the database, a collisional entry is created. In order to ensure data-traceability and reproducibility, a version number is attributed to the freshly created set and a comment describing the details of this creation operation is stored. Alternatively, when an existing "collisional data set" is updated, the metadata not linked to the uniqueness concept and affected by the change are modified, a new version number of the "collisional data set" is created, and a comment describing the type of change is stored. Detailed information on the versioning is provided in Appendix B.

Once the import file has been processed, its scientific content, i.e., the "collisional data set", is copied into the database of the BASECOL "ingestion instance". This "collisional data set" is visible on the website of the BASECOL "ingestion instance", and the administrator can check the imported

data. Once the information is checked, the administrator allows its visibility in the "production instance" of BASECOL. Indeed, by default, any new "collisional data set" status is "non visible" in the production database, thus protecting the "production instance" of the database from non checked information when the "ingestion" database is dumped into the "production" database. The visibility on the "production instance" website of BASECOL implies that VAMDC can access the "collisional data set".

The processed import ASCII file is additionally copied into a git versioning system. The git versioning system allows traceability in case of a later update and allows restoration of the database in case of corruption and/or loss of the database. This process guarantees the data integrity.

4. Public Access to BASECOL via the Web User Interface (WebUI) at basecol.vamdc.org

The public interface of BASECOL has been simplified, and we describe here its current features. The "Home" section is the landing page for the URL: https://basecol.vamdc.org; it provides the citation policies and the units. There are two ways to query the database: a guided query with the "Browse Collisions" (see Section 4.1.1) section that corresponds to a simplified version of BASECOL2012 collisional query browser and the "Search Collisions" (see Section 4.1.2) that provides queries by several free search criteria. The "Search articles" section (cf. Figure 3) provides a search on the bibliographic entries which are associated with numerical data. The "Tools" section gives access to information on the SPECTCOL tool [4] and on a scientific package for the water-H_2 collisions [19–24]. The search on energy tables has been removed following recommendations from the French PCMI community[18].

Figure 3. Query Interface for the bibliography associated with the numerical data. There is a search by year, journal, authors, and keywords. The search by keywords is possible because of the "keywords" metadata of Table 6.

4.1. Query Part

4.1.1. Browse Collisions

The query part of the "Browse Collision" section is a four-step process: from a single page, the user successively chooses a target, its symmetry, then a collider and its symmetry. The interface provides an auto-completion for the species name and can suggest available colliders according to the selected target. The available datasets corresponding to the restrictions submitted by the user are listed immediately below the query section. Figure 4 shows an example for the request of CO colliding with ortho–H_2. In this example, we see that the data set of Yang et al. [25] has the flag recommended = "yes" as it is more complete and more recent than the previous sets [26,27] which have recommendation = "no".

---

[18] https://www.pcmi.cnrs.fr/

Figure 4. Query Interface with "Browse Collision" for collisional rate coefficients of CO–ortho-H_2.

4.1.2. Search Collisions

The query part of the "Search Collision" section allows for freely querying the database with the year of publication and the authors of paper attached to the datasets, the target and collider species. Therefore, it is possible to see the complete content of the database when no criteria are chosen, it is possible to see all collisional sets with electrons as the collider, and so on. This is an extremely useful tool for both the users and the managers of the database. We provide an example in Figure 5.

Figure 5. Query Interface with "Search Collision". It should be noted that it is not necessary to write the full surname of the authors. The retrieved data correspond a H_2O-electron data sets by Faure et al. [14].

4.2. Returned Information

4.2.1. General Display

Once the user has selected a collisional data set, the user interface displays on the same single page the complete numerical and documentary information. Figure 6 shows the return page for the selection of Yang et al. [25] data set. From this page, one can access the rate coefficients, the fits to the rate coefficients, the energy tables with state labels that are used to identify states in the rate coefficient tables, the bibliographic references attached to the collisional data set (the main reference where the

dataset is published is shown in red), and then a set of information that allows for rapidly characterizing the methodology used in the calculations. In BASECOL2012 [5] user interface, those items were split between several tabs.

Rotational de-excitation of CO by ortho-H$_2$ (j=1), 1K < T < 3000K - 41 levels (Yang et al., 2010)

Presentation

Rotational state-to-state rate coefficients are obtained among the 41 lowest rotational levels of CO for temperature between 1K and 3000K, and for o-H$_2$(j=1). The calculations include the j(H$_2$)=3 level, therefore other state-to-state rate coefficients must have been obtained corresponding to a transition from j(H$_2$)=1 to j(H$_2$)=3. Those state-to-state rate coefficients are not provided by the authors. The 4D PES of Jankowski et al, 2005 is used as in Wernli et al, 2006.

Coefficients

Rate

Display rate coefficients as :

- **HTML**
- **Text**

Fit

No fit coefficient for this collision

Data information

- CO initial level labelled from 2 to 41.
- CO final level labelled from 1 to 40.
- H2 initial level labelled from 1 to 1.
- H2 final level labelled from 1 to 1.
- 25 temperatures between 2 K and 3000 K.
- Units : $cm^3 s^{-1}$

Labelling Energy Tables

CO

- Basecol data :
  View -Text version

H2

- Basecol data :
  View -Text version

History

- Version 1 - 2017-10-12 - Collision added into the database

(version note : First creation)

References

NB : the main reference is displayed in red

- Yang,B.; Stancil,P. C.; Balakrishnan,N.; Forrey,R.~C.. **Rotational Quenching of CO due to H$_2$ Collisions.** *ApJ* Vol.718 : p1062-1069, 2010, August
- Wernli,M.; Valiron,P.; Faure,A.; Wiesenfeld,L.; Jankowski,P.; Szalewicz,K.. **Improved low-temperature rate constants for rotational excitation of CO by H$_2$** *A&A* Vol.446 : p367-372, 2006, January
- Jankowski,P.; Szalewicz,K.. **A new ab initio interaction energy surface and high-resolution spectra of the H$_2$-CO van der Waals complex** *JChPh* Vol.123 : p4301-+, 2005, September

Range Of Energy

Total energy between 10^{-5} and 10000 cm^{-1}.

Basis Set

UP to 50 levels of CO are included. For o-H$_2$, j=1, 3 are included.

Method

Close coupling calculations below total energy of 800cm-1 and Coupled states calculations above
No references

Precision

PES

The PES of Jankowski et al. 2005 is used. This surface has been obtained with CCSD(T) method using correlation-consistent basis sets complemented by bonding finctions and basis set extrapolation. It has been obtained by averaging over the intramolecular ground state vibration of H$_2$ with CO fixed at its vibrational ground-state averaged geometry.

- Jankowski,P.; Szalewicz,K.. **A new ab initio interaction energy surface and high-resolution spectra of the H$_2$-CO van der Waals complex** *JChPh* Vol.123 : p4301-+, 2005, September

Figure 6. Return page containing the CO–ortho-H$_2$ data set when the reference of Yang et al. [25] is selected. In a future release we will simplify the interface so that the sections "Range of Energy", "Basis Set", "Method" be combined into a single item called "Methodology".

The user can visualize the rate coefficients and energy tables on the WebUI and can download the data as text files. When fitting coefficients are available, a new graphical display allows for comparing the calculated rate coefficients with the fitting function.

4.2.2. History and Versioning

In the "History" part of the return page, the different versions of the collisional datasets will be available. At the time of publication, none of the sets have been updated since the creation of the

versioning system. Therefore, we show how it works in Figure 7. The versions correspond to major changes in the database (numerical data), and each version has a creation date and might have some minor modifications done at a later date. The currently displayed version is explicitly notified to the user by the "(this page)" tag after the corresponding version number. Clicking on any of the other versions will bring a similar return page (such as Figure 6) corresponding to the selected version. That history part reflects the complex versioning system described in this paper and is an important achievement as it allows full traceability and reproductibility of the data.
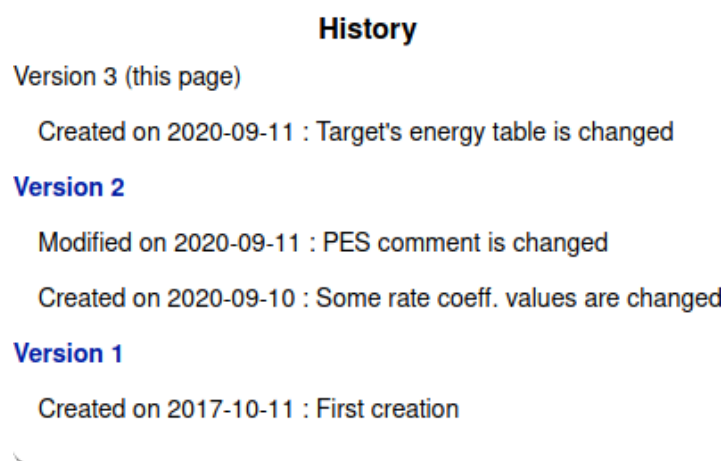
History

Version 3 (this page)

     Created on 2020-09-11 : Target's energy table is changed

Version 2

     Modified on 2020-09-11 : PES comment is changed

     Created on 2020-09-10 : Some rate coeff. values are changed

Version 1

     Created on 2017-10-11 : First creation

Figure 7. View of the history part when three citable versions are stored in the database.

5. VAMDC Access to BASECOL

The BASECOL database is interoperable within the VAMDC ecosystem[19], and the interoperability is ensured by the metadata associated with the imported data (cf. Section 3.1), and by the implementation of VAMDC software configured for the BASECOL database data model.

5.1. BASECOL Implementation of VAMDC Standards

By implementing the node software[20], the BASECOL database becomes a VAMDC-federated resource. BASECOL is currently the sole VAMDC connected database that implements the Java version of the node software[21]. The node software is a wrapping between the internal BASECOL structure and the VAMDC standards: it implements a data access which is a variant of the TAP protocol from the IVOA[22], deals with the query language [23] used in the VAMDC e-infrastructure, returns data in the XSAMS format[24], and uses standardized field names described in a dictionary: the "Restrictables"[25] contain the list of field names that are used to filter the query to the database and the "Returnables"[26] refer to the type of data that the database returns to VAMDC. The current BASECOL "Restrictables" and "Returnables" are reported in Tables 7 and 8.

---

19 http://www.vamdc.org/activities/research/
20 http://www.vamdc.eu/documents/nodesoftware/concepts.html
21 http://www.vamdc.org/activities/research/software/java-nodesoftware/
22 https://standards.vamdc.eu/dataAccessProtocol/index.html
23 https://standards.vamdc.eu/queryLanguage/index.html
24 https://standards.vamdc.eu/#data-model
25 https://standards.vamdc.eu/dictionary/restrictables.html
26 https://standards.vamdc.eu/dictionary/returnables.html

Table 7. List of BASECOL Restrictables metadata on the date of submission.

Category: List of Returnables
Atom: AtomSymbol, IonCharge
Molecule: MoleculeStoichiometricFormula, InchiKey, MoleculeStateNuclearSpinIsomer, StateEnergy
Particule: Particule name
Source: SourceYear
Collision data: CollisionCode, EnvironmentTemperature

Table 8. List of BASECOL Returnables metadata on the date of submission.

Returnables	Category in XSAMS
MethodCategory, MethodDescription, MethodID	Method
CollisionComment, CollisionID, CollisionCode, CollisionRef	Collision
CollisionMethod	Collision
SourceAuthorName, SourceCategory, SourceComments	Source
SourcePageBegin, SourcePageEnd	Source
SourceID, SourceName, SourceTitle, SourceURI	Source
SourceVolume, SourceYear	Source
AtomSpeciesID, AtomSymbol, AtomNuclearCharge	Atom
AtomInchiKey, AtomIonCharge	Atom
AtomMassNumber, AtomMass	Atom
AtomStateEnergy, AtomStateTotalAngMom	Atom
AtomStateRef, AtomStateID	Atom
MoleculeSpeciesID, MoleculeInchi, MoleculeInchiKey	Molecule
MoleculeChemicalName, MoleculeOrdinaryStructuralFormula	Molecule
MoleculeStoichiometricFormula	Molecule
MoleculeMolecularWeight, MoleculeComment	Molecule
MoleculeStateID, MoleculeStateDescription	Molecule
MoleculeStateEnergy, MoleculeStateEnergyOrigin	Molecule
MoleculeStateNSIName, MoleculeStateNSILowestEnergyStateRef	Molecule
MoleculeStateAuxillary	Molecule

Finally, the database must be registered in the VAMDC registry[27]. The VAMDC registry is a database of metadata describing the VAMDC nodes and web applications/services. These metadata are collected by querying directly the services' endpoints. The information stored into the registries allows humans and/or machines to find the address of a given node, in order to select nodes by the kinds of data they offer and to find out which query terms are supported at a node. The parameters of Tables 7 and 8 can be found in the registries description of BASECOL[28].

Members of the VAMDC Consortium have a password allowing them to register a new database or a web application, once they have followed the quality procedures concerning the resource to be registered.

5.2. BASECOL and VAMDC Ecosystem

Within the VAMDC ecosystem, BASECOL can be accessed from different VAMDC tools such as the VAMDC portal[29] and the SPECTCOL tool [4], and from any other software that has implemented the VAMDC protocols and standards. On the VAMDC portal, the BASECOL data may be displayed with the "Collisional XSAMS to HTML" display, and Figure 8 gives an example of output that corresponds to querying the CO molecule on the VAMDC portal without any selection of the collider. When the BASECOL data are queried through the VAMDC portal and the SPECTCOL tool, each query is stored in

---

27 https://standards.vamdc.eu/registry/index.html
28 http://registry.vamdc.eu/registry-12.07/main/viewResourceEntry.jsp?IVORN=ivo%3A%2F%2Fvamdc%2Fbasecol%2Fvamdc-tap_12.07
29 https://portal.vamdc.eu

the *VAMDC query store*[30] and can be linked to a DOI. When this DOI is used in publication, the citations associated with the datasets are acknowledged, therefore increasing the impact factor of the data producers [28–30].

Results from BASECOL database VAMDC node

1: Rotational de-excitation of CO by para-H$_2$ (j=0), 5K < T < 70K - 6 levels (Wernli et al., 2006)
Rate Coefficients
Energy Table of CO
Energy Table of H2
2: Rotational excitation of CO by ortho-H$_2$ (Flower, 2001)
3: Rotational excitation of CO by para H$_2$ (Flower, 2001)
4: Rotational de-excitation of CO by H for 5K < T< 100K (Balakrishnan & al, 2002)
5: Vibrational de-excitation of CO by H (Balakrishnan & al, 2002)
6: Rotational de-excitation of CO (v=0) by He (Cecchi-Pestellini & al, 2002)
7: Vibrational de-excitation of CO by He (Cecchi-Pestellini & al., 2002)
8: Rotational de-excitation of CO by H for 100K< T <3000K (Balakrishnan & al, 2002)
9: Rotational de-excitation of CO by ortho-H$_2$ (j=1), 5K < T < 70K - 6 levels (Wernli et al., 2006)
10: Rotational de-excitation of CO by para-H$_2$ (j=0), 1K < T < 3000K - 41 levels (Yang et al., 2010)

Sources

Id	Title	Origin	Authors	Year	Link
BBAS920	A new ab initio interaction energy surface and high-resolution spectra of the H$_2$-CO van der Waals complex	journal:jcp,Vol:123,Page Begin:4301	P. Jankowski;K. Szalewicz;	2005	http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2005JChPh.123j4301J&db_key=PHY
BBAS0		database:BASECOL database,Vol:,Page Begin:	M.-L. Dubernet;	2020	http://basecol.obspm.fr
BBAS921	Improved low-temperature rate constants for rotational excitation of CO by H$_2$	journal:aap,Vol:446,Page Begin:367	P. Valiron;P. Jankowski;K. Szalewicz;A. Faure;L. Wiesenfeld;M. Wernli;	2006	http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2006A%26A...446..367W&db_key=AST

Rate Coefficients of Rotational de-excitation of CO by para-H$_2$ (j=0), 5K < T < 70K - 6 levels (Wernli et al., 2006)

Unselect all	I1	I2	F1	F2	5.0	10.0	20.0	30.0	40.0	50.0	60.0	70.0
☑	2	1	1	1	3.24E-11	3.31E-11	3.26E-11	3.27E-11	3.3E-11	3.35E-11	3.39E-11	3.44E-11
☑	3	1	1	1	2.92E-11	3.05E-11	3.12E-11	3.15E-11	3.17E-11	3.18E-11	3.19E-11	3.2E-11
☑	3	1	2	1	7.93E-11	7.21E-11	6.46E-11	6.12E-11	5.98E-11	5.94E-11	5.97E-11	6.03E-11
☑	4	1	1	1	4.83E-12	5.21E-12	5.56E-12	6.02E-12	6.56E-12	7.14E-12	7.74E-12	8.35E-12
☑	4	1	2	1	4.75E-11	5.01E-11	5.09E-11	5.09E-11	5.08E-11	5.07E-11	5.08E-11	5.09E-11
☑	4	1	3	1	8.98E-11	7.95E-11	7.06E-11	6.7E-11	6.56E-11	6.52E-11	6.54E-11	6.6E-11
☑	5	1	1	1	2.8E-12	3.11E-12	3.42E-12	3.66E-12	3.86E-12	4.05E-12	4.22E-12	4.38E-12
☑	5	1	2	1	7.95E-12	9.59E-12	1.08E-11	1.18E-11	1.28E-11	1.38E-11	1.49E-11	1.61E-11
☑	5	1	3	1	5.9E-11	6.2E-11	6.22E-11	6.12E-11	6.04E-11	5.98E-11	5.94E-11	5.92E-11
☑	5	1	4	1	8.49E-11	8.22E-11	7.51E-11	7.13E-11	6.94E-11	6.87E-11	6.86E-11	6.89E-11
☑	6	1	1	1	8.17E-13	9.71E-13	1.19E-12	1.39E-12	1.61E-12	1.84E-12	2.1E-12	2.37E-12
☑	6	1	2	1	6.16E-12	6.22E-12	6.51E-12	6.81E-12	7.12E-12	7.42E-12	7.72E-12	8.02E-12
☑	6	1	3	1	9.86E-12	1.25E-11	1.47E-11	1.6E-11	1.71E-11	1.82E-11	1.94E-11	2.06E-11

Figure 8. VAMDC portal display of BASECOL data.

6. BASECOL and FAIR Principles

We would like to conclude this paper with a discussion about the compliance of BASECOL with the FAIR principles [7].

- Findable: On one hand, as described in Section 5, BASECOL is registered into public registries and may be found via the VAMDC discovering facilities. On the other hand, once BASECOL is discovered as a useful data repository, the collisional data it contains may be searched (cf. Section 4.1) using the collision search engine or browsed using the collisional browser.
- Accessible: Once found, data may be accessed directly on the Basecol Web site (cf. Section 4.2) or extracted by the VAMDC infrastructure.
- Interoperable: the BASECOL interoperability comes from its integration within the VAMDC infrastructure. BASECOL data extracted via VAMDC may be handled using VAMDC compliant tools and processors (cf. Section 5.2).
- Reusable: Trust in data is a crucial element for data reusability. With BASECOL, scientists are able to cite where data are coming from, with a fine-grained granularity that allows to achieve scientific reproducibility. Indeed data are both versioned and timestamped, and a user may refer to a specific version. All the "collisional data set" versions are stored in the BASECOL database and are available via the web interface. Another aspect that increases the trust in the BASECOL data is the direct link between a given collisional set and the refereed publication where the data set was published: the methods, hypothesis, and algorithms used for producing a given data set always come with the data.

7. Conclusions

This paper provides a detailed analysis of the latest technical developments on the BASECOL database and its environment. The major upgrades concern its very careful versioning system linked to an ingestion system that verifies the imported data with respect to the content of the database, as well as the possibility for users to retrieve previous versions of the collisional data sets. Another major

---

30   https://cite.vamdc.eu

update is the dual system of a validation and of a production environment coupled to the storage of the collisional data sets ascii files in a github repository. BASECOL is now designed to be a database fully compliant with the VAMDC e-infrastructure, in particular through the import script that contains the parameters necessary for the VAMDC interoperability. Citing one referee: "The 2020 version is undoubtedly a major update compared to the 2012 version: simpler, clearer with excellent traceability and reproducibility. The website is well designed and user friendly." Nevertheless, some issues must still be solved, they concern a more friendly import management environment and, of course, the scientific update of the database. In the future, we might foresee partnerships with websites that provide the above cited RADEX files, and we might adopt a different business model for the sustainability of the database.

Glossary

Spring	Java framework to build a web application (https://spring.io/)
React	Javascript library to build a web interface (https://reactjs.org/)
MariaDB	Open source relational database system (https://mariadb.org/)

Abbreviations

The following abbreviations are used in this manuscript:

API	Application Programming Interface
DOI	Digital Object Identifier
FAIR	Findability, Accessibility, Interoperability, and Reuse
HTTP	Hypertext Transfer Protocol
IOS	Infinite Order Approximation
IVOA	International Virtual Observatory Alliance
JDBC	Java Database Connectivity
JSON	JavaScript Object Notation
PES	Potential Energy Surface
RDBMS	Relational Database Management System
REST	Representational state transfer
SAMP	Simple Application Messaging Protocol
TAP	Table Access Protocol
VAMDC	Virtual Atomic and Molecular Data Center
VAMDC-TAP	VAMDC Tap Access Protocol
XSAMS	XML Schema for Atoms, Molecules and Solids

Appendix A. Examples or Templates of the Import File Sections

The import ASCII file is composed of several "Data" sections. Examples or templates of those sections are provided in this appendix.

Table A1 provides an example of an "Element" section input for N_2H^+; it corresponds to the metadata of Table 1.

Table A1. Example of a "Element" section input for N_2H^+ as requested by the import script. The metadata are in red and the values in blue. The metadata "elementType" have fixed values chosen among [molecule | atom | negative molecular ion | positive molecular ion | particle]. The metadata \$\$mass give the molecular weight in atomic mass unit (in the species database, it is used to infer the mass number). The metadata "molecularConstant" is historical, concerns the rotational constant of a linear molecule, and has units in MHz. "htmlName" uses normal HTML language and is used in the BASECOL web interface.

```
$$name = N2H+
$$inchikey = JGRMHOSHXDMSA-UHFFFAOYSA-O
$$inchi = InChI=1S/N2/c1-2/p+1
$$stoichiometricFormula = HN2+
$$htmlName = N<sub>2</sub>H<sup>+</sup>
$$latex = N$_2$H$⁺$
$$mass = 29.01
$$molecularConstant = 46586.87
$$charge = 1
$$elementType = molecule
```

Table A2 provides a template of an "Energy table" input; it includes the metadata of Table 2.

Table A2. Template of an energy table input as requested by the import script. The metadata are in red and the values in blue. The "case" metadata are dcs | hunda | hundb | ltcs | nltcs | stcs | lpcs | asymcs | asymos | sphcs | sphos | ltos | lpos | nltos. Q1 and Q2 are characters for quantum numbers. Both the case and the quantum numbers are described in the case-by-case document (http://www.vamdc.eu/documents/cbc-1.0/). In the second part of the table, all fields are mandatory. The first column gives the label attached to the energy level; this label identifies the levels in the rate coefficient table. The number of levels must be the number of levels necessary to identify the energy levels of the rate coefficients table. The second column gives the energy (in cm^{-1}), the two following columns give the values of quantum numbers identified in the parameter \$\$quantumNumbers. In the current template, there are two quantum numbers and two levels.

```
$$case = case (see caption)
$$symmetry= none | A | E | meta | ortho | para (for atoms it is "none")
$$title = Free Text (for example: Rotational energy levels of HCN)
$$comments = Free Text
$$type = theoretical | experimental
$$energyUnit = cm⁻¹ | Kelvin
$$quantumNumbers = Q1; Q2
              #                          [M]
              1   value   value   value  [M]
              2   value   value   value  [M]
```

Table A3 provides a template of an "Energy Origin" input; it includes the metadata of Table 3.

Table A3. Template of an "Energy origin" section input as requested by the import script. The metadata are in red and the values in blue. The "case" metadata are dcs | hunda | hundb | ltcs | nltcs | stcs | lpcs | asymcs | asymos | sphcs | sphos | ltos | lpos | nltos. Q1 and Q2 are characters for quantum numbers. Both the case and the quantum numbers are described in the case-by-case document (http://www.vamdc.eu/documents/cbc-1.0/). In the second part of the table, all fields are mandatory. The logic of the numerical line is the same as in Table A2.

$$case = case (see caption)
$$symmetry = none | A | E | meta | ortho | para (for atoms it is "none")
$$quantumNumbers = Q1; Q2
```
        #                        [M]
          1   0.0   value   value   [M]
```

Table A4 provides a template of a "Rate Coefficients" input; it includes the metadata of Table 4.

Table A4. Template of a "Rate Coefficients" input as requested by the import script. The metadata are in red and the values in blue. In the second part of the table, all fields are mandatory. The first column corresponds to the target's initial level, the second column to the target's final level, the third column to the collider's initial level, and the fourth column to the collider's final level. Then, the rate coefficients at the different temperatures indicated as values of the parameter $$temperatures are provided. In the current template, the transition is between the third and the first rotational levels of a target with the collider in its fundamental state. Two temperatures are considered at T1 and T2. The values 0.130607E-10 and 0.161225E-10 are the rate coefficients in $cm^3 s^{-1}$.

$$title = Free text (Author1 & al, year)
$$comment = Free text
$$pesComment = Free text
$$methodComment = Free text
$$precision = Free text
$$rateType = state to state | effective | thermalized
$$process = none | fine | hyperfine | ro-torsional | ro-vibrational | rotation | vibration
$$year = value
$$contributor = Free text
$$temperatures = T1 ; T2; .. (for example: 5; 10)
```
        #                                        [M]
          3   1   1   1   0.130607E-10   0.161225E-10   [M]
```

Table A5 provides a template of a "Fitting Coefficients" input; it includes the metadata of Table 5.

Table A5. Template of a "Fitting coefficients" input as requested by the import script. The metadata are in red and the values in blue. None of the information is requested by VAMDC. The metadata $$designation has the following choices explained in the text (common fit equation | Faure et al. 2004 | Lim et al. 1999 | Faure et al. 2001 | Sarpal et al. 1993 | Not available). In the second part of this entry, the first four columns indicate the energy level's labels in the same way as as in the rate coefficient table (see Table A4); this is followed by the values of fitting coefficients whose symbols are indicated in the parameter $$coefficients. [M] stands for mandatory. In the current template, the transition is between the third and the first energy levels of the target with the collider in its fundamental state, and the fitting functions have two coefficients.

$$title = Free text
$$function = [latex] mathematical expression of the fitting function[/latex]
$$comment = Free text
$$designation = choices
$$coefficients = coefficient1_symbol; coefficient2_symbol
```
# [M]
        3   1   1   1   coefficient1_value   coefficient2_value   [M]
```

Appendix B. Versioning System

Appendix B.1. Introduction

The BASECOL versioning information is stored in a "Changelogs" table that lists the new additions or modifications brought to the database. Each entry of the "Changelogs" table refers to a given version of an "object" and is timestamped with the timestamps indicated in the "Versions" table. Five types of versioned "objects" are defined: "Element" (Table 1), "Energy Table" (Table 2), "Rate Coefficients" (Table 4), "Fitting Coefficients" (Table 5) and "Publication" (Table 6). The versioning of the "Energy Origin" object (Table 2) is considered as part of the versioning of the "Energy Table" object, since a different origin changes the meaning of an energy table.

The «Changelogs» table contains 10 columns:

1. IdVersion: a micro version ID, linked to a timestamp in the « Versions » table.
2. IdGlobalVersion: a global version ID. Objects belonging to the same version of the database share the same global version id. This id is timestamped in the Versions table.
3. IdVersionedElement: ID of the versioned object
4. IdLinkedElement: ID that refers to a versioned object. In BASECOL the modifications of an object are classified as "minor" (corresponding to the optional fields of the input files) and "major" when the changes occur on fields that define the uniqueness unicity of the object (the tables numerical values and some of the mandatory fields of the input files). In the case of a major change, a new entry (with a new ID) is created in the table of the concerned object. To maintain tracing of that major change, we create a relationship between the old and the new object using the "IdLinkedElement" that allows for storing the ID of the old object and linking it to the new one. For example, if the element "EnergyTable1" with an ID = 1 undergoes a major change, an element "EnergyTable2" is created with an ID = 2, and, in the "Changelogs" table, we put the number 2 in "idVersionedElement" and 1 into "idLinkedElement".
5. ElementName: the type of the versioned object
6. Action: The nature of the action (modification of existing data or a new addition)
7. Description: describes what has changed (automatically generated from the modification history)
8. Comment: comments given by the administrator (free text)
9. CitableVersion: a user-friendly version number compiled from the idVersion; it is used to identify the version of a collisional set and is aimed at public display (for example if a table is inserted with IdVersion = 450 and the first update occurs with IdVersion = 570, then citableVersion = 1 for idVersion = 450 and citableVersion = 2 for idVersion = 570).

We describe below how the versioning procedure behaves when changes occur in the database.

Appendix B.2. A New "Collisional Data Set" Is Created from Scratch

We re-call that a "collisional data set" corresponds to a complete import file as described in Section 3.1.

- How is the database version assigned? We create a new global version of the DB for every new dump into the production database (timestamp of the day). Before this moment, this field is set to NULL in the ingestion database.
- How is the "Collision" version assigned? A new entry linked to the Collision is created in the "Changelogs" table. A new timestamp is created in the "Versions" table and linked to "Changelogs". In addition, the "citableVersion" in "Changelogs" table is incremented for every later update, if the user wishes to track it.

Appendix B.3. The "Element" Object of the Target/Collider Is Changed

- How is its version updated? If the change is a minor change, we keep the ID of the "element" object and create a timestamp linked to the "Element". Otherwise, we create a new ID, include a relationship between the new species ID and the old "Element" ID, and we create a timestamp linked to the new element.
- How are the versions of the other objects impacted? When a new ID of "Element" object is created, it involves the creation of a new "Energy Table" object and of a new "collisional data set".
- How are the versions seen by the users impacted? In case of major changes, the "citableVersion" of the "collisional data set" is incremented.

Appendix B.4. The "Energy Table" Object of the Target/Collider Is Changed

- How is its version updated? As it is a major change, both a new ID of the "Energy Table" object and a new timestamp related to the "Energy Table" object are created.
- How are the versions of the other objects impacted? We create a new "collisional data set" ID, a new timestamp related to the "collisional data set".
- How are the versions seen by the users impacted? The "citableVersion" of the "collisional data set" is incremented.

Appendix B.5. The "Rate Coefficients" Object Is Changed

- How is its version updated? When the "Rate Coefficients" object is changed, a new "collisional data set" ID is created with its own timestamp and with a link to the previous version of the "collisional data set".
- How are the versions seen by the users impacted? The "citableVersion" of the "collisional data set" is incremented.

Appendix B.6. The "Fitting Coefficients" Object Is Changed

- How is its version updated? If the change is minor, a timestamp linked to the "Fitting Coefficients" object is created. If the change is major, a new ID of the "Fitting Coefficients" object is created together with a new timestamp.
- How are the versions of the other objects impacted? no impact.
- How are the versions seen by the users impacted? no impact.

Appendix B.7. A "Publication" Object Is Changed

The "Publications" section might be composed of several "Publication" objects, each of them including the metadata of Table 6. Each "Publication" object is versioned.

- How is a "Publication" object version updated? When the change is minor, the ID of the "Publication" object is left unchanged. When the change is major, a new entry is created in the publication table with its own ID. In that case, a timestamp linked to the "Publication" object is created in "Changelogs".
- How are the versions seen by the user updated? If the "Publication" object is linked to a "Rate Coefficients" object, and, if the change is major, the "citableVersion" of the "collisional data set" is incremented.
- Are there impacts on the other versions ID? yes, if the "Publication" object is linked to a "Rate Coefficients" object or to an "Energy Table" object and, if the change is major, a new "collisional data set" ID; yes, if the "Publication" object is linked to an "Energy Table" object and, if the change is major, a new "Energy Table" ID is created.

Appendix C. WebUI for the Creation of the Import Files

This appendix shows two sections of the WebUI that create import files. This WebUI is private to the group that prepares the import datafile. This user interface allows for extracting information from

the BASECOL database, in order to modify and to add information, and then to create an import file in a flawless format that can be treated by the import script. This is particularly useful for the retrieval of metadata associated with species, for the discovery of the right format for quantum numbers, and in general for the retrieval of any information from the BASECOL database in order to create a new import file. Figure A1 shows an example for "species": a species is selected in the auto-completion bar of "Available elements", and then the fields are automatically filled. Figure A2 shows an example for "quantum numbers": a case is chosen; then, the list of standardized VAMDC quantum numbers appears, and the administrator can choose the relevant ones for his/her application.

Figure A1. Extract from the webUI for the creation of an "Element" input file (Table A1). In this example, the information related to the H_2O species is retrieved from BASECOL.

Figure A2. Extract from the WebUI for the creation of an "Energy Table" input (Table A2). In this example, the information related to the case "dcs" (diatomic close shell molecules) is retrieved from BASECOL. On the right of the "Quantum numbers" box, a list of quantum numbers appears. The administrator chooses the relevant quantum numbers and includes them in the box.

References

1. Dubernet, M.L.; Boudon, V.; Culhane, J.L.; Dimitrijevic, M.S.; Fazliev, A.Z.; Joblin, C.; Kupka, F.; Leto, G.; Le Sidaner, P.; Loboda, P.A.; et al. Virtual atomic and molecular data centre. *J. Quant. Spectrosc. Radiat. Transf.* **2010**, *111*, 2151–2159. [CrossRef]
2. Dubernet, M.L.; Antony, B.K.; Ba, Y.A.; Babikov, Y.L.; Bartschat, K.; Boudon, V.; Braams, B.J.; Chung, H.K.; Daniel, F.; Delahaye, F.; et al. The virtual atomic and molecular data centre (VAMDC) consortium. *J. Phys. B* **2016**, *49*, 074003. [CrossRef]
3. Albert, D.; Antony, B.K.; Ba, Y.A. The VAMDC SPECTCOL Tool. Unpublished work, 2020.

4.　Dubernet, M.L.; Nenadovic, L. *SPECTCOL: Spectroscopic and Collisional Data Retrieval*; Record ascl:1111.005; Astrophysics Source Code Library: Houghton, MI, USA, 2011. Available online: http://ascl.net/1111.005 (accessed on 7 October 2020).

5.　Dubernet, M.L.; Alexander, M.H.; Ba, Y.A.; Balakrishnan, N.; Balança, C.; Ceccarelli, C.; Cernicharo, J.; Daniel, F.; Dayou, F.; Doronin, M.; et al. BASECOL2012: A collisional database repository and web service within the Virtual Atomic and Molecular Data Center (VAMDC). *Astr. Astroph.* **2013**, *553*, A50. [CrossRef]

6.　Van der Tak, F.F.S.; Lique, F.; Faure, A.; Black, J.H.; van Dishoeck, E.F. The Leiden Atomic and Molecular Database (LAMDA): Current Status, Recent Updates, and Future Plans. *Atoms* **2020**, *8*, 15. [CrossRef]

7.　Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]

8.　Taylor, M.B.; Boch, T.; Taylor, J. SAMP, the Simple Application Messaging Protocol: Letting applications talk to each other. *Astron. Comput.* **2015**, *11*, 81–90, [CrossRef]

9.　Vastel, C.; Bottinelli, S.; Caux, E.; Glorian, J.M.; Boiziot, M. CASSIS: A tool to visualize and analyse instrumental and synthetic spectra. In *Proceedings of the SF2A-2015: Annual Meeting of the French Society of Astronomy and Astrophysics*; Martins, F., Boissier, S., Buat, V., Cambrésy, L., Petit, P., Eds.; SF2A: Paris, France, 2015; pp. 313–316.

10.　Endres, C.P.; Schlemmer, S.; Schilke, P.; Stutzki, J.; Mueller, H.S.P. The Cologne Database for Molecular Spectroscopy, CDMS, in the Virtual Atomic and Molecular Data Center, VAMDC. *J. Mol. Spect.* **2016**, *327*, 95–104. [CrossRef]

11.　Pickett, H.M.; Poynter, R.L.; Cohen, E.A.; Delitsky, M.L.; Pearson, J.C.; Muller, H.S.P. Submillimeter, millimeter and microwave spectral line catalog. *J. Quant. Spectrosc. Radiat. Transf.* **1998**, *60*, 883–890. [CrossRef]

12.　Hill, C.; Gordon, I.E.; Rothman, L.S.; Tennyson, J. A new relational database structure and online interface for the HITRAN database. *J. Quant. Spectrosc. Rad. Transf.* **2013**, *130*, 51–61. [CrossRef]

13.　Rothman, L.S.; Gordon, I.E.; Babikov, Y.; Barbe, A.; Chris Benner, D.; Bernath, P.F.; Birk, M.; Bizzocchi, L.; Boudon, V.; Brown, L.R.; et al. The HITRAN2012 molecular spectroscopic database. *J. Quant. Spectrosc. Rad. Transf.* **2013**, *130*, 4–50. [CrossRef]

14.　Faure, A.; Gorfinkiel, J.D.; Tennyson, J. Electron-impact rotational excitation of water. *MNRAS* **2004**, *347*, 323–333. [CrossRef]

15.　Faure, A.; Tennyson, J. Electron-impact rotational excitation of linear molecular ions. *MNRAS* **2001**, *325*, 443–448. [CrossRef]

16.　Sarpal, B.K.; Tennyson, J. Calculated vibrational excitation rates for electron-H_2^+ collisions. *MNRAS* **1993**, *263*, 909–912. [CrossRef]

17.　Lim, A.J.; Rabadan, I.; Tennyson, J. Electron-impact rotational excitation of CH^+. *MNRAS* **1999**, *306*, 473–478. [CrossRef]

18.　Skinner, F.M.; Gordon, I.E.; Hill, C.; Hargreaves, R.J.; Lockhart, K.E.; Rothman, L.S. Referencing Sources of Molecular Spectroscopic Data in the Era of Data Science: Application to the HITRAN and AMBDAS Databases. *Atoms* **2020**, *8*, 16. [CrossRef]

19.　Daniel, F.; Dubernet, M.L.; Grosjean, A. Rotational excitation of 45 levels of ortho/para-H_2O by excited ortho/para-H_2 from 5 K to 1500 K: State-to-state, effective, and thermalized rate coefficients. *Astron. Astrophys.* **2011**, *536*, A76. [CrossRef]

20.　Daniel, F.; Dubernet, M.L.; Pacaud, F.; Grosjean, A. Rotational excitation of 20 levels of para-H_2O by ortho-H_2 (j_2 = 1, 3, 5, 7) at high temperature. *Astron. Astrophys.* **2010**, *517*, A13. [CrossRef]

21.　Dubernet, M.L.; Daniel, F.; Grosjean, A.; Lin, C.Y. Rotational excitation of ortho-H_2O by para-H_2 (j_2 = 0, 2, 4, 6, 8) at high temperature. *Astron. Astrophys.* **2009**, *497*, 911–925. [CrossRef]

22.　Dubernet, M.L.; Daniel, F.; Grosjean, A.; Faure, A.; Valiron, P.; Wernli, M.; Wiesenfeld, L.; Rist, C.; Noga, J.; Tennyson, J. Influence of a new potential energy surface on the rotational (de)excitation of H_2O by H_2 at low temperature. *Astron. Astrophys.* **2006**, *460*, 323–329. [CrossRef]

23.　Valiron, P.; Wernli, M.; Faure, A.; Wiesenfeld, L.; Rist, C.; Kedzuch, S.; Noga, J. R12-calibrated H_2O-H_2 interaction: Full dimensional and vibrationally averaged potential energy surfaces. *J. Chem. Phys.* **2008**, *129*, 134306. [CrossRef]

24.　Faure, A.; Valiron, P.; Wernli, M.; Wiesenfeld, L.; Rist, C.; Noga, J.; Tennyson, J. A full nine-dimensional potential-energy surface for hydrogen molecule-water collisions. *J. Chem. Phys.* **2005**, *122*, 221102. [CrossRef]

25. Yang, B.; Stancil, P.C.; Balakrishnan, N.; Forrey, R.C. Rotational Quenching of CO due to H_2 Collisions. *Astrophys. J.* **2010**, *718*, 1062–1069. [CrossRef]
26. Wernli, M.; Valiron, P.; Faure, A.; Wiesenfeld, L.; Jankowski, P.; Szalewicz, K. Improved low-temperature rate constants for rotational excitation of CO by H_2. *Astron. Astrophys.* **2006**, *446*, 367–372. [CrossRef]
27. Flower, D.R. The rotational excitation of CO by H_2. *J. Phys. B* **2001**, *34*, 2731–2738. [CrossRef]
28. Zw'olf, C.M.; Moreau, N.; Ba, Y.; Dubernet, M.L. Implementing in the VAMDC the New Paradigms for Data Citation from the Research Data Alliance. *Data Sci. J.* **2019**, *18*, 4. [CrossRef]
29. Moreau, N.; Zwolf, C.M.; Ba, Y.A.; Richard, C.; Boudon, V.; Dubernet, M.L. The VAMDC Portal as a Major Enabler of Atomic and Molecular Data Citation. *Galaxies* **2018**, *6*, 105. [CrossRef]
30. Zwölf, C.M.; Moreau, N.; Dubernet, M.L. New model for datasets citation and extraction reproducibility in VAMDC. *J. Mol. Spectrosc.* **2016**, *327*, 122–137. [CrossRef]