Atomic and Molecular Databases Open Science for a sustainable world

M L Dubernet, G.B. Berriman, P. Barklem, K. Choi, A. Foster, I. Gordon, C. Hill, J. Kim, D.H. Kwon, H. Linnartz, et al.

IAU 371 Symposium
XXIXth IAU General Assembly, August 2015
D. Soderblom & G. Nave, eds.

DOI: 00.0000/X000000000000000X

Atomic and Molecular Databases Open Science for a sustainable world

M.L. Dubernet[1], G.B. Berriman[2], P. Barklem[3], K. Choi[4], A. Foster[5], I. Gordon[5], C. Hill[6], J. Kim[4], D.H. Kwon[7], H. Linnartz[8], Yu. Ralchenko[9], F. Salama[10], H. Shmagun[4], P. Schilke[11], D. Seo[4], J. Shim[12], S. Shin[4], M.-Y. Song[13], J. Tennyson[14], and C. Vastel[15]

[1]LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne University, 5 Place Janssen, 92190 Meudon, France
email: marie-lise.dubernet@observatoiredeparis.psl.eu

[2] Caltech/IPAC—NExScI, 1201 East California Blvd, Pasadena, CA. 91125, USA
email : gbb@ipac.caltech.edu

[3]Theoretical Astrophysics, Department of Physics and Astronomy, Uppsala University, Box 516, S 75120 Uppsala, Sweden; email: paul.barklem@physics.uu.se

[4] Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea
email : knchoi@kisti.re.kr, jaesoo@kisti.re.kr
email : hanna.shmagun@kisti.re.kr, dmseo@kisti.re.kr, maximus74.shin@gmail.com

[5]Harvard-Smithsonian Center for Astrophysics, 60 Garden St, Cambridge MA 02138, USA
email: afoster@cfa.harvard.edu, igordon@cfa.harvard.edu

[6] International Atomic Energy Agency, Wagramer Strasse 5, PO Box 100, Vienna A-1400, Austria; email: ch.hill@iaea.org

[7] Nuclear Data Center, Nuclear Physics Application Research Division, Korea Atomic Energy Research Institute, 111, Daedeok-daero 989 beon-gil, Yuseong-gu, Daejeon, 34057, Republic of Korea; email: hkwon@kaeri.re.kr

[8] Laboratory for Astrophysics, Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, the Netherlands; email : linnartz@strw.leidenuniv.nl

[9] National Institute of Standards and Technology Gaithersburg, MD 20899, USA
email : yuri.ralchenko@nist.gov

[10]Space Science and Astrobiology Division, NASA Ames Research Center, Moffett Field, California 94035, USA; email: farid.salama@nasa.gov

[11] I. Physikalisches Institut der Universitaet zu Koeln, Zuelpicher Str. 77, D-50937 Koeln, Germany; email: schilke@ph1.uni-koeln.de

[12] School of Computing, Korea Advanced Institute of Science and Technology, Republic of Korea; email : hl1iit@kaist.ac.kr

[13] Institute of Plasma Technology, Korea Institute of Fusion Energy(KFE), 37, Dongjangsan-ro, Gunsan, Jeollabuk-do, 54004, Republic of Korea
mysong@kfe.re.kr

[14] Department of Physics and Astronomy, University College London, London WC1E 6BT, UK
email: j.tennyson@ucl.ac.uk

[15] IRAP, Université de Toulouse, CNRS, CNES, UPS, (Toulouse), France
email : cvastel@irap.omp.eu

Abstract. The building of online atomic and molecular databases for astrophysics and for other research fields started with the beginning of the internet. These databases have encompassed different forms: databases of individual research groups exposing their own data, databases providing collected data from the refereed literature, databases providing evaluated compilations, databases providing repositories for individuals to deposit their data, and so on. They were, and are, the replacement for literature compilations with the goal of providing more complete and

in particular easily accessible data services to the users communities. Such initiatives involve not only scientific work on the data, but also the characterization of data, which comes with the "standardization" of metadata and of the relations between metadata, as recently developed in different communities. This contribution aims at providing a representative overview of the atomic and molecular databases ecosystem, which is available to the astrophysical community and addresses different issues linked to the use and management of data and databases. The information provided in this paper is related to the keynote lecture "Atomic and Molecular Databases: Open Science for better science and a sustainable world" whose slides can be found at DOI : doi.org/10.5281/zenodo.6979352 on the Zenodo repository connected to the "cb5-labastro" Zenodo Community (zenodo.org/communities/cb5-labastro).

Keywords. standards, atomic data, molecular data, laboratory astrophysics, experiment, theory, interdisciplinary studies, data network, data analysis

1. Introduction

Scientific research requires dissemination of findings and results, and today most of the exchanges take place electronically. So it requires scientists to adapt to conducting research with data that come in increasing quantities, varieties and modes of dissemination, often for purposes far more interdisciplinary than in the past. In addition, in many countries public funding organizations promote Open Science and the usage of various "technological Open Science" platforms. As a result, we are nowadays not only dealing with well established databases, but also with numerous initiatives of small databases and science portals, as well as with the storage of data on national and on international repositories.

The experimental and theoretical fields related to Atomic and Molecular (A&M) Data provide such a wealth of data that is used and applied across a wide range of scientific and technological applications. Indeed, progress in many scientific and technological areas is underpinned by the availability of accurate quantitative information on the collisional properties and spectroscopic characteristics of interacting species. A&M data are indispensable for such diverse applications as astrophysics, atmospheric science, the development of fusion energy, semiconductor manufacturing and other plasma based technologies, the lighting industry, etc.

This paper starts with a non-exhaustive presentation of Open Science initiatives. Then, following the thread of the life cycle of A&M data for astrophysics, the paper describes the various challenges involved in organising, publishing, finding, and re-using A&M data. Finally, the paper aims at proposing some ideas for improvement of the current system with respect to enabling a more sustainable transmission of information and data.

Please note that background hyperlinks are used throughout the paper for convenience of the reader.

2. Open Science Initiatives

The UNESCO recommendation on Open Science, which was adopted by the General Conference of UNESCO at its 41st session in November 2021, provides an international framework for Open Science policy and practice. Indeed, worldwide Open Science platforms and clouds have been or are being created, we can cite:

- the European Open Science Cloud (EOSC) that the European Commission supports through its science funding programs;
- the China Science and Technology Cloud (CSTCloud);
- the Australian Research Data Commons (ARDC);

- the Malaysian Open Science Platform;
- the African Open Science platform;
- the planned broadening of LA Referencia in Latin America;
- the Digital Research Alliance of Canada;
- the Nederlands plan for Open Science;
- Germany's Nationale Forschungsdaten Infrastruktur funding initiative that supports the building of thematic Open Science platforms in different areas of science, thus creating consortia in chemistry : NFDI4Chem, for "Particles, Universe, Nuclei and Hadrons: PUNCH4NFDI " and in physics : NDFI4phys;
- the 2nd French National Plan for Open Science, which combines policies to encourage Open Science and the usage of Open Science platforms with 4 pillars: 1) "The generalisation of open access to publications" with the obligation to publish all articles and books resulting from publicly funded calls in open access, and with an open access repository for publications HAL, 2) "Structuring, sharing and opening up research data", 3) "Opening up and promoting source codes produced by research", 4) "Transforming practices to make Open Science the default principle" through developing and promoting Open Science skills throughout the educational and career pathways of students and research staff, through promoting Open Science and the diversity of scientific productions in the assessment of researchers, of projects, of universities and of research organisations.
- The Republic of Korea's national Open Science approach proposes an overarching infrastructure, the ScienceON, which supports the whole research lifecycle by connecting various knowledge and information resources, platforms and services. In particular, this integrated infrastructure federates information on research outputs (e.g. papers, patents, research reports) and data from government-funded projects and makes them accessible in one place; it harvests metadata from foreign knowledge infrastructures, including the Japan's Institutional Repositories DataBase, the Australia's ARDC platform and the Europe's OpenAIRE platform; it provides an open collaboration environment for researchers [Shmagun *et al.* (2022), Han *et al.* (2022), Shin *et al.* (2019)].

At the transnational level CODATA launched the Global Open Science Cloud initiative (GOSC) that aims to encourage cooperation, and ultimately alignment and interoperability, between the above cited national initiatives and similar initiatives. There is an interplay between different actors at the national levels and in international organisations such as the Research Data Alliance (RDA), or in long term established association such as the International Union of Pure and Applied Chemistry (IUPAC); they work through different working groups and they propose recommendations that support the free and safe circulation of data, and they participate in different funded projects such as those supported by the European Commission. The latest example is the major new project "WorldFAIR : Global cooperation on FAIR data policy and practice" funded through the Horizon Europe Framework Programme. One of their case study covers chemistry which is relevant for atomic and molecular data.

3. Life cycle of Atomic and Molecular Data

The life cycle of atomic and molecular data is illustrated by the life cycle of astrophysical data, outlined below.

3.1. *From astrophysics to A&M Data*

The astrophysical community is very much organised around "observations" and therefore the preparation, development and the exploitation of space missions, and the exploitation of ground facilities. For the different instruments the science teams provide science use cases corresponding to key questions related to the observed astrophysical objects: sun, stars, planets, molecular clouds, protoplanetary disks, etc..

A simple view of the astrophysical work is that once observed spectra and/or images are analyzed, they provide information and constraints on the involved models. In many cases, both the analysis of spectra/images and the modelling of the astrophysical media involve A&M processes and therefore A&M data.

Discussions on "needs" for astrophysics can be organised in the different fields of research that are reflected by the International Astronomical Union (IAU) Divisions: for example, Division E deals with the sun and the heliosphere, Division F deals with planetary systems and astrobiology, Division H deals with the interstellar matter and the local universe, and so on. In addition Division B hosts the B5 commission on Laboratory Astrophysics that promotes astronomy and international collaboration through community-issued reports [Salama (2020), Barklem *et al.* (2023)] and the creation of working groups that produce regular reports on both developments and needs in the fields of atomic and molecular data for astrophysics.

In some countries such as the USA, the laboratory astrophysics community is well represented as a full Division of the American Astronomical Society (AAS) and is particularly active with the organisation of workshops such as the 2018 NASA Laboratory Astrophysics workshop [Brickhouse *et al.* (2020)] in preparation to the "Decadal Survey on Astronomy and Astrophysics 2020" that includes several documents related to the needs in Astrophysics: "The Need for Laboratory Measurements and Ab Initio Studies to Aid Understanding of Exoplanetary Atmospheres"[Fortney *et al.* (2020)],"From Interstellar Ice Grains to Evolved Planetary Systems: The Role of Laboratory Studies" [Gudipati *et al.* (2019)],"Unlocking the Capabilities of Future High-Resolution X-ray Spectroscopy Missions Through Laboratory Astrophysics" [Betancourt-Martinez *et al.* (2019)],"Astrophysical Science enabled by Laboratory Astrophysics Studies in Atomic, Molecular, and Optical (AMO) Physics" [Savin *et al.* (2019)],"Astromineralogy of interstellar dust with X-ray spectroscopy" [Corrales *et al.* (2019)]. In addition, the "Critical Laboratory Studies to Support and Advance Planetary Science and Planetary Missions: Overview, Challenges, and Recommendations" [Fayolle *et al.* (2021)] has been submitted as a white paper to the "Planetary Science and Astrobiology Decadal Survey 2023-2032". Some groups provide roadmaps, for example, the "Roadmap on cosmic EUV and X-ray spectroscopy" [Smith *et al.* (2020)].

Such published as well as other initiatives on "A&M data needs for astrophysics" should be generalized at the international scale and this requires a central indexing of "Needs," so that laboratory astrophysics groups and young researchers can be informed in a timely manner. IAU Commission B5 is the perfect tool to achieve this goal.

3.2. *Experimental and Theoretical Challenges*

The expression of "Needs" for A&M data can lead to specific funded projects such as individual large grants or national or international collaborative projects. Such projects include the description of the objectives, the milestones, the methodologies, the expected results and a plan for the dissemination of results, called the research data management (RDM) plan. Considering how sparse the laboratory astrophysics resources generally are, it would be useful that a minimum set of metadata concerning the description of the

projects as well as the relevant data sets, be published and indexed in a central repository. Indeed, such information is useful for other groups to easily obtain an overview of the existing knowledge, to decide on future research directions, and for young researchers to learn more about the current activities in laboratory astrophysics.

The needs for A&M data in astrophysics usually concern "secondary data" obtained by processing of "primary data" that are measured by experiment or are calculated via numerical models. As an example of primary and secondary A&M data, one can take the examples of measured experimental spectra which are primary data: these data are not usually made publicly available, and they are further post-treated to obtain secondary data such as line lists, energy levels, possibly line parameters such as broadening and shifting coefficients. These secondary quantities are the ones useful for the analysis of astronomical spectra and they are the quantities that are found in databases such as the Vienna Atomic Line Database (VALD) [Ryabchikova *et al.* (2018)] compilation, which is a critically assessed collection of radiative transition data aimed primarily at the stellar astrophysics community. Nevertheless, it may also turn out to be useful to keep the primary data obtained for some range of frequencies, so as to combine the past information with new information.

Another example is the numerical simulation of the collisional excitation of molecules by para/ortho–H_2 for application to the interstellar medium (see, for example, Daniel *et al.* (2011)). Such calculations involve several steps: the first step involves the calculations of the potential energy surfaces for some geometries of the atoms, meaning the interaction between the molecular species, then those data points are further fitted by function forms; the second step corresponds to solving the nuclear motion Hamiltonian and to obtaining the diffusion matrices or cross sections as a function of the relative collisional energy; the fourth step corresponds to taking an energy Boltzmann average of the cross sections in order to obtain state-to-state rate coefficients, and the final step is to sum the rate coefficients over the final state of para/ortho–H_2 and to average over the initial state of para/ortho–H_2 to obtain "thermalized" rate coefficients that are going to be used in radiative transfer models for the analysis of non-LTE (non-local thermodynamic equilibrium) media in the interstellar medium. In several databases such as the LAMDA database [Schöier *et al.* (2019)], the new EMAA database or the CASSIS database (see section 5.0.1), only the final step is stored. In the BASECOL database [Dubernet *et al.* (2013)] one can find the fourth and the last step; the other steps, that could be considered as "primary data" are not always publicly available, though their open publication would be useful for the purpose of re-use and comparisons.

In the above examples that span experimental and theoretical data, spectroscopic and collisional processes, the publication of primary data would request that the data be associated to rich metadata with a full explanation about the context of the experiments or of the calculations, that the data be associated to adequate provenance, that the data be indexed in some general repository, that the data be stored in sustainable and machine readable format, so that they could be findable, accessible, interoperable and re-usable (FAIR) [Wilkinson *et al.* (2016)]. Apart from the technical issues, the key scientific issue is that users not familiar with the data interpretation process may connect good data to bad conclusions. Currently the community is far from having defined the necessary context to usefully publish those primary data, and an international effort would be necessary to achieve such goal.

4. Data and Databases Challenges

This section focuses on database issues. The onset of collecting A&M data for astrophysics has been described in the presentation of the achievements of Dr Charlotte Sitterly Moore who has been a pioneer in creating consistent and evaluated collections of data [Devorkin (2023), Kramida (2023)]. With the start of internet many electronic databases have been built, nevertheless the scientific work of preparing data has not changed.

4.1. *Data preparation and ingestion in databases*

The scientific work involved in the construction of a database involves many steps: the collection of the data and sometimes the critical evaluation of those data as is done for the NIST ASD database [Kramida *et al.* (2019)], a database containing atomic data including energy levels, radiative transitions probabilities, oscillator strengths, observed and accurately calculated wavelengths of spectral lines. Most of the time, the scientific work requires the verification of the consistency of data sets and possibly the aggregation of different datasets; it involves the association of numerical data with the experimental, fitting, theoretical methodologies that have been used to obtain those data; the association of relevant references to the data; and finally the addition of metadata so that the data can be found and retrieved. This work is highly specialized with the involvement of experts, is very time consuming, but is not very attractive for the careers of (young) researchers as there is very little recognition of the importance of this scientific activity in the evaluation rounds and during the recruitment processes. Therefore the community should be aware of these issues and make the relevant efforts to provide the most complete relevant information and data to the scientific maintainers of the databases. In order to address these issues, some databases have designed input templates that producers of data can fill, those templates help to secure and speed up the processing of the information. The BASECOL database [Dubernet *et al.* (2013)], which provides state-to-state inelastic collisional rate coefficients with energy transfer in both the target and the collider mainly for the interstellar medium and cometary atmospheres, has put together a template [Ba *et al.* (2020)] that includes the numerical data: rate coefficients values and energy levels values with quantum numbers, the metadata that allow to search and to display information on the website, metadata that follow the Virtual Atomic and Molecular Data Center (VAMDC) standards conventions and that allow interoperability within the VAMDC e-infrastructure.

The International Atomic Energy Agency's CollisionDB, a database of plasma collisional processes, cross sections and rate coefficients for nuclear fusion energy and astrophysics research, provides an input template whose principle is similar to the BASECOL template; the data providers use a simplified, key-value pair format and the input of metadata and information are partly standardized: the processes, the species and the states are expressed in string that can be interpreted by the Pyvalem software designed by the developers.

The SSHADE database, which provides information to the planetary and astrophysics community spectral and spectro-photometric data, bandlists related to materials such as ices, minerals, etc., uses a complex input XML file that allows to include all the relevant data in the database. The different groups can access the database to implement their data, and their data are identified as the group's database. In this example a single input file and database welcome very different types of data that are described by a single data model.

It should be mentioned that in the three above cases the users providing the data

often protest that the requested work is time consuming. However, they should realize/be aware that those procedures ensure the maximum integrity and the sustainability of the databases, as well as ultimately promoting citation of their work upon the data's use. The last item requires that the databases attach to the data the references of the initial manuscript and linking to the original authors. Fortunately, nowadays this is mostly the case.

4.1.1. *Access to data, standards and e-infrastructures*

All the databases offer the possibility to download data from their websites with output format that may vary. Sometimes those output formats are designed to fit the needs of various users. An example is the ExoMol database [Tennyson *et al.* (2020)], which provides molecular line lists for exoplanet and other atmospheres with an emphasis on hot atmospheres, as well as many other parameters such as opacities and pressure broadening coefficients, for example. The various output files for a single dataset fit the modelling needs of different communities. As part of the spectroscopic input to atmospheric codes, the HITRAN molecular spectroscopic database output format [Gordon *et al.* (2022)], which also includes data for collision-induced absorption [Karman *et al.* (2019)] is already internationally recognized as a standard in the planetary community and is also used in its high-temperature version HITEMP [Rothman *et al.* (2010)] HITEMP. For the planetary community the GEISA database is also widely recognized.

Some databases are provided with a suite of software and modelling codes that make use of the data. A typical example is the NASA Ames Polycyclic Aromatic hydrocarbons (PAH) IR Spectroscopic Database (PAHdb) [Bauschlicher*et al.* (2018)] which is a web-accessible database with accompanying models and tools to readily analyze and interpret astronomical PAH observations. Another example is the AtomDB project [Foster *et al.* (2012)] which is a combination of atomic data and of the plasma models required to convert these atomic data into spectra useful for analysing astrophysicsal X-ray spectra. Finally, a third example is the Astrochem-tools website that provides the access to the KIDA database[Wakelam *et al.* (2012)] for astrochemistry and to the Nautilus gas-grain code that embeddes networks with KIDA's data.

The VAMDC Consortium [Dubernet *et al.* (2010), Dubernet *et al.* (2016), Albert *et al.* (2020)] has developed robust standards such as the XSAMS standards and has built a robust technical infrastructure that allows to find and retrieve A&M data across nowadays 41 databases. Section 2 of Albert *et al.* (2020) provides a short description of 37 of those databases as well as a table indicating the fields of application in astrophysics. In addition to the access from their usual website, those databases are accessible via any software that interrogates the VAMDC platform. For example they are accessible from the VAMDC portal [Moreau *et al.* (2018)] and partly from a user tool such as the CASSIS software (see section 5.0.1) or a VAMDC tool such as the SPECTCOL tool [Ba *et al.* (2023)]. In addition a recent work accessed the level of FAIRness of the VAMDC infrastructure as the FAIR principles (Wilkinson *et al.* (2016)) are essential for intra- and interdisciplinary exchanges. The result of the analysis is provided in the proceedings of the IAU B2-B5 inter-commission Working group on "Laboratory Astrophysics Data" session [Dubernet (2023)].

4.1.2. *A few examples of other initiatives*

A website at Leiden Observatory provides the basic molecular data needed to compute photo-destruction rates in various environments, including diffuse and translucent clouds, dense PDRs, the surface layers of protoplanetary disks, and cometary and exoplanetary atmospheres. Timely with the recent launch of the James Webb Science Telescope, the

Leiden Ice Database for Astrochemistry (LIDA)Rocha *et al.* (2022) has been recently upgraded; it contains high quality vibrational spectra of solid-phase molecules in ice mixtures and for temperatures of astrophysical relevance. This information is needed to interpret infrared observations toward protostars and background stars.

The Heidelberg-Jena-St Petersburg database [Henning *et al.* (1999), Jäger *et al.* (2003)] contains references to papers and links to internet resources related to measurements or calculations of the optical constants of materials of astronomical interest.

Among other portals that might be useful for astrophysics, one can find links to many databases for atomic and plasma physics on the website of the Weizmann Institute, one can find many datasets in the LXCat project which is aimed at the low temperature plasmas community. Other databases such as the PEARL database and the DCCP database in Korea, whose initial target audience is the plasma community, could be also useful to the astrophysical community.

For high energy astrophysics, there are several initiatives such as the above cited AtomDB project [Foster *et al.* (2012), Smith *et al.* (2001)], the Interactive Kronos Charge Exchange Database project [Cumbee *et al.* (2021)], as well as several initiatives linked to opacities for kilonovae and neutron star mergers: the NIST-LANL Lanthanide Opacity Database and the Japan-Lithuania Opacity Database for Kilonova.

Among other new initiatives of databases or portals, we can cite the Cosmic PAH Portal that includes molecular databases linked mostly to PAH and various tools to model and analyse spectra, the new Lille Spectroscopic Database that provides molecular spectroscopic data produced by the group and their collaborators.

The conclusion that can be drawn from the many but definitely not complete list of databases and portals is that there is space for improvement to further improve the general indexing of A&M resources. This is a major job, that becomes clear when realizing that a large consortium such as VAMDC has spent 6 years to build and provide the software and standards that now interconnect 41 databases, but that this still does not include other resources that would be useful to connect to.

5. Usage of A&M Data and Databases

5.0.1. *Astronomical Tools and Software using A&M Data*

Many software tools and numerical codes process and use A&M data. These might retrieve data directly from databases, prepare their own static files using different sources such as databases, papers, data sent by producers, or interpolate or extrapolate the data depending upon their usage. We list here some examples of existing tools and numerical codes in addition to the suite of tools already cited.

The CASSIS software is an example of an astrophysical application that both accesses external A&M databases via VAMDC and contains internally compiled files for the analysis of media that are not in local thermodynamic equilibrium (from the LAMDA database, producers, and BASECOL). An interesting feature of the CASSIS software is the usage of two e-infrastructures: the VAMDC infrastructure based on VAMDC standards for the access to A&M data and the International Virtual Observatory (IVOA) that has developed and implemented its VO standards on the major astronomical archives. Indeed, the CASSIS software and the Aladin application have recently interfaced their VO applications in order to create a tool able to explore simultaneously both the spatial and spectral dimensions of data cubes. The XCLASS software (eXtended CASA Line Analysis Software Suite) [Möller *et al.* (2017)] is a toolbox for the Common Astronomy Software Applications package (CASA), aimed at

fitting spectral line data from astronomical sources observed both with interferometers or single dish telescopes. XCLASS models a synthetic spectrum that is automatically compared to the data with the aim of providing a measurement of physical quantities of the media. Molecular data required by XCLASS are taken from an embedded database, populated and updated via the Python library vamdclib, which queries the data directly from the VAMDC nodes [Regandell *et al.* (2018)] of CDMS [Müller*et al.* (2018)]] and JPL [Pickett *et al.* (2018)] databases. The MADCUBA suite of tools [Martín *et al.* (2019)] is a package developed to import, visualise, manipulate, process, and analyse molecular and recombination line astronomical data from both 3D spectroscopic cubes and single-pointing spectra. They use data from the JPL and the CDMS databases and other datasets listed in their publications. The ARTEMIX service that searches and displays ALMA data includes a line search tool called YaFits that makes use of VAMDC standards to access CDMS and JPL databases. The ENIIGMA tool [Rocha *et al.* (2021)] is a fitting tool that decomposes the infrared spectrum of protostars containing ice features by using a linear combination of infrared laboratory data of molecules in the solid phase, this tool is linked to the LIDA database (see section 4.1.2).

There are many numerical codes that simulate different media in astrophysics and that include internal A&M datasets. We can cite the spectral synthesis code CLOUDY [Ferland *et al.* (2017)] for Interstellar Medium, Stellar/Solar Astrophysics, the MARCS code for 1D LTE model atmosphere production [Gustafsson *et al.* (2008)] for stellar/solar astrophysics, the PHOENIX code [Hauschildt *et al.* (1999)] for 1D+3D atmosphere modelling of astrophysical objects such as exoplanets, all types of stars, novae, supernovae, including EOS, NLTE, radiative transfer. The Interstellar Medium (ISM) service makes available many codes that include A&M data: the Meudon PDR code [Le Petit *et al.* (2006)] that can be used to study the physics and chemistry of diffuse clouds, photodissociation regions (PDRs), dark clouds, the Dustem code [Compiègne *et al.* (2006)] which is a numerical tool that computes the extinction and the emission of interstellar dust grains heated by photons, the Paris-Durham shock code is a software dedicated to the modelling of magnetized molecular shocks propagating in interstellar environments.

5.0.2. *Noteworthy issues with usage of A&M data*

This paragraph is not intended to highlight issues on the specific tools and numerical codes cited in this paper. The remarks deal with general issues that come with the use of A&M databases.

The general concepts of FAIR [Wilkinson *et al.* (2016)] are encouraged at the international level as they will ease finding, accessing, comparing and re-using data. The FAIR principles can be resumed as : 1) Findable: unique and persistant identifiers are assigned to data and metadata, the metadata are rich, there is a registration of metadata and data; 2) Accessible: there are standardised, open, free communication protocols to retrieve data and metadata, metadata are still available when the data are no longer available; 3) Interoperable: the data and metadata use a formal, accessible, shared and broadly applicable language for knowledge representation, vocabularies follow FAIR principles; 4) Re-usable: the data and metadata are richly described with accurate and relevant attributes, associated with detailed provenance, and they meet domain-relevant community standards. The IAU inter-commission B2-B5 working group "Laboratory Astrophysics Data Compilation, Validation and Standardization : from the Laboratory to FAIR usage in the Astronomical Community" goal [Dubernet (2023)] is to encourage the FAIR publication and usage of A&M data. To this extend

one of their recent surveys, among maintainers of numerical codes and tools, shows that these maintainers care about the versioning of the A&M datasets used in their application, but that the A&M data citation is less an issue and a concern. Indeed within a given application, the versioning of A&M datasets is a key component of the reproducibility of simulations, but comparison of results from simulations using different codes should be able to identify the provenance of the data, and data citation is part of this provenance concept. Another parallel survey by the same working group shows that databases maintainers and A&M producers are extremely sensitive to data citation, as the citation of the producers of A&M data and of the databases has a strong impact on the funding of their activities, on the other hand they are less sensitive to the versioning of their data. One obvious recommendation from the working group will be to encourage versioning and citation of A&M data on both sides: users and databases.

The adoption of FAIR principles will certainly improve many aspects of data exchange, but it will not solve the issue of guiding the users in choosing the relevant datasets for their applications. Indeed the concept of data quality is a human concept that varies with needs. Data quality for A&M producers is linked to the methodologies and the uncertainties: this is formalized by acceptance in published papers and by a proper description of the relevant medatada, so FAIR principles should be enough. For users of A&M data their perception of data quality might vary with their needs, and therefore only a combined work of producers and users can provide the necessary guidance in choosing and accessing the data relevant to their applications. Therefore we could imagine that the solution to "choosing datasets" will create new "third generations of databases" aimed at specific users (or applications), and that those third generation databases will contain versioned datasets with full citation and rich metadata indicating the limits of application of those datasets.

6. Conclusion and Perspective

This paper is certainly incomplete with respect to references to different A&M resources and even more regarding numerical codes and tools for astrophysical applications. This is basically an impossible job and emphasizes even more that the following issues should be addressed in the future :

- for the laboratory A&M activities: to have a central indexing of all activities related to laboratory A&M data for astrophysics, and possibly other fields;
- for data: to define the commons in order to index, to find and to understand all the available A&M resources, thus following FAIR principles both for the publication and the usage of A&M data; indeed among other benefits we should make sure that the A&M data producers are cited when A&M data are used in applications.
- for user communities: to make the effort to maintain publicly available third generation databases that serve some specific user's purpose and that follow FAIR principles. This is challenging as it requires a strong interdisciplinary approach between chemists, physicists, astrophysicists and data curators.
- for databases: to ease at the maximum the work of the databases scientific maintainers and to have a structural approach that supports the long term sustainability of the databases.

Solving such issues will create de facto a "Global Network of Laboratory A&M activities and data for astrophysics (and possibly other applications)", that would go in the direction of better science and a sustainable world.

Acknowledgements

MLD thanks Ms I. Blanc from the French Ministry of Higher Education and Research for providing a full presentation of the 2nd French Plan for Open Science. MLD thanks Dr Kwon Duck-Hee and Dr Song Mi-Young for welcoming her and supporting her visits respectively at KAERI and at KFE. MLD thanks Dr Jaesoo Kim, the president of KISTI, and Dr Kwangnam Choi, the head of the KISTI's Division of National Science & Technology Data, for allowing her visit at KISTI, Ms H. Shmagun from KISTI for providing information on Open Science in Korea. MLD thanks Dr B. Schmitt for providing information on the SSHADE project; Dr R. Smith for exchanges on Laboratory astrophysics in the USA; Dr P. Salomé for providing information on the ALMA/Artemix/Yafits suite of software. MLD thanks Dr C. Jäger, Dr M. Rengel, Dr W. Rocha who participated to the IAU B2-B5 working group session, and Dr B. Barbuy, Dr D. Ryu, Dr Dr P.K. Tan, N. Watanabe who participated to the IAU B5 commission business meeting, both meetings being held at the IAU 2022 GA in Busan 8th August. MLD thanks the VAMDC consortium, and in particular the Paris team : Ms Y.A. Ba, Mr N. Moreau, Dr C.M. Zwölf for their long term collaboration.

References

Albert, D., Anthony, B.L. , Ba, Y.A. & al. 2020, *Atoms*, 8, 76. DOI = 10.3390/atoms8040076

Ba, Y.A. , Dubernet, M.L., Moreau, N., Zwölf, C.M. 2020, *Atoms*, vol. 8, number 4, article 69 ; DOI : 10.3390/atoms8040069

Ba, Y.A. , Dubernet, M.L., Moreau, N., Zwölf, C.M. 2023, *to be published in A&A*

Barklem, P.S., Barbuy, B., Dubernet, M.L., Ryu, D. & al. 2023, *IAU 371 Symposium Proceedings*, "IAU B5 commission Activities : Laboratory Astrophysics", 8th August 2022, IAU 2022 General Assembly, Busan, Eds Soderblom & G. Nave; to be published

Bauschlicher, C.W., Ricca, A., Boersma, C., Allamandola, L.J. 2018, *AJSS*, 234, 32; DOI: 10.3847/1538-4365/aaa019

Betancourt-Martinez, G., Akamatsu, H., Barret, D., Bautista, M. & al. 2019, *BAAS*, 51, 537; Bibcode : 2019BAAS...51c.337B

Brickhouse, N., Ferland, G. J., Milam, S. & al. 2020, *BAAS*, 52, 202; DOI : 10.3847/25c2cfeb.10bdb63d

Compiègne, M., Verstraete, L., Jones, A. & al. 2011, *A&A*, 525, A103; DOI: 10.1051/0004-6361/201015292

Corrales, L.,Valencic, L.,Costantini, E., Garcia, J., Gatuzz, E., Kallman, T. & al. 2019, *BAAS*, 51, 264; Bibcode : 2019BAAS...51c.264C

Cumbee, R. and Stancil, P. and Mcilvane, S. 2021, *American Astronomical Society Meeting Abstracts*, 53, 126.01; Bibcode: 2021AAS...23812601C

Daniel, F., Dubernet, M.-L., Grosjean, A. 2011, *A&A*, 536, A76+; DOI : 10.1051/0004-6361/201118049

Devorkin, D. 2023, *IAU 371 Symposium Proceedings*, IAU 2022 General Assembly, Busan, 371 Symposium: "Honoring Charlotte Moore Sitterly: Astronomical spectroscopy in the 21st century", Eds Soderblom & G. Nave; to be published

Dubernet, M.-L., Boudon, V., Culhane, J. L., & al. 2010, *JQRST*, 111, 2151-2159; DOI : 10.1016/j.jqsrt.2010.05.004

Dubernet, M.-L. , Alexander, M. H., Ba, Y. A., Balakrishnan, N. & al. 2016, *A&A*, 553, A50; DOI : 10.1051/0004-6361/201220630

Dubernet, M.-L., Antony, B. K., Ba, Y.-A., & al. 2016, *J. of Physics B*, 49, 074003; DOI:10.1088/0953-4075/49/7/074003

Dubernet, M.L., Berriman, G.B., Boersma, C. & al. 2023, *IAU 371 Symposium Proceedings*,

"IAU B2-B5 inter-commission working group session", 8th August 2022, IAU 2022 General Assembly, Busan, Eds Soderblom & G. Nave; to be published

Edith C. Fayolle, E.C., Barge, L., Cable, M. & al. 2021, *BAAS*, 53, 170; DOI: 10.3847/25c2cfeb.f51744bb

Ferland, G. J. and Chatzikos, M. and Guzmán, F. & al. 2017, *Revista Mexicana de Astronomía y Astrofísica* , 53, 385-438; Bibcode : 2017RMxAA..53..385F

Fortney, J.,Robinson, T. D., Domagal-Goldman, S., Genio, A. D. Del, Gordon, I. E. & al. 2020, *Astro2020: Decadal Survey on Astronomy and Astrophysics*, 2020, 146; Bibcode: 2019astro2020T.146F

Foster, A.R., Ji, L. , Smith, R.K., Brickhouse, N.S. 2012, *ApJ*, 756, 128; DOI: 10.1088/0004-637x/756/2/128

Gordon, I. E., Rothman, L. S., Hargreaves, R. J., et al. 2022, *JQRST*, 277, 107949. doi:10.1016/j.jqsrt.2021.107949

Gudipati, M., Milam, Stefanie N., Hendrix, A. R., Henderson, B. L., Linnartz, H.V. J. & al. 2019, *BAAS*, 51, 518; Bibcode : 2019BAAS...51c.518G

Gustafsson, B. and Edvardsson, B. and Eriksson, K. and Jørgensen, U. G. and Nordlund, Å. and Plez, B. 2008, *A&A*, 486, 951; DOI: 10.1051/0004-6361:200809724

Han, Sangjun, Shin, Jaemin, Lee, Seokhyoung, Park Junghun 2022, *J. of Information Science Theory and Practice*, 10, 1-11; DOI: 10.1633/JISTaP.2022.10.S.1

Hauschildt, P. H. and Baron, E. 1999, *Journal of Computational and Applied Mathematics*, 109, 41; Bibcode: 1999JCoAM.109...41H

Henning, Th. and Il'in, V. B. and Krivova, N. A. and Michel, B. and Voshchinnikov, N. V. 1999, *AASS*, 136, 405; DOI: 10.1051/aas:1999222

Jäger, C. and Il'in, V. B. and Henning, Th. and Mutschke, H. and Fabian, D. and Semenov, D. and Voshchinnikov, N. 2003, *JQSRT*, 79-80, 765; DOI: 10.1016/S0022-4073(02)00301-1

Karman T., Gordon, I. E., et al. 2019, *Icarus*, 328, 160-175. doi:10.1016/J.ICARUS.2019.02.034

Kramida, A. 2023, *IAU 371 Symposium Proceedings*, IAU 2022 General Assembly, Busan, 371 Symposium: "Honoring Charlotte Moore Sitterly: Astronomical spectroscopy in the 21st century", Eds Soderblom & G. Nave; to be published

Kramida, A., Ralchenko, Yu., Reader, J., and NIST ASD Team 2020, *NIST Atomic Spectra Database (ver. 5.7.1), [Online]. Available:* `https://physics.nist.gov/asd` *[2020, July 9]. National Institute of Standards and Technology, Gaithersburg, MD.*

Le Petit, F., Nehmé, C., Le Bourlot, J., Roueff, E. 2006, *ApJS*, 164, 506; DOI: 10.1086/503252

Martín, S. and Martín-Pintado, J. and Blanco-Sánchez, C. and Rivilla, V. M. and Rodríguez-Franco, A. and Rico-Villas, F. 2019, *A&A*, 631, A159; DOI: 10.1051/0004-6361/201936144

Möller, T. and Endres, C. and Schilke, P. 2017, *EDP Sciences*, 598, A7; DOI : 10.1051/0004-6361/201527203

Moreau, N., Zwölf, C.M., BA, Y.A., & al. 2018, *Galaxies*, 6, 105; DOI : 10.3390/galaxies6040105

Müller, H. S. P. and Schlöder, F. and Stutzki, J. and Winnewisser, G. 2005, *J. of Molecular Structure*, 742, 215; DOI: 10.1016/j.molstruc.2005.01.027

Pickett, H. M., Poynter, R. L., Cohen, E. A., Delitsky, M. L., Pearson, J. C.,Müller, H. S. P. 1998, *JQRST*, 60, 883-890; DOI: 10.1016/S0022-4073(98)00091-0

Regandell, S., Marquart, T., et Piskunov, N. 2018, *Physica Scripta* 93, 3 035001; DOI: 10.1088/1402-4896/aaa268

Rocha, W. R. M., Perotti, G., Kristensen, L. E., Jørgensen, J. K. 2021, *A&A*, 654, A158; DOI: 10.1051/0004-6361/202039360

Rocha, W. R. M., Rachid, M. G., Olsthoorn, B. & al. 2022, *arXiv*, 0, 0; DOI : 10.48550/ARXIV.2208.12211

Rothman, L. S., Gordon, I. E., Barber A., et al. 2010, *JQRST*, 111, 2139-2150. doi:10.1016/j.jqsrt.2010.05.001

Ryabchikova, T., Pakhomov, Y., Piskunov, N. 2018, *Galaxies*, 5, 93; DOI: 10.3390/galaxies6030093

Salama, F. 2020, *Proceedings IAU Symposium No. 350*, "Laboratory Astrophysics: from Observations to Interpretation", F. Salama & H. Linnartz, eds., CUP, IAU 2020; doi: 10.1017/S1743921320004123

Savin, D.W., Babb, J.F. , Bellan, P.M., Brogan, C., Cami, J., Caselli, P. & al. 2019, *BAAS*, 51, 96; Bibcode : 2019BAAS...51c..96S

Schöier, F. L. and van der Tak, F. F. S. and van Dishoeck, E. F. and Black, J. H. 2005, *A&A*, 432, 369-379; DOI: 10.1051/0004-6361:20041729

Shin, Youngho, Um, Junho, Seo, Dongmin, Shin, Sungho 2022, *J. of Information Science Theory and Practice*, 10, 25-38; DOI: 10.1633/JISTaP.2022.10.S.3

Shmagun, H., Shim, J., Choi, K.-N., Shin, S. K., Kim, J., & Oppenheim, C. 2022, *Journal of Information Science*, 0(0); DOI: 10.1177/01655515221107336

Smith, R., Hahn, M., Raymond, J., Kallman, T., Ballance, C. P. & al. 2020, *J. of Physics B*, 53, 092001; DOI : 10.1088/1361-6455/ab69aa

Smith, R.K., Brickhouse, N.S., Liedahl, D.A., Raymond, J.C. 2012, *ApJ*, 556, L91 ; DOI: 10.1086/322992

Tennyson, J., Yurchenko, S.N., Al-Refaie abd, A.F., Clark, V.H.J., Chubb, K.L. & al. 2020, *JQRST*, 255, 107228; DOI: 10.1016/j.jqsrt.2020.107228

Wakelam, V. and Herbst, E. and Loison, J.-C. and Smith, I. W. M. & al. 2012, *ApJS*, 199, 21; DOI: 10.1088/0067-0049/199/1/21

Wilkinson, M., Dumontier, M., Aalbergersberg, I. & al. 2016, *Sci Data* 3, 160018; DOI: 10.1038/sdata.2016.18