Chapter II: Databases and Community Activities

https://doi.org/10.1017/S1743921323000327 Published online by Cambridge University Press https://doi.org/10.1017/S1743921323000327 Published online by Cambridge University Press

Honoring Charlotte Moore Sitterly: Astronomical Spectroscopy in the 21st Century
Proceedings IAU Symposium No. 371, 2024
D. R. Soderblom & G. Nave, eds.
doi:10.1017/S1743921323000327

Laboratory Astrophysics Data Working Group IAU 2022 GA session

M.L. Dubernet[1], G.B. Berriman[2], C. Boersma[3], C. Jäger[4,5], M. Rengel[6], W. R. M. Rocha[7], F. Salama[8] and C.M. Zwölf[1]

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

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

[3]NASA Ames Research Center, MS 245-6, Moffett Field, CA 94035-1000, USA
email: Christiaan.Boersma@nasa.gov

[4]Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany;

[5]Institut für Festkörperphysik, Friedrich-Schiller-Universität Jena, Helmholtzweg 3, D-07743 Jena, Germany;
email: Cornelia.Jaeger@uni-jena.de

[6]Max-Planck-Institut für Sonnensystem Forschung, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany,
email: rengel@mps.mpg.de

[7]Laboratory for Astrophysics, Leiden Observatory, Leiden University, P.O. Box 9513, NL 2300 RA Leiden, The Netherlands,
email: rocha@strw.leidenuniv.nl

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

Abstract. The paper corresponds to the session organised by 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" at the IAU 2022 General Assembly. The session included talks about the usage and implementation of FAIR concepts in VAMDC and in the IVOA, then domain specific talks oriented towards planetology, dust and ices. The program (doi.org/10.5281/zenodo.7050654) and the various talks can be found in the ZENODO "cb5-labastro" community (zenodo.org/communities/cb5-labastro).

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

1. The Laboratory Astrophysics Data working group Activities

The overall objective and goals of 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"† is to provide a platform where to discuss the FAIR [Wilkinson *et al.* (2016)] (Findability, Accessibility,

† www.iau.org/science/scientific_bodies/working_groups/335/

Interoperability, Reuse) usage of laboratory Atomic and Molecular (A&M) data in astronomy and astrophysics. The working group collects information on items such as managing, archiving, and sharing atomic and molecular data of interest to astronomers, on the validation processes of those data, on the documentation, on the referencing and citation processes, on the available tools and codes using A&M data in the astronomical community. The deliverable involves to provide a state-of-the art report with the above information, to identify the bottlenecks in providing the data to the astronomical community and in the FAIR usage of the A&M data by the astronomical community, to provide practical and political recommendations related to optimizing the process from laboratory data to astrophysics and vice-versa. Slides are here†.

2. VAMDC and FAIR principles

The Virtual Atomic and Molecular Data Centre VAMDC (vamdc.org), developed since 2009 [Dubernet *et al.* (2010), Dubernet *et al.* (2016), Albert *et al.* (2020)], interconnects ∼41 heterogeneous A&M databases that are mainly used for the interpretation of astronomical spectra and for the modeling in media of many fields of astrophysics, and offers a common entry point through the VAMDC portal [Moreau *et al.* (2018)]. From a technical point of view the VAMDC infrastructure relies on four pillar components: **1) The set of federated data-nodes** [Regandell *et al.* (2018)], the federated databases accepting VAMDC queries and producing as an output a standard XML file : VAMDC XML Schema for Atomic Molecular and Solid Data, VAMDC-XSAMS‡. In VAMDC jargon a federated database is called a *data-node*; **2) The registries**§ where data-nodes metadata are registered. Those metadata include the resolvable identifiers of each node, the contact details for the scientific and technical maintainers, the version of implemented standard and protocols; **3) The Species Database**¶, a central service harvesting from the data-nodes the species information : name (in natural language), stoichiometric formula, formula, InChi/InChiKey, mass number and charge; **4) The Query Store**‖ [Zwölf *et al.* (2019)], a central repository of all the queries answered by the data-nodes. The Query Store assigns a persistent resolvable identifier (optionally this identifier may be a DOI) to each query. It stores the original query, together with the version of the data-node and of the standards, the extracted XSAMS data file, the list of publications associated to the retrieved dataset. With those four pillars, data in VAMDC are easy to discover and to find, they are accessible, interoperable and easy to reuse; so VAMDC implemented FAIR principles before their formal definition [Wilkinson *et al.* (2016)].

Nevertheless we performed an audit of the FAIR-ness of VAMDC using the assessment framework produced by the Research Data Alliance∗∗ [Bahim *et al.* (2020)]. The details and methodology of this audit will be described in a further work. In this text we will only focus on the improvement items identified during the assessment: **a)** Systematically define conditions of usage and licenses for all the services and data in the infrastructure; **b)** A single entry point should be provided to get all the information about a given node. Today this information is fragmented between the Specied Database and the Registries; **c)** The references pointing from one technical pillar to other should be in the form of persistent resolvable machine actionable identifiers; **d)** Register VAMDC standards and format into ad hoc registries of types (e.g. FairSharing ones) and assign persistent resolvable machine actionable identifiers to each standard, to easily refer to it; **e)** Systematically use FAIR

† https://doi.org/10.5281/zenodo.6979299
‡ https://vamdc.org/standards
§ https://registry.vamdc.org
¶ https://species.vamdc.org
‖ https://cite.vamdc.eu/
∗∗ https://www.rd-alliance.org/

compliant dictionaries to express knowledge; **f)** Embed provenance information in all the element of the infrastructure, not only into the XSAMS output.

As a conclusion, the FAIR-ness level of VAMDC infrastructure is satisfying for experts aware of the infrastructure conventions and standards, but may be significantly increased for newcomers or for trans-disciplinary activities. Slides from the session are here (https://doi.org/10.5281/zenodo.6979299), and are further expanded in another presentation†.

3. FAIR principles in the IVOA

The Virtual Observatory is best described as a multi-wavelength digital sky that can be searched, visualised and analysed in new and innovative ways. It is built atop protocols for data discovery and access that are delivered by the International Virtual Observatory Alliance IVOA ((https://ivoa.net). These protocols, now built into the architecture of all major astronomy archives and data centers, together form a framework for astronomical datasets, tools, services to work together in a seamless way. From an architectural point of view, the VO can be thought of as a collection of interoperating tools and services.

Founded in 2002, the IVOA itself today has 24 national VO member projects and Inter Governmental Organizations (IGOs). The newest members, the Square Kilometer Array Observatory (SKAO) and VO-Kazakhstan, joined in 2022. From the outset, the goal of the IVOA was enable seamless interoperability of distributed data sets, and promote open data and open science. It was implementing FAIR principles before they were formalized as such. [O'Toole & Tocknell (2022)] describe how a service provider building a VO-compliant service can in fact make it fully compliant with FAIR principles by the addition of a DOI and an Open Source license. Slides are here‡.

4. The NASA Ames PAH IR Spectroscopic Database

The NASA Ames PAH IR Spectroscopic Database [(PAHdb)§ [Boersma *et al.* (2014); Bauschlicher, Jr. *et al.* (2018); Mattioda *et al.* (2020)] is a web-accessible database with accompanying models and tools to readily analyze and interpret astronomical PAH observations [e.g. Boersma *et al.* (2018)]. PAHdb contains the world's foremost collection of genuine laboratory-measured and quantum-chemically computed spectra from over 4000 PAHs and PAH clusters. PAHdb offers models and software tools, which can be obtained from GitHub (github.com/pahdb). Comprehensive manuals describing the data, models and software can be found via the PAHdb Documentation Portal¶. Taken together, this makes the PAH spectroscopic data findable, accessible, interoperable, and reusable (FAIR). Slides can be found here‖.

5. Laboratory Astrophysics Databases on grains and ices : from the laboratory to the users

Interstellar dust and ice play a major role in the thermodynamical and chemical evolution of the interstellar medium (ISM), protoplanetary and planetary disks, and planets. The analysis of observed spectral features delivers important information on grain sizes, composition, and structure as well as on the temperature and spatial distribution of the material. The current state of databases of optical constants or of absorption data of solids, containing information on the samples, including references to relevant papers

† https://zenodo.org/record/7112590#.Y06pTy8itB0
‡ https://doi.org/10.5281/zenodo.7049804
§ https://www.astrochemistry.org/pahdb/
¶ https://pahdb.github.io
‖ https://doi.org/10.5281/zenodo.7042959

as well as information on the measurements or on the calculations, have been discussed for the Leiden Ice Database for Astrochemistry - LIDA [Rocha *et al.* (2022)]†, the SSHADE database (https://www.sshade.eu) and the Heidelberg-Jena-St Petersburg database‡ [Henning *et al.* (1999), Jäger *et al.* (2003)]. One final remark: though the databases are well structured, an important issue is that astrophysicists need more guidance to choose the materials that are relevant for simulation and modelisation. Slides are here (https://doi.org/10.5281/zenodo.7040441).

6. About the atomic and molecular databases in the planetary community

A&M databases compile and provide detailed spectral information for A&M to feed codes that predict and simulate radiation in gaseous media. Between the applications, we find atmospheric physics: (exo) planetary atmospheres, comets and small bodies. These databases are a critical input for the codes which predict and interpret spectra of planetary atmospheres (hydrostatic equilibrium atmospheres and expanding comas), and space and ground-based telescopes facilities depend on the quality and extent of reference A&M parameters.

Several groups worldwide generate A&M data through measurement and/or calculation (e.g. HITRAN§, GEISA¶, JPL Molecular Spectroscopy‖, CDMS**, VAMDC (https://vamdc.org), ExoMol††, HITEMP‡‡, etc.). Several secondary databases and information services are fed with data from such sources in a fragmented manner. A variety of data formats (cross sections, K-tables, line-by-line, super-lines) and file formats (e.g. .hdf5, .pickle, .mp4, .txt, .npy) are generated. There are online tools such as HAPI (https://hitran.org/hapi/), exo-k library (https://pypi.org/project/exo-k/) that enables conversion between different formats. As part of the spectroscopic input to atmospheric codes, the HITRAN molecular spectroscopic database [Gordon *et al.* (2022)] is already internationally recognised as standard in the planetary community, and the ExoMol database [Tennyson *et al.* (2020)], valid over extended temperature ranges, is widely used in the exoplanetary community.

In spite of the tremendous advances and current efforts in the generation of databases, there are still many aspects in progress. A growing demand for spectroscopic data for (exo) planetary studies and other atmospheres is being driven by scientists who are interested in modelling as well as observing diverse bodies. Line lists are generated from experiments and/or ab initio calculations and may be incomplete or contain errors. Databases differ in completeness, and some ones do not accurately characterise high-frequency spectral regions. Atmospheric codes used by the planetary and exoplanetary characterisation communities, that are designed to solve the radiative transfer equation by looking at the propagation of radiation through a medium and simulate observations and infer parameters, have their own methods for the computation of opacities and there are no community standards. Furthermore, there are also no community standards in the selection of atmospheric codes in mission planning. There are needs to increase accessibility of opacities (computation, access, visualisation, manipulation), laboratory

† https://icedb.strw.leidenuniv.nl
‡ https://www2.mpia-hd.mpg.de/HJPDOC/index.php
§ https://hitran.org
¶ https://geisa.aeris-data.fr/
‖ https://spec.jpl.nasa.gov
** https://cdms.astro.uni-koeln.de
†† https://www.exomol.com
‡‡ https://hitran.org/hitemp

measurements of molecular cross-sections, and pressure broadening description for some species, among many other aspects. Slides are here†.

7. ENIIGMA: A Python package for ice spectral decomposition of protostars

The ENIIGMA fitting tool‡ [Rocha *et al.* (2021)] is a public Python package to fit laboratory-measured ice spectra to astronomical observations. The code handles a large amount of laboratory-measured ice spectra and uses genetic modelling algorithms to search for the global minimum solution. Additionally, a robust statistical module allows to assess the quality of the solutions. Slides are here§.

Acknowledgments

We thank the members of the SOC of this session : Prof. Beatriz Barbuy, Prof. Ohishi Masatoshi, Prof. Dongsu Ryu, and the other members of the inter-commission B2-B5 working group members. MLD and CMZ thank the VAMDC consortium and in particular the other members of the Paris team who are strongly involved in the maintenance of the core components of the VAMDC e-infrastructure : Nicolas Moreau and Yaye-Awa Ba.

References

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

Bahim, C., Casorrzn-Amilburu, C., Dekkers, M., Herczog, E. *et al.* 2020, *Data Science Journal*, 19, 41 10.5334/dsj-2020-041

Bauschlicher, Jr., C. W., Ricca, A., Boersma, C., Allamandola, L. J. 2018, *ApJSS*, 234, 32 10.3847/1538-4365/aaa019

Boersma, C., Bauschlicher, C. W., Ricca, A., Mattioda, A. L., Cami, J., Peeters, E., Sánchez de Armas, F., Puerta Saborido, G., Hudgins, D. M., Allamandola, L. J., 2014, *ApJSS*, 211, 8 10.1088/0067-0049/211/1/8

Boersma, C., Bregman, J., Allamandola, 2018, *ApJ*, 858, 67 10.3847/1538-4357/aabcbe

Dubernet, M.-L., Boudon, V., Cullhane, J. L., *et al.* 2010, *JQRST*, 111, 2151–2159; 10.1016/j.jqsrt.2010.05.004

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

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

Henning, Th. and Il'in, V. B. and Krivova, N. A. and Michel, B. and Voshchinnikov, N. V. 1999, *AASS*, 136, 405 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 10.1016/S0022-4073(02)00301-1

Mattioda, A. L., Hudgins, D. M., Boersma, C., Bauschlicher, C. W., Ricca, A., Cami, J., Peeters, E., Sánchez de Armas, F., Puerta Saborido, G., Allamandola, L. J., 2020, *ApJSS*, 251, 2 10.3847/1538-4365/abc2c8

Moreau, N., Zwölf, C.M., BA, Y.A., *et al.* 2018, *Galaxies*, 6, 105 10.3390/galaxies6040105

O'Toole, S. & Tocknell, J. 2022 arXiv:2203.10710

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

† https://doi.org/10.5281/zenodo.7040446
‡ https://eniigma-fitting-tool.readthedocs.io
§ https://doi.org/10.5281/zenodo.7040458

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

Rocha, W. R. M., Rachid, M. G., B. Olsthoorn, *et al.* 2022, *A&A* in press. 10.1051/0004-6361/202244032

Tennyson, J., Yurchenko, S. N., Al-Refaie, A. F., *et al.* 2020, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 255, 107228 10.1016/j.jqsrt.2020.107228

Zwölf, C.M., Moreau, N., Ba, Y.-A., Dubernet, M.-L. 2019, *Data Science Journal*, 18, 4 10.5334/dsj-2019-004

Wilkinson, M., Dumontier, M., Aalbergersberg, I. *et al.* 2016, *Sci Data* 3, 160018 10.1038/sdata.2016.18