

COMMISSIONS B5-B2 / WORKING GROUP

Laboratory Astrophysics Data Compilation, Validation and Standardization: from the Laboratory to FAIR usage in the Astronomical Community

CHAIR

VICE-CHAIR

ORGANIZING COMMITTEE

Marie-Lise Dubernet

Graham Bruce Berriman

Florence Conception, Friedrich Kupka ,

Farid Salama, Randall Smith, Jiayong Zhong

TRIENNIAL REPORT 2021-2024

1. Background

A large variety of atomic, molecular and solid-state data, commonly called “Laboratory Astrophysics Data”, are essential for the analysis of astronomical spectra and for the development of astronomical models of a wide variety of astronomical objects. Databases useful to the astronomical community have been built over the years. Many analysis tools and modelling codes include Laboratory Astrophysics Data, and both the astronomical and physics/chemistry communities have been working on interoperable access to data (IVOA[†]), the VAMDC[‡][Albert *et al.* (2020), Dubernet *et al.* (2016)] and the EuroPlanet[¶] consortiums).

The Inter-commission B5-B2 working group^{||} focuses on Laboratory Astrophysics Data Compilation, Validation, and Standardization, emphasizing FAIR (Findability, Accessibility, Interoperability, Reuse) usage in the Astronomical Community. FAIR principles are a cornerstone of Open Science, and address issues of transparency, sustainability and credit. Established in November 2021, the group includes members from France, the USA, Austria, the UK, and China.

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 summarized as : 1) Findable: unique and persistent 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. Our WG is to encourage the FAIR publication and usage of A&M data.

The 3 core milestones of the working group were to establish an overview of the state of the art of the communities (astronomy, physics, chemistry, planetary, etc....), to identify

[†] <https://ivoa.net/>

[‡] <https://vamdc.org/>

[¶] <http://www.europlanet-vespa.eu/>

^{||} https://www.iau.org/science/scientific_bodies/working_groups/335/

the bottlenecks and challenges encountered for FAIR exchange and usage of laboratory astrophysics data, and to provide recommendations to the IAU community and to the data providers.

2. Developments within the past triennium

2.1. Activities

The initial actions of the WG were linked to the first milestone of the WG†, i.e., to establishing an overview of the state of the art for the involved communities. Those activities led to one contributed and 2 invited presentations in conferences, to 5 seminars at institutes in France and in Korea, to presentations to the IUPAC, to the IVOA, to the VAMDC Annual Meeting and to its Board of Directors and to 3 refereed proceedings to be published in 2024, one of them being a white paper. In addition through those presentations we met the objectives of connecting to various communities and to international bodies that deal with standardisation in physics/chemistry and in astronomy.

Here is a detailed list of our activities :

- A Commission B5 Zenodo community has been created (cb5-labastro)‡. It includes all activities of the B5 commission encompassing the B2-B5 WG activities. This community can be used for any purpose linked to the "Laboratory Astrophysics Data" activities.
- Study of the state of the art of practices in the database community as well as in the astrophysics community related to the tools commonly used to analyse astronomical data and to numerical codes used to model astronomical media. Two on-line surveys were designed for this purpose: A database survey and an astro Codes/Tools survey. The preliminary results of the surveys were presented at the WG session organised at the IAU GA in Busan, 2022.
- The WG organised a session¶ at the IAU GA, Busan 2022, entitled "Laboratory Astrophysics Databases: from the provider to the user: encouraging FAIRness". The speakers came from the "low energy spectroscopy" communities and from the "data" communities. That session and some of its talks led to proceedings [Dubernet *et al.* (2024a), Rengel (2024)] that are included in the proceedings of the Symposium "IAUS 371: Honoring Charlotte Moore Sitterly: Astronomical spectroscopy in the 21st century" held at the Busan 2022 general assembly. The program can be found at (<https://zenodo.org/records/7050655>), it provides the zenodo DOI of the various presentations.
- A large survey on state-of-the art was presented in the Keynote Lecture|| "Atomic and Molecular Databases - Open Science for better science and a sustainable world" at the IAUS 371 symposium . The conclusions of our WG were also presented at the SPIG2022 conference†† "Towards a Global Network for Laboratory Astrophysics Activities and Data: Better science and a sustainable world"
- The WG contacted the VAMDC Consortium and suggested to perform a FAIR analysis of the VAMDC infrastructure. Dr C.M. Zwölf (Paris Observatory, France) performed the FAIR analysis of the VAMDC using a FAIR model from RDA. This led to a contribution at the IAU GA Busan 2022, to a contributed talk at the ICAMDATA conference‡‡

† https://www.iau.org/static/science/scientific_bodies/working_groups/335/wg-335-annual-report-2022-2023.pdf

‡ <https://zenodo.org/communities/cb5-labastro/>

¶ <https://zenodo.org/records/6979299>

|| <https://zenodo.org/records/7070971>

†† <http://spig2022.ipb.ac.rs/>

‡‡ <https://www.icamdata2022.it/>

and to a proceeding [Zwölf *et al.* (2023)], as well as to a presentation at the french section of the IVOA (ASOV) in Spring 2023¶¶.

- The WG interacted with the IVOA, the major international body for the standardisation of interoperability of astrophysical data, so that A&M data citation be considered in the standards of the IVOA in astro tools and numerical codes (this is an outcome of the surveys of databases providers).

- Finally a white paper [Dubernet *et al.* (2024b)] has been written in collaboration with an extended community, and it will be published in the IAUS 371 symposium proceedings. We provided our unpublished white paper to the large survey† that has been performed on “Laboratory Astrophysics” in the USA. The key conclusions of our white paper are developed below.

3. Recommendations

3.1. *WG conclusions following the WG surveys*

The WG surveys, among maintainers of numerical codes and tools, show 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.

3.2. *Recommendations extracted from the White Paper*

The White Paper [Dubernet *et al.* (2024b)] used the conclusions of our surveys, and proposed additional recommendations that we emphasise here : 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 formalised by acceptance in published papers and by a proper description of the relevant metadata, 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.

As a conclusion the white paper [Dubernet *et al.* (2024b)] suggests 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;

¶¶ <https://asov.obspm.fr/journees-asov-2023/>
 † <https://www.nsf.gov/mps/ast/aaac/labastro.jsp>

- 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.

4. Closing remarks

The Working Group is opened to new members as means to further widen its impact and long-term action. We believe that the future inter-commission B2-B5 inter-commission working group could help to solve the various issues that were raised at the end of the white paper.

Marie-Lise Dubernet
Chair of the Working Group

References

- Albert, D., Anthony, B.L., Ba, Y.A. & al. 2020, *Atoms*, 8, 76. doi = 10.3390/atoms8040076
- 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. 2024a, *IAU 371 Symposium Proceedings*, “IAU B2-B5 inter-commission working group session”, 8th August 2022, IAU 2022 General Assembly, Busan, Eds Soderblom & G. Nave; doi = 10.1017/S1743921323000327, to be published April 2024
- Dubernet, M.L., Berriman, G.B., Barklem, C. & al. 2024b, *IAU 371 Symposium Proceedings*, “Atomic and Molecular Databases Open Science for a sustainable world”, 8th August 2022, IAU 2022 General Assembly, Busan, Eds Soderblom & G. Nave; doi = 10.1017/S1743921323000261, to be published April 2024
- Rengel, M. 2024, *IAU 371 Symposium Proceedings*, “About the atomic and molecular databases in the planetary community – A contribution in the Laboratory Astrophysics Data WG IAU 2022 GA session”, 8th August 2022, IAU 2022 General Assembly, Busan, Eds Soderblom & G. Nave; doi = 10.1017/S1743921323000327, to be published April 2024
- Wilkinson, M., Dumontier, M., Aalbergersberg, I. & al. 2016, *Sci Data* 3, 160018; doi = 10.1038/sdata.2016.18
- Zwölf, C.M., Moreau, N. 2023, *Eur. Phys. J. D* 77, 70; doi = 10.1140/epjd/s10053-023-00649-x