

# Inter-Commission B2-B5 WG Activities and Plans

## “Laboratory Astrophysics Databases: from the provider to the user: encouraging FAIRness”

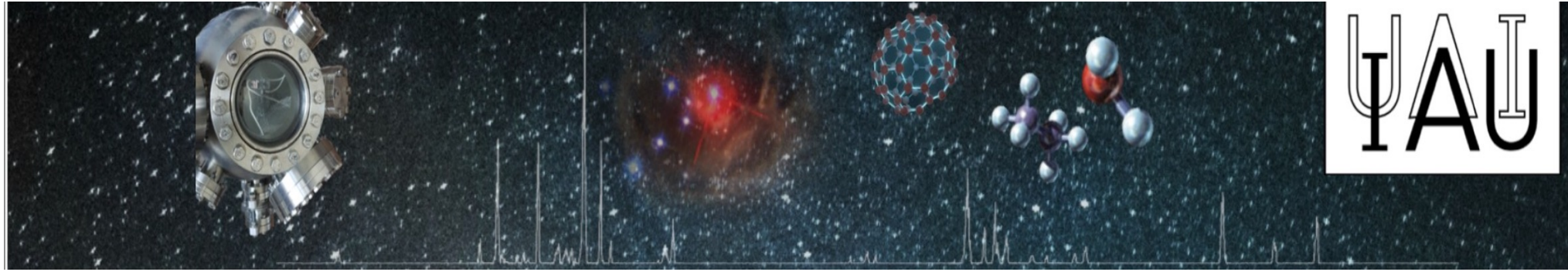
M.L. Dubernet

LERMA, Observatory of Paris, PSL University, CNRS

[marie-lise.dubernet@observatoiredeparis.psl.eu](mailto:marie-lise.dubernet@observatoiredeparis.psl.eu)



# Members



- Marie-Lise Dubernet, Observatory of Paris, France - chair
- Bruce Berriman, IVOA representative, California Institute of Technology, USA - co-chair
- Farid Salama, NASA-Ames Research Center, USA
- Jiayong Zhong, Beijing Normal University, Dept Astronomy, China
- Friedrich Kupka, University of Applied Sciences Technikum Wien, Austria
- Randall Smith, Smithsonian Astrophysical Observatory, USA

# This afternoon program : F.A.I.R. Aspects and session scientifically oriented towards molecules/ice/dust

## **13.30 – 15.00 Session 5A: Laboratory Astrophysics Databases: from the provider to the user: encouraging FAIRness.**

**Chair: ML Dubernet and B. Berriman Room 101**

13.30 – 13.40 WG Activities and Plans (ML Dubernet)

13.40 – 13.45 FAIR principles in VAMDC (ML Dubernet)

13.45 – 13.55 FAIR principles in IVOA (B. Berriman)

13.55 – 14.10 The NASA Ames PAH IR Spectroscopic Database (Christiaan Boersma)

14.10 – 14.25 Laboratory Astrophysics Databases on Grains and Ices: From the Laboratory to the Users (Cornelia Jäger)

14.25 – 14.40 About the atomic and molecular databases in the planetary community (Miriam Rengel)

14.40 – 14.55 Examples of astro analysis tools: ENIIGMA (Will Rocha)

14.55 – 15.00 General Discussion and concluding remarks

This talk :

**Inter-Commission B2-B5 WG Activities and Plans**

**“Laboratory Astrophysics Databases: from the provider to the user: encouraging FAIRness” - 8th August – 1:30pm-1:45pm**

SOC : WG Members and Prof. Beatriz Barbuiz, Prof. Masatoshi Ohishi, Prof. Dongsu Ryu

# Introduction

- A large variety of atomic, molecular, solid data are used in Astrophysics for the analysis of astronomical spectra and for the development of astronomical models.
- Many resources have been created over time :
  - Many databases that contain spectroscopic data and data for other microscopic processes, for stellar, interstellar, planetary, galactic, solar and plasmas environments.
  - Many Astrophysical data analysis software and modelisation software that include usage of laboratory A&M&Solid data are available
  - Efforts have been made concerning the interoperability of databases through different initiatives:
    - **the International Virtual Observatory Alliance (IVOA)** community develop standards including standard description of gas-phase spectral data collected with astronomical observatories : still on-going efforts
    - **Virtual Atomic & Molecular Data Center (VAMDC)** defined standards, develop software and databases, interconnect heterogeneous databases for a wide range of processes and chemical species
- And standards are further developed with CODATA, IUPAC, NIST, ..

# F.A.I.R. Issues

→ FAIR principles ensure Tracability of A&M&Solid Data that is an essential component to reproduce the analysis of observed data and the modelisation of the astrophysical objects

- **F**indable : Unique & persistant identifier for data and metadata, rich metadata, registration of metadata and data
- **A**ccessible : standardised, open, free communication protocols to retrieve data/metadata & metadata still available when data no longer available
- **I**nteroperable : data/metadata use a formal, accessible, shared and broadly applicable language for knowledge representation, FAIR vocabularies
- **R**e-usable : data/metadata richly described with accurate & relevant attributes, associated with detailed provenance, meet domain-relevant community standards





- Objectives : provide a platform where to discuss the FAIR publication and usage of laboratory astrophysics data in astronomy and astrophysics
- Timing :
  - Year 1 : to provide a state-of-the-art report on the existing infrastructures, databases, codes, virtual environments, standards
  - Year 2 : to identify the bottlenecks in providing the data to the astronomical community and in the FAIR usage of the laboratory astrophysics data by the astronomical community
  - Year 3 : to provide practical and political recommendations related to optimizing the process from laboratory data to astrophysics and vice-versa

# Year 1 - State-of-the art : databases, codes, tools, virtual environments, F.A.I.R initiatives

- 1) Collect of Information :
  - [A preliminary overview at the Keynote Lecture “Atomic and Molecular Databases”, Symposium 371, 9th August](#)
  - [Report Year 1](#) : available 15th December 2022 from IAU Website
- 2) 2 surveys : Identifying resources and some questions related to FAIR launched mid-June 2022
  - Databases : currently about 19 answers (14 sent directly + VAMDC)
  - Codes/Tools : currently about 8 answers (12 asked directly + mail to IVOA)
  - Preliminary analysis of the surveys
- 3) FAIR Survey of VAMDC e-infrastructure (asked to C.M. Zwölf, Paris Observatory)
- 4) Contacts with two international body for standardisation
  - **IVOA for Astro** : “How to include the citation of data in their standards” : to be reviewed by their science committee → crucial issue for the A&M&Solid community that provides data.
  - **IUPAC for Chemistry FAIR standards** : VAMDC as a usecase of the WorldFAIR project

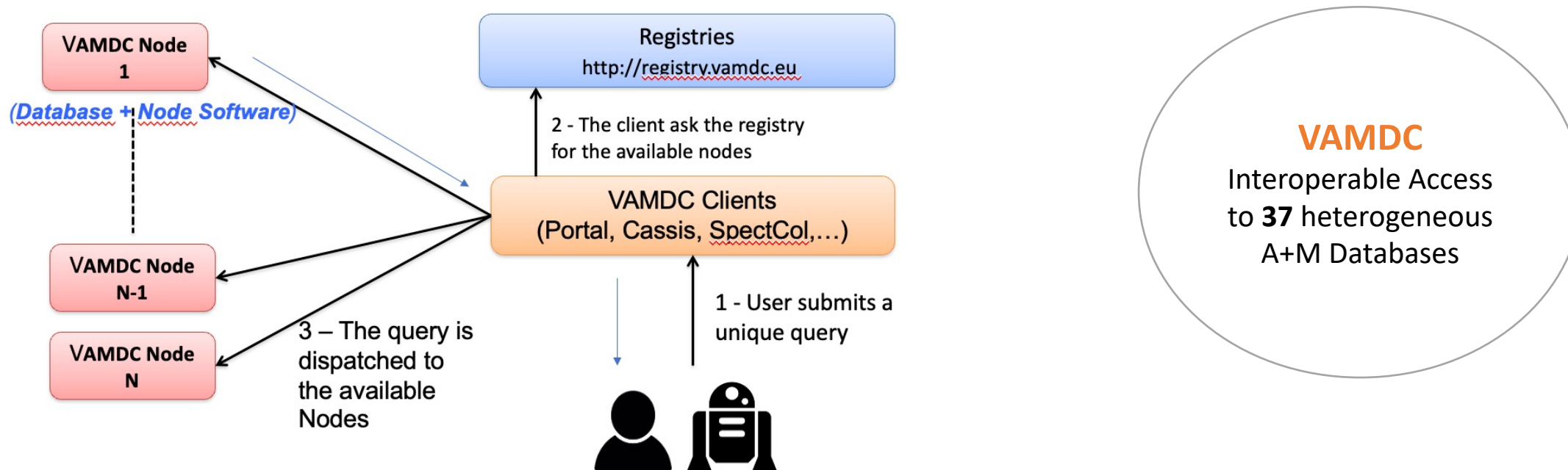
# Preliminary results From surveys : first criteria for reproducible, citable and quality science

- Among the 19 databases
  - 100% : “Data Citation”
  - 10% : versioning of datasets
- Among the 8 codes/tools
  - 25% : “Data Citation”
  - 65% : “versioning of datasets”
- First obvious recommendations :
  - Databases : have versioning (and keep previous versions of datasets)
  - Codes/Tools : implement easy/pertinent data citation



## FAIR Principles in VAMDC and Preliminary Analysis

M.L. Dubernet & C.M. Zwölf



Paper « A decade with VAMDC : results and ambition, Atoms, 2020 »

<http://dx.doi.org/10.3390/atoms8040076>.

# VAMDC and the FAIR principle

## Conceptually

FAIR (Findability, Accessibility, Interoperability, Reuse) principles underlined the design of VAMDC

- **Findable** : data coming from the infrastructure can be tagged with persistent unique identifiers, are described with rich metadata and are indexed into public registries
- **Accessible** : the extraction query relies on open, documented standards
- **Interoperable** : the data extracted from VAMDC are formatted using the XSAMS standard (some aspects need improvement)
- **Re-usable** : the provenance and sources of all the data are documented in each data set extracted from VAMDC. Data tools are provided to convert VAMDC data into widely adopted community data formats (can be further developed)

Based on RDA recommendations for FAIR across disciplines

[FAIR Data Maturity Model Specification and Guidelines](#) (RDA recommendation : DOI: 10.15497/rda00050, **Co-chairs**: Edit Herczog, Keith Russell, Shelley Stall)

Analysis of FAIRness in VAMDC based by C.M. Zwölf using that recommendation (July 2022)

- On 20 Essential Criteria : 17 are mature
- On 14 Important Criteria : 6 are mature, the other in implementation, under consideration or not applicable
- On 7 Useful Criteria : 4 are mature

VAMDC is FAIR within its ecosystem but lack of machine actionable metadata necessary cross-disciplinary FAIRness

\*\* Recommendation : to be improved

# To go further for VAMDC.org

Use previous analysis to improve inter-disciplinary FAIRness :

VAMDC is FAIR within its ecosystem but lack of machine actionable metadata necessary cross-disciplinary FAIRness

**Recommendation** : improve machine actionable metadata

Step 1 : VAMDC.org as a usecase in the Chemistry WP for general indexing of chemical metadata



<https://iupac.org/worldfair-global-cooperation-on-fair-data-policy-and-practice/>





## Short term plan of the WG

- Get more resources into the surveys
- Finalize report on State-of-the art & Analyse Results – Yr1 & 2
- Include news members from Plasmas, Nuclear & Particle Physics who can provide info about their own activities with respect to FAIR issues

### About interaction with community

- Do we need an external website of the WG ? → If yes, we **would need inputs from community about what kind of info is useful**
- **Anybody is welcome to participate to the work of the WG !**
- **Any suggestions/Ideas are most welcome !**

## Mid-term/Final plans of the WG

- To publish guidances/recommendations in 2 years to
  - Encourage evolution of current standards or develop new standards
  - Encourage inclusion of data in existing databases, infrastructures, repositories that are FAIR or nearly FAIR compliant
  - Encourage the databases/infrastructure to improve FAIR aspects for distribution
  - Encourage codes/tools as secondary publishers of A&M&Solid data to follow FAIR principles



# Long term Goal

“A better science and sustainable world”

Global Network of A&M&Solid Data for Astrophysics (and connected to other application fields as well)

- ◆ Part I : Build an Information system on lab. Groups, projects, needs, etc ..
  - To be built upon the generalisation of this morning's presentations
  - Support/Extend the <http://astrochemistry.eu> initiative
- ◆ Part II : Build a General indexing of resources
  - To be built upon the analysis of the WG

What next ?

- To start an Interest Group at Research Data Alliance (<https://www.rd-alliance.org/>) in order to define the specifications → Brainstorming

Questions/Ideas from the audience : slack channel, Q&A, end of session, send mails to [marie-lise.dubernet@observatoiredeparis.psl.eu](mailto:marie-lise.dubernet@observatoiredeparis.psl.eu)