3rd FAIR Digital Objects Conference

Poster Session

https://doi.org/10.52825/xxxx..... DOI placeholder (WILL BE FILLED IN BY TIB Open Publishing)

Published: (WILL BE FILLED IN BY TIB Open Publishing)

Poster: GROMACS MetaDump: Automatic FAIR-Ready Metadata Extraction for Molecular-Dynamics Simulations

Poster Session

Adrián Rošinec[1,2] https://orcid.org/0000-0002-7748-5590, Terézia Slanináková[2,*] https://orcid.org/0000-0003-0502-1145, Matej Antol[2] https://orcid.org/0000-0002-1380-5647 and Jakub Čillík[2,*] https://orcid.org/0009-0001-7780-3317

[1]CEITEC - Central European Institute of Technology, Masaryk University, Kamenice 753/5, 625 00 Brno, Czech Republic

[2]Institute of Computer Science, Masaryk University, Šumavská 525/33, 602 00 Brno, Czech Republic

*Correspondence:

Terézia Slanináková, slaninakova@ics.muni.cz

Jakub Čillík, cillik@ics.muni.cz

Abstract. Molecular dynamics (MD) simulations are being deposited in increasing numbers across generalist and emerging domain-specific repositories. However, the absence of consistent, machine-actionable metadata limits the findability, comparability, and reuse of these datasets. Critical information about simulation setup — such as force fields, water models, system composition, simulation length, and provenance — is often embedded only in binary input files or informal documentation. This makes it difficult for researchers, data infrastructures, and automated workflows to reliably assess or integrate simulation results.

GROMACS MetaDump[1] is a tool that automatically extracts structured metadata directly from GROMACS simulation input files and outputs them in JSON or YAML formats. The generated metadata describe system parameters, simulation conditions, provenance, and software configuration in a standardized and interoperable form. This enables MD datasets to be indexed, compared, validated, and reused with significantly reduced manual effort.

In this poster, we show how the metadata schema used in MetaDump can serve as the basis for a FAIR Digital Object (FDO) Type for MD simulations. When combined with persistent identifiers already present in common repositories (e.g., DOIs), MetaDump metadata provides all components required to instantiate MD simulations as FDOs: stable identity, machine-actionable description, and resolvable access to related data files. The approach supports STAP trust principles by improving transparency and accountability in simulation workflows.

We present example use cases and discuss how this schema can be adopted in federated data spaces and MD community infrastructures. MetaDump therefore offers a concrete and incremental pathway toward FDO-aligned, interoperable, and reusable MD simulation data.

Keywords: Molecular Dynamics, GROMACS, Metadata Extraction, FAIR Digital Object, Machine-Actionable Metadata, Data Interoperability, STAP Trust Principles, Federated Data Spaces, Repository Indexing, Simulation Reproducibility

References

[1] Rošinec, A., Slanináková, T., Pavlík, T. et al. Gromacs MetaDump: a tool for extracting GROMACS simulation metadata. J Cheminform 17, 160 (2025). https://doi.org/10.1186/s13321-025-01082-5