The Journal of Open Source Software

Project Chameleon: A Python Package for Interoperability in Experimental Materials Science Research Data

Peter Cauchy [1,2,3,4], **David Elbert** [1,2], **and Tyrel McQueen** [1,2,3,4]

1 PARADIM, William H. Miller III Department of Physics and Astronomy, Johns Hopkins University, Baltimore 21218, MD **2** Hopkins Extreme Materials Institute, Johns Hopkins University, Baltimore, MD 21218 **3** Department of Chemistry, Johns Hopkins University, Baltimore 21218, MD **4** Department of Materials Science and Engineering, Johns Hopkins University, Baltimore 21218, MD

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Rachel Kurchin

Reviewers:

- @Grant-Xue
- @mxwalbert

Submitted: 29 July 2025
Published: unpublished

Summary

Project Chameleon is a versatile Python package for extensible data transformation and interoperability. Chameleon streamlines analysis and combination of scientific data from different sources and provides a foundational pillar for implementation of the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles (Wilkinson et al., 2016). In the context of FAIR, interoperability refers to the ability of diverse datasets, tools, and systems to seamlessly interact, integrate, and exchange data. Interoperability ensures that scientific data can be readily combined or used alongside related datasets, software, and analytical tools across different domains and disciplines.

A central barrier to interoperability arises from heterogeneous data formats and differing metadata conventions encoded in the varied file formats of different laboratories. Project Chameleon transformations eliminate this barrier to provide effective interoperability that facilitates comprehensive analyses and machine learning that accelerate scientific discovery and enhance reproducibility of scientific findings.

Project Chameleon provides a collection of conversion scripts accessed through a REST API that enables researchers to simply convert specialized, often manufacturer-specific file formats for data such as RHEED electron diffraction patterns, scanning transmission electron microscopy (STEM) arrays, molecular beam epitaxy (MBE) logs, magnetic property measurement system analyses (PPMS/MPMS), and X-ray diffraction patterns into universally accessible formats like CSV for tabular data and PNG for image data. These universal formats provide the research community the ability to use open and readily available tools to analyze and visualize data from different instruments or laboratories. The API for Project Chameleon provides a function-based, user-friendly interface that allows simple conversions of some common file types. The API is also designed to handle multiple input and output types, including raw bytes, JSON, and URLs, giving flexibility in how it can be implemented. Project Chameleon itself leverages several open source packages for conversion and processing of specialized formats (HTMDEC, 2025; Peña & others, 2025; Savitzky & others, 2021; Wojdyr, 2025).

Statement of need

The U.S. interagency Materials Genome Initiative (Brinson et al., 2024; National Academies of Sciences, Engineering, and Medicine, 2024; Odegard et al., 2023; Pablo et al., 2019; U.S. White House Office of Science and Technology Policy, 2021) demands a leap in FAIR-compliant data to accelerate materials discovery, AI/ML applications, high-throughput experimentation,

advanced digital twins, and growing interest in autonomous laboratories and manufacturing.

Fundamental research in materials science, chemistry, and solid-state physics, however, depends on structure and property characterizations that use a wide range of techniques including, but not limited to, diffraction, microscopy, and spectroscopy. While these methods are a common thread in modern science, they rely on a wide variety of laboratory instrumentation developed by a wide range of manufacturers, creating a multitude of data formats demanding use of specialized analysis tools. The wide variation in data formats hinders FAIR interoperability; the inability to combine data from different sources or reproduce analyses done in different labs thwarts Open Science and MGI goals.

Project Chameleon was designed to be used by materials science researchers, but is applicable to any field of research using instrumental characterization tools. Project Chameleon is already in use in the NSF PARADIM Materials Innovation Platform (Cornell and Johns Hopkins universities) as well as the McQueen Lab at Johns Hopkins where the ARPES, Bruker RAW, Bruker BRML, 4D STEM, Non-4D STEM, RHEED, MBE, PPMS/MPMS, Laue HS2, and JEOL SEM functions have been implemented. These functions are used in the lab file server for McQueen Lab and to create FAIR datasets published with DOIs in the PARADIM Data Portal (Platform for the Accelerated Realization, Analysis, and Discovery of Interface Materials (PARADIM), 2025), allowing for the download of the raw file or converted file without any work done on the user's end.

Acknowledgements

We acknowledge contributions and support from Maggie Eminizer, Matt Turk, Ali Rachidi, Steve Zeltman, Ben Redemann, Abby Neil, and Evan Crites.

References

Brinson, L. C., Bartolo, L. M., Blaiszik, B., & others. (2024). Community action on FAIR data will fuel a revolution in materials research. *MRS Bulletin*, *49*, 12–16. https://doi.org/10.1557/s43577-023-00498-4

HTMDEC. (2025). *Htmdec/htmdec_formats*. https://github.com/htmdec/htmdec_formats.

National Academies of Sciences, Engineering, and Medicine. (2024). *Foundational research gaps and future directions for digital twins*. The National Academies Press. https://doi.org/10.17226/26894

Odegard, G. M., Liang, Z., Siochi, E. J., & others. (2023). A successful strategy for MGI-inspired research. *MRS Bulletin*, *48*, 434–438. https://doi.org/10.1557/s43577-023-00525-4

Pablo, J. J. de, Jackson, N. E., Webb, M. A., & others. (2019). New frontiers for the materials genome initiative. *Npj Computational Materials*, *5*, 41. https://doi.org/10.1038/s41524-019-0173-4

Peña, F. de la, & others. (2025). *Hyperspy/hyperspy: V2.3.0*. Zenodo. https://doi.org/10.5281/zenodo.14956374

Platform for the Accelerated Realization, Analysis, and Discovery of Interface Materials (PARADIM). (2025). *PARADIM data portal*. https://data.paradim.org/index.php.

Savitzky, B. H., & others. (2021). py4DSTEM: A software package for four-dimensional scanning transmission electron microscopy data analysis. *Microscopy and Microanalysis*, *27*, 712. https://doi.org/10.1017/S1431927621003251

U.S. White House Office of Science and Technology Policy. (2021). *Materials*

genome initiative strategic plan. https://www.mgi.gov/sites/default/files/documents/MGI-2021-Strategic-Plan.pdf.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., & others. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*. https://doi.org/10.1038/sdata.2016.18

Wojdyr, M. (2025). *Wojdyr/xylib*. https://github.com/wojdyr/xylib.