Semantic Data Management for Synchrotron Science: Materials Data Science Ontology (MDS-Onto) for FAIR, Learning and Reasoning

Roger H. French[1,2,3,*], Redad Mehdi[1,2], Van D. Tran [1,2], Gabriel O. Ponon [1,2], Jonah A. Bachman [1,2], Weiqiu Yue[1,3], Kyle R. Henrikson [1,2], Johan Buecker[1,2], Laura S. Bruckman [1,2], Frank Ernst [1,2], Neamul H. Khansur [1,2], Matthew A. Willard [1,2], Yinghui Wu [1,3], Quynh D. Tran [1,2], Erika I. Barcelos[1,2] and Ozan Dernek[1,2]

[1]*Materials Data Science for Stockpile Stewardship Center of Excellence, Case Western Reserve University, 10900 Euclid Ave, Cleveland OH 44106, USA)*

[2]*Materials Science and Engineering Department, Case School of Engineering, Case Western Reserve University, 10900 Euclid Ave, Cleveland OH 44106, USA)*

[3]*Computer and Data Sciences Department, Case Western Reserve University, 10900 Euclid Ave, Cleveland OH 44106, USA)*

Abstract

The foundations of Materials Data Science (MDS) computing, including 1) the integration of distributed (e.g., Cloudera, Hadoop) and high performance computing, 2) neural network training engines (e.g., Nvidia AISC), and 3) RDF triplestore Graph Databases (e.g., Ontotext's GraphDB), mean that statistical and deep learning on sparse, massive, and historical datasets is becoming possible. Study Protocols are an essential feature of robust, data-centric, MDS studies, where data FAIRification is enabled by MDS-Onto [1] ontology and tools, transforming messy data into linked data [2,3] across Study Stages, including sample, tool, recipe, pre-processing, analysis, modeling and results publishing. These concepts, relationships and stages guide the research process allowing exchange of JSON-LD files and their deserialization with Apache Arrow into language-independent dataframes for analysis, modeling and reasoning.

In Synchrotron Science, we have focused on Xray diffraction and Xray Compute Tomography studies, where terabyte size datasets are commonly acquired, and scientist users have struggled to completely analyze these massive datasets. Using a Study Protocol/Study Stage approach enables automated data analysis pipelines (e.g. FAIRshake [4]) and Apache Airflow workflow automation. We illustrate this for three different materials studies, one on Xray diffraction of Ti3Nb and two on Ti-6Al-4V, done respectively at DOE's Advanced Photon Source (APS) and NSF's Cornell High Energy Synchrotron Source (CHESS).

Keywords

FAIRification, Low-level Ontology, Machine Learning, Machine Reasoning Distributed Computing High Performance Computing Xray Diffraction Xray Computed Tomography

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

SeMatS 2025: The 2nd International Workshop on Semantic Materials Science, co-located with the 24th International Semantic Web Conference (ISWC 2025), November 2, 2025, Nara, Japan

*Corresponding author.

rxf131@case.edu (R. H. French); mxm1684@case.edu (R. Mehdi); vxt101@case.edu (V. D. T.); gop2@case.edu (G. O. P.); jab433@case.edu (J. A. B.); wxy215@case.edu (. W. Yue); krh117@case.edu (K. R. H.); jxb1224@case.edu (J. Buecker); lsh41@case.edu (L. S. B.); fxe5@case.edu (F. E.); nxk709@case.edu (N. H. K.); maw169@case.edu (M. A. W.); yxw1650@case.edu (Y. W.); qdt@case.edu (Q. D. T.); eib14@case.edu (E. I. Barcelos); oxd37@case.edu (O. Dernek)

https://mds3-coe.com/ (R. H. French); https://mds3-coe.com/ (R. Mehdi); https://mds3-coe.com/ (V. D. T.); https://mds3-coe.com/ (G. O. P.); https://mds3-coe.com/ (J. A. B.); https://mds3-coe.com/ (. W. Yue); https://mds3-coe.com/ (K. R. H.); https://mds3-coe.com/ (J. Buecker); https://mds3-coe.com/ (L. S. B.); https://mds3-coe.com/ (F. E.); https://mds3-coe.com/ (N. H. K.); https://mds3-coe.com/ (M. A. W.); https://mds3-coe.com/ (Y. W.); https://mds3-coe.com/ (Q. D. T.); https://mds3-coe.com/ (E. I. Barcelos); https://mds3-coe.com/ (O. Dernek)

0000-0002-6162-0532 (R. H. French); 0000-0002-0593-5222 (R. Mehdi); 0009-0008-4355-0543 (V. D. T.); 0000-0002-3253-7991 (. W. Yue); 0009-0009-4765-4945 (K. R. H.); 0009-0004-6150-3223 (J. Buecker); 0000-0003-1271-1072 (L. S. B.); 0000-0002-4823-2041 (F. E.); 0000-0001-8769-3329 (N. H. K.); 0000-0001-5052-8012 (M. A. W.); 0000-0003-3991-5155 (Y. W.); 0009-0006-4025-1834 (Q. D. T.); 0000-0002-9273-8488 (E. I. Barcelos); 0000-0002-5071-9684 (O. Dernek)

1. Citations

1. B. P. Rajamohan et al., “Materials Data Science Ontology(MDS-Onto): Unifying Domain Knowledge in Materials and Applied Data Science,” Sci Data, vol. 12, p. 628, Apr. 2025, doi: 10.1038/s41597-025-04938-5.

2. Van D. Tran et al., “FAIRLinked: Transform research data into FAIR-compliant RDF Data.” July 2025. Available: https://pypi.org/project/FAIRLinked/

3. Gabriel O. Ponon, et al., “CEMENTO: A package to view and write ontologies directly from draw.io diagram files.” July 27, 2025. Available: https://pypi.org/project/cemento/.

4. Finley Holt, et al., “FAIRshake: For the FAIR analysis of XRay Datasets.” Oct. 08, 2024. Available: https://pypi.org/project/FAIRshake/.