Materials Data Repository metadata schema and cross-database federation

A. Matsuda[1*], K. Tanabe[1], M. Ishii[1], T. Kadohira[1]

[1] National Institute for Materials Science
*MATSUDA.Asahiko@nims.go.jp

To accelerate collection, accumulation, distribution, and utilization of materials data, the National Institute for Materials Science (NIMS, Japan) has been developing and operating DICE [1], a materials data platform and a collection of databases/data-oriented systems. One of the DICE services is the Materials Data Repository (MDR) [2,3], released in 2020. The public front-end of MDR is based on Hyrax by Samvera community. Works deposited to MDR are assigned DataCite DOIs while their creators are identified by ORCIDs, allowing integration with external services such as Clarivate's Data Citation Index.

Significant effort was devoted to implement a metadata schema with a focus on persistent identifiers and support for specimen, methods, and instrument metadata. The first schema was intended to be the common versatile format among multiple DICE services [4]. Despite some of its success in inter-system data transfer, its complexity induced negative user feedback. It was superseded in favor of a leaner YAML-based schema [5], developed alongside a new separate system to natively support the new schema. Works deposited to this system are automatically transferred to the public-facing Hyrax system. The new UI works in conjunction with NIMS's research registration system and makes use of CrossRef API, minimizing metadata input from researchers. This led to a tenfold increase in active users and weekly deposits, highlighting the importance of usability in software solutions.

Using works deposited to MDR as instances, we can define an ontology for data linking and federation [6]. Figure 1 shows a representation of an x-ray absorption fine structure (XAFS) spectrum in MDR XAFS DB [7]. The last line defines the specimen, where the wd: prefix refers to our vocabulary system, MatVoc. This graph can be linked with the graph defining "x-ray absorption edge" as shown in Figure 2. The link can be swapped with "core level" for hard x-ray photoelectron spectroscopy (HAXPES) [8] while retaining the schema. This interoperability allows us to easily obtain the HAXPES spectrum for a sample in MDR XAFS DB, demonstrating the ability for cross-database queries.

```
@prefix mdr-xafs: <http://dice.nims.go.jp/ontology/mdr-xafs-ont/Schema#>
@prefix obo: <http://purl.obolibrary.org/obo/>
@prefix prism: <http://prismstandard.org/namespaces/1.2/basic/>
@prefix wd: <http://matvoc.nims.go.jp/entity/>

<http://dice.nims.go.jp/ontology/mdr-ont#8a02714e-46a7-4fdc-95a5-1acb18338d7d> a mdr-xafs:Work ;
  rdfs:seeAlso <https://mdr.nims.go.jp/concern/datasets/h128nh653> ;
  rdfs:label "XAFS spectrum of Gold(III) hydroxide"@en ;
  prism:doi "https://doi.org/10.48505/nims.1602"^^xsd:string ;
  obo:RO_0000057 wd:Q1304, wd:Q1308 .
```

Figure 1. RDF representation of a XAFS spectrum in MDR.

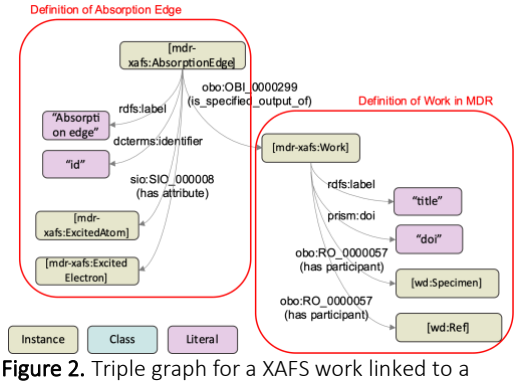

Figure 2. Triple graph for a XAFS work linked to a graph defining x-ray absorption edge.

This work was partly supported by MEXT Program: Data Creation and Utilization-Type Material Research and Development Project (Digital Transformation Initiative Center for Magnetic Materials) Grant Number JPMXP1122715503.

References

[1] DICE, National Institute for Materials Science, https://dice.nims.go.jp/
[2] MDR (NIMS Materials Data Repository), https://mdr.nims.go.jp/
[3] P. Walk, M. Tanifuji, K. Tanabe, A. Matsuda, R. Jones, A. Ranganathan, *14th Int. Conf. Open Repos.*, **2019**. doi:10.5281/zenodo.3553963
[4] DICE Common Message Format Schema, **2019**, doi:10.48505/nims.3240
[5] MDR Schema, **2022**, doi:10.48505/nims.3239
[6] DICE Common Namespace, https://dice.nims.go.jp/ontology/about.html (accessed February 28, 2023)
[7] M. Ishii, H. Nagao, K. Tanabe, A. Matsuda, H. Yoshikawa, "MDR XAFS DB" **2021**, doi:10.48505/nims.1447
[8] S. Yasuno, T. Matsumoto, H. Nagao, H. Yoshikawa, "MDR HAXPES DB" **2021**, doi:10.48505/nims.3056