

- PubChem subproject

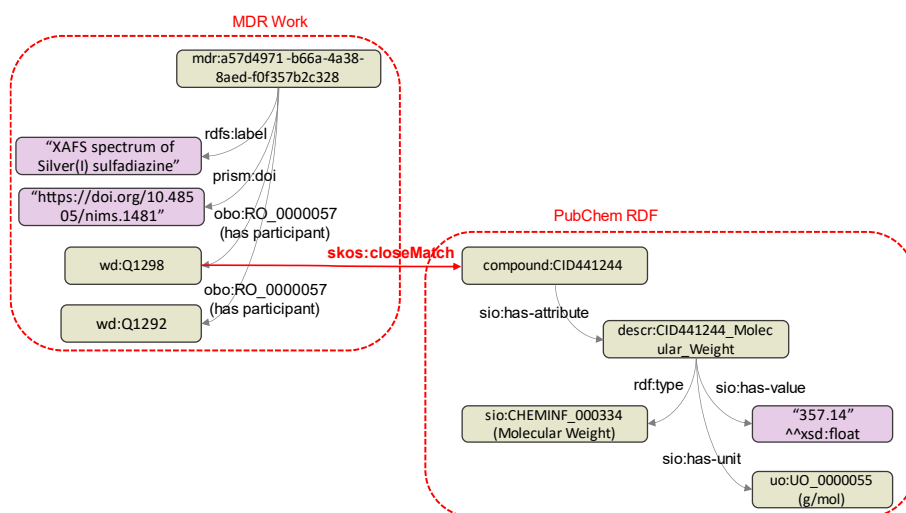
<http://dice.nims.go.jp/ontology/mdr-xafs-ont/Subproject#toPubChem>

As discussed in

<http://dice-internal.nims.go.jp/ontology/about.html#mdr-xafs-ont>

we defined namespace of MDR XAFS DB (<https://doi.org/10.48505/nims.1447>) to work with external data and to enable advanced searching. Here, we introduce the toPubChem subproject, which utilizes that namespace.

The aim of this subproject is to add chemical information to XAFS spectral data by linking the MDR XAFS DB with the external database PubChem. An instance level Resource Description Framework (RDF) schema to realize this aim is shown below. This example shows the RDF linkage for Silver(I) sulfadiazine in the MDR XAFS DB, when the molecular weight in PubChem is used in the linkage.



The material ID defined in the MDR XAFS DB (wd:Q1298 in this example) is mapped to the Compound ID in the PubChem RDF (compound:CID441244 in this example) using skos:closeMatch.

Once we are in the PubChem RDF schema

https://pubchemdocs.ncbi.nlm.nih.gov/rdf#_1-3

we can use the 23 descriptors, including the molecular weights shown in this example, and related information from that database.

The prefixes in this schema are as follows:

Prefix	Name space
compound	http://rdf.ncbi.nlm.nih.gov/pubchem/compound/

descry	http://rdf.ncbi.nlm.nih.gov/pubchem/descriptor/
mdr	http://dice.nims.go.jp/ontology/mdr-ont#
obo	http://purl.obolibrary.org/obo/
prism	http://prismstandard.org/namespaces/1.2/basic/
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
sio	http://semanticscience.org/resource/
skos	http://www.w3.org/2004/02/skos/core#
uo	http://purl.obolibrary.org/obo/
wd	http://dice.nims.go.jp/ontology/mdr-xafs-ont/Item#
xsd	http://www.w3.org/2001/XMLSchema#

The published RDF files in ttl format according to this schema are as follows:

File name	Version	Triples	Description
Open_RDF_toPubChem.ttl	0.1	55	An RDF set for link to PubChem sub-project, Definition of semantic links between materials measured in MDR XAFS DB and PubChem compound ID.
Open_RDF_MdrWork.ttl	0.1	13,051	An RDF set for On Top of Single Figure (OTSF) project, Definition of works in MDR.

SPARQL examples linked to PubChem can be found in the following file.

query_PubChem.txt

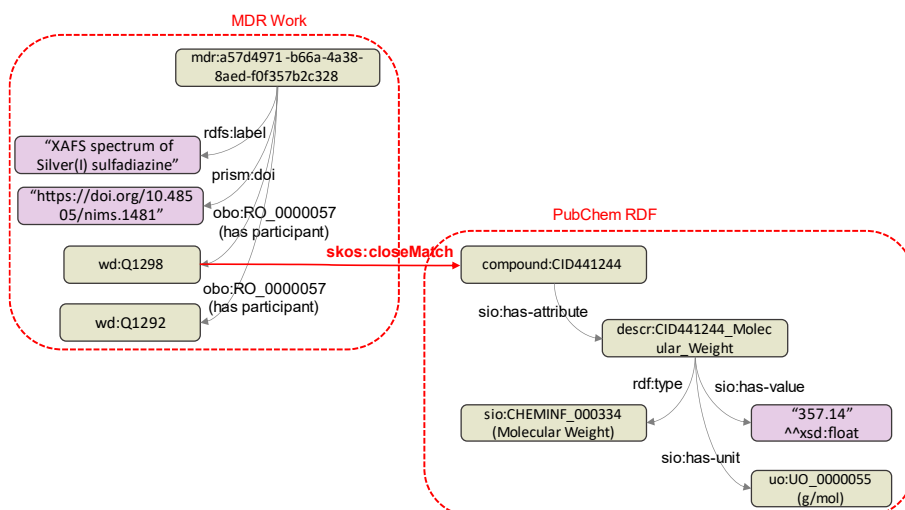
- PubChem サブプロジェクト

<http://dice.nims.go.jp/ontology/mdr-xafs-ont/Subproject#toPubChem>

<http://dice-internal.nims.go.jp/ontology/about.html#mdr-xafs-ont>

で紹介したように、MDR XAFS DB (<https://doi.org/10.48505/nims.1447>) は外部のデータと連携し、高度な検索を可能にするために、名前空間を定義した。ここでは、その名前空間を活用する toPubChem サブプロジェクトを紹介する。

このサブプロジェクトの狙いは、MDR XAFS DB を外部のデータベース PubChem と連携させることにより、XAFS スペクトルデータに化学情報を付加することである。それを実現するインスタンスレベルの RDF (Resource Description Framework) スキーマは次のとおりである。この例では、MDR XAFS DB の Silver(I) sulfadiazine について、PubChem の分子量を利用する場合の RDF 連携を表している。



MDR XAFS DB で定められた材料の ID (この例では wd:Q1298) が skos:closeMatch で PubChem RDF の Compound ID (この例では compound:CID441244) に対応付けられている。一旦 PubChem RDF のスキーマに入れば、

https://pubchemdocs.ncbi.nlm.nih.gov/rdf#_1-3

に従って、この例で示す分子量を含めた 23 の記述子や同データベースの関連情報を利用することができる。

このスキーマにおける prefix は以下のとおりである：

Prefix	Name space
compound	http://rdf.ncbi.nlm.nih.gov/pubchem/compound/
descr	http://rdf.ncbi.nlm.nih.gov/pubchem/descriptor/

mdr	http://dice.nims.go.jp/ontology/mdr-ont#
obo	http://purl.obolibrary.org/obo/
prism	http://prismstandard.org/namespaces/1.2/basic/
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
sio	http://semanticscience.org/resource/
skos	http://www.w3.org/2004/02/skos/core#
uo	http://purl.obolibrary.org/obo/
wd	http://dice.nims.go.jp/ontology/mdr-xafs-ont/Item#
xsd	http://www.w3.org/2001/XMLSchema#

このスキーマに従って公開されている ttl 形式の RDF ファイルは以下のとおりである：

ファイル名	バージョン	トリプル数	記述
Open_RDF_toPubChem .ttl	0.1	55	An RDF set for link to PubChem sub-project, Definition of semantic links between materials measured in MDR XAFS DB and PubChem compound ID.
Open_RDF_MdrWork.ttl	0.1	13,051	An RDF set for On Top of Single Figure (OTSF) project, Definition of works in MDR.

PubChem と連携させた SPARQL 例は、次のファイルに記載してある。
query_PubChem.txt