Exploring the FAIR Digital Object Framework within Materials Science and Engineering

Zachary Trautt, Material Measurement Laboratory, NIST

ORCID: 0000-0001-5929-0354
GitHub: ztrautt
Twitter: @ztrautt
Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.
Materials are Technology Enablers

There would be no...

- Skyscrapers without **steel girders**
- Commercial aviation industry without **high-strength aluminum alloys** and **polymer composites**
- Information age without **silicon**
- Mobile phones without **functional ceramics**
- Solar electricity without **photovoltaic materials**
- Modern medicine without **biocompatible soft materials**
Research Priorities

Current:
• Ubiquitous use of persistent identifiers for all research objects
• Digital Object Metadata, Linked Data, and Semantic Web
• Digital Object Architecture Vision

Future:
• Operations on Digital Objects
• Versioning Digital Objects
• Data Type Registries
• PID Kernel Metadata
• Connection to RO-Crate
Pilot Projects

Materiom – Material recipes/data for locally-abundant supply chains

Microstructure/PSP Repository

HTE Materials Data Portal
Microstructure Repository Workshops

Workshop #1 (November 2018)
- Define scope
- Survey landscape
- **Metadata** for diversity of images and multi-spectral data
- **Metadata** to support queries for
  - Images and data
  - Analysis
  - Samples

Workshop #2 (May 2019)
- Prototype demo
- **Metadata** for
  - Images and data
  - Samples
  - Instruments
  - Analytics
- Presentations of stakeholder projects
Instruments

- Inst. Type (SEM, TEM, ...)
- Name/Model
- Setup/State
- Operation Parameters
- Divisible Components

Concerns
- File Type/Format
- Last Calibration/History
- Accuracy/Confidence
- Automated Population as much as possible

Filtering Queries
- How to Collect Data w/o Burdening User
- Instrument Registry
- Facilities & Vendors
- Responsive Interface (Basic, Adv.)

How to Incentivize User?
Instruments

**What does it take/need to find and access data?**

- **Instrument Type** (SEM, TEM, ...
- **Name/Model**
- **Set up / hardware**
- **Operation/parameter**

**Concerns**
- **File type/format**
- **Filtering queries**

**Last calibration/history**

**Accuracy/confidence**

How to incentivise user?
Instruments

- Inst. Type (SEM, TEM, ...
- Name/Model
- Setup/State
- Operation Parameters
- Divisible Components

Concerns
- File type/format
- Last calibration/history
- Accuracy/confidence
- Automated population as much as possible

Filtering Queries

What does it take/need to find & access data?
Samples

SAMPLE META For the Repository

HISTORY
Source Material
Derived Material
Age

Level 2
Composition [e.g. dopants]
Complex Type
Form (Solid, Amorphous...)
Crystal Systems (cubic, monoclinic...)
Polymorphic Class (thermoset, thermoplastic...)
Phase
Composition
Crystal Structure

Level 1
PID
Material
Syn & Processing
Characteristics
Property (Y/N)
Microstructure
(particulate size, grain size)
Availability

G1: Find images with <Material>
G2: Find images with <synthesis:processing>
G3: Find images with <char. tech>
G4: Find images with <properties>

Cancer:
G5 + G2 + G3 + G4
G1x: G1 and composition
G2y: G2 and form
G5: Images with range [x,y] with <prop: strength>

Range, grain size, range size, and prop: strength
Micro, particle size dist., micro. dispersion
Samples

![Diagram of SAMPLE META and HISTORY sections]

- Level 1
  - PID
  - Material
  - Syn & Processing
  - Characteristics
  - Property (Y/N)
  - Microstructure (dissolution, matrix, etc., particle size, grain size)
  - Availability

- Level 2
  - Composition (e.g., dopants)
  - Complex Type
  - Form (Sols, Amorphous...)
  - Crystal System (cubic, monoc...)
  - Polymer Class (thermoset, thermoplastic)
  - Phase
    - Composition
    - Crystal Structure

- HISTORY
  - Source Material
  - Derived Material
  - Age
Example
• Diffractometer
  • X-ray Source
    • Interchangeable different anode materials: Ag, Mo, Cu, Co, Cr, etc.
  • X-ray Detector
    • Interchangeable different types: area, multi-mode 0D/1D/2D, etc.
• Sample Stage
  • Interchangeable many types
• Sample Holder
  • Custom made by staff machinist
Example

• AM metal powders are input to AM Laser Powder Bed Fusion Process
• In-situ Process monitoring of melt pool generates high-frame rate image data
• AM Laser Powder Bed Fusion Process has output of a part or test artifact
Material Characterization

Example
- Material sample(s) may undergo many measurements via many instruments:
  - Microscopy
  - Mechanical testing
  - Differential scanning calorimetry
  - Corrosion resistance
  - Etc.
Opportunity: Leverage/Extend Schema.org

Welcome to Schema.org

Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond.

Schema.org vocabulary can be used with many different encodings, including RDFa, Microdata and JSON-LD. These vocabularies cover entities, relationships between entities and actions, and can easily be extended through a well-documented extension model. Over 10 million sites use Schema.org to markup their web pages and email messages. Many applications from Google, Microsoft, Pinterest, Yandex and others already use these vocabularies to power rich, extensible experiences.

Founded by Google, Microsoft, Yahoo and Yandex, Schema.org vocabularies are developed by an open community process, using the public-schemaorg@w3.org mailing list and through GitHub.

A shared vocabulary makes it easier for webmasters and developers to decide on a schema and get the maximum benefit for their efforts. It is in this spirit that the founders, together with the larger community have come together – to provide a shared collection of schemas.

Research Object Crate

RO-Crate has been developed as a schema.org-based JSON lightweight approach to the next generation Research Object serialization.
Full Hierarchy

Schema.org is defined as two hierarchies: one for textual property values, and one for the things that they describe.

This is the main schema.org hierarchy: a collection of types (or "classes"), each of which has one or more parent types. Although a type may have more than one super-type, here we show each type in one branch of the tree only. There is also a parallel hierarchy for data types.

Types:

Close hierarchy / Open hierarchy

- Thing -
  - Action +
  - CreativeWork +
  - Event +
  - Intangible +
  - MedicalEntity +
  - Organization +
  - Person +
  - Place +
  - Product +
# Dataset

A Schema.org Type

Thing > CreativeWork > Dataset

A body of structured information describing some topic(s) of interest.

## Properties from Dataset

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>distribution</td>
<td>DataDownload</td>
<td>A downloadable form of this dataset, at a specific location, in a specific format.</td>
</tr>
<tr>
<td>includedInDataCatalog</td>
<td>DataCatalog</td>
<td>A data catalog which contains this dataset. Supersedes includedDataCatalog, catalog. Inverse property: dataset</td>
</tr>
<tr>
<td>issn</td>
<td>Text</td>
<td>The International Standard Serial Number (ISSN) that identifies this serial publication. You can repeat this property to identify different formats of, or the linking ISSN (ISSN-L) for, this serial publication.</td>
</tr>
<tr>
<td>measurementTechnique</td>
<td>Text or URL</td>
<td>A technique or technology used in a Dataset (or DataDownload, DataCatalog), corresponding to the method used for measuring the corresponding variable(s) (described using variableMeasured). This is oriented towards scientific and scholarly dataset publication but may have broader applicability; it is not intended as a full representation of measurement, but rather as a high level summary for dataset discovery. For example, if variableMeasured is: molecule concentration, measurementTechnique could be: “mass spectrometry” or “hmr spectroscopy” or “colorimetry” or “immunofluorescence”. If the variableMeasured is “depression rating”, the measurementTechnique could be “Zung Scale” or “HAM-D” or “Beck Depression Inventory”. If there are several variableMeasured properties recorded for some given data object, use a PropertyValue for each variableMeasured and attach the corresponding</td>
</tr>
</tbody>
</table>
### Simple Example (no PID yet)

<table>
<thead>
<tr>
<th>JSON-LD Object</th>
<th>is-a schema.org/Dataset</th>
<th>is-a Diffraction Dataset</th>
</tr>
</thead>
</table>
| `{ "@context": "https://schema.org/"},  
  "@type": "Dataset",  
  "measurementTechnique": [{  
    "@type": "DefinedTerm",  
    "name": "Diffraction"  
  }]} | True | True |
| `{ "@context": "https://schema.org/"},  
  "@type": "Dataset",  
  "measurementTechnique": [{  
    "@type": "DefinedTerm",  
    "name": "Scanning Electron Microscopy"  
  }]} | True | False |
Simple Example (backend representation)

```
{
  "@context": "https://schema.org/",
  "@type": "Dataset",
  "@id": "20.500.12772/9d61ede2fe2744f32f39",
  "name": "Example TEM Measurement",
  "material": ["20.500.12772/7c6a05450e3f56dce3aa"],
  "measurementTechnique": [
    "20.500.12772/9c7aecb1-382c-419f-89d5-95392fa32bd6",
    "20.500.12772/1b21f72952bd33ca835b"
  ],
  "variableMeasured": [{
    "@type": "PropertyValue",
    "propertyID": ["20.500.12772/5c72b85931ffd955a08c"],
    "value": 0.9,
    "unitCode": "20.500.12772/9ad4d84e879b8a718e0d"
  }]
}
```
Simple Example (backend representation)

```json
{
  "@context": "https://schema.org/",
  "@type": "Dataset",
  "@id": "20.500.12772/9d61ede2fe2744f32f39",
  "name": "Example TEM Measurement",
  "material": ["20.500.12772/7c6a05450e3f56dce3aa"],
  "measurementTechnique": ["20.500.12772/9c7aecb1-382c-419f-89d5-95392fa32bd6",
                           "20.500.12772/1b21f72952bd33ca835b"],
  "variableMeasured": [{
    "@type": "PropertyValue",
    "propertyID": ["20.500.12772/5c72b85931ffd955a08c"],
    "value": 0.9,
    "unitCode": "20.500.12772/9ad4d84e879b8a718e0d"
  }]
}
```
Simple Example (backend representation)

```json
{
  "@context": "https://schema.org/",
  "@type": "Dataset",
  "@id": "20.500.12772/9d61ede2fe2744f32f39",
  "name": "Example TEM Measurement",
  "material": ["20.500.12772/7c6a05450e3f56dce3aa"],
  "measurementTechnique": [
    "20.500.12772/9c7aecb1-382c-419f-89d5-95792fa32bd6",
    "20.500.12772/1b21f72952bd33ca835b"
  ],
  "variableMeasured": [{
    "@type": "PropertyValue",
    "propertyID": ["20.500.12772/5c72b85931ffd955a08c"],
    "value": 0.9,
    "unitCode": "20.500.12772/9ad4d84e879b8a718e0d"
  }]
}
```
Link to material sample during data curation. (this can be automated)
Dataset
This schema is for describing a Dataset in Cordra.

Material
This field is for linking to any substance (e.g., sample, specimen, material, chemical, etc.), or a larger artifact, item within a collection of specimens, etc.

Material 1 *

20.500.12043/fd81c4f5aa4a6a0da165

Name: Demo ZnO Sample Synthesized by CVD
Link to specific instrument during data curation. (this can be automated)
Dataset

This schema is for describing a Dataset in Cordra.

Synthesis/Measurement/Computational Technique or Instrument/Software Used

**Referenced Item 1**

20.500.12043/5bbe5db9e9e1eaa0726e7

Name: NIST Bruker D8 Discover 223/A232

**Referenced Item 2**

20.500.12043/4a3c1392ea1247f8b26b

Name: Cu K-alpha radiation source for NIST Bruker D8 Discover 223/A232

**Referenced Item 3**

20.500.12043/6f14aa624f0e00ba76a

Name: VÁNTEC-500 Area Detector for NIST Bruker D8 Discover 223/A232
Graph View:
Linked Research Objects (can be bi-directional)
Pilot Projects

Materiom – Material recipes/data for locally-abundant supply chains

Microstructure/PSP Repository

HTE Materials Data Portal
Combinatorial Measurement

Objects Linked to 1 NIST-Synthesized Library

Objects Linked to 3 NREL-Synthesized Libraries

Objects Linked to MRR Vocabulary

Sample Position

Deposition Ratio

Material 1

Material 2

Library
To Do: Generate Crate of Research Objects
Research Priorities

Current:
• Ubiquitous use of persistent identifiers for all research objects
• Digital Object Metadata, Linked Data, and Semantic Web
• Digital Object Architecture Vision

Future:
• Operations on Digital Objects
• Versioning Digital Objects
• Data Type Registries
• PID Kernel Metadata
• Connection to RO-Crate

Anything Missing?
Thank You!