# Materials metadata:

# as a custom schema, as directories, or in a data package

Asahiko Matsuda, Kosuke Tanabe, Takuya Kadohira, and Mikiko Tanifuji Materials Data and Integrated System (MaDIS), National Institute for Materials Science (NIMS), Japan

## We are developing a materials data platform



A combination of systems to support every stage of data's lifecycle for materials science



Metadata for materials data

- "Bibliographic" metadata Title, Authors, Identifiers... "Scientific" metadata
  - Instruments, Specimen,





# 1: Metadata as a custom schema

"One JSON schema to rule them all"?



### 2: Metadata as directories

Researchers already manage their data using folders.



#### platform systems



X Difficulties in implementing a complex structure. X Massive edit forms  $\rightarrow$  Users were overwhelmed.



### 3: Metadata in a data package

Making full use of package metadata files.

Data packaging:

ro-crate

Making data self-documenting and self-contained. RO-Crate, BagIt, etc.

data.csv ro-crate-preview.html ro-crate-metadata.jsonld

 $\checkmark$  Assists wide distribution beyond our platform.

```
"@graph": [
{ "@id": "./", //metadata for whole
  /* Bibliographic metadata */
  "name": ...,
  "author": ...,
   /* Scientific metadata */
  "variableMeasured": ...,
```

"hasPart": {"@id": "data.csv"} ر { "@id": "data.csv", /\* Different metadata for parts of the dataset \*/

IoT-assisted data collection system. Mostly positive researcher feedback.

Thoughts

X Prone to human errors.

X Only simple common metadata.

X Different mapping for every research project.



 Bibliographic metadata are well standardized, while domain-specific scientific metadata require discussions in respective communities.

✓ Parts can have different metadata from the whole.

X Limited vocabulary for

domain-specific metadata. → Potential topic for community discussions.

(See Bioschemas, an extension of Schema.org)

 Dealing with heterogeneous data requires a common metadata schema, but it's not always a good idea to expose the whole to the users.

There's more than one way to deal with metadata. Look for a practical combination of methods.



National Institute for Materials Science (NIMS) Materials Data Platform Center (DPFC), Materials Data and Integrated System (MaDIS)

