Implementing workflows for FAIR data

Variables dictionary

Pawel Gancarski
CENER
WESE 2019, 3 October 2019
The title explained

CONTEXT
Context – FAIR data

• Findable
  – Can you find the data which you didn’t know that existed?

• Accessible
  – Once you know a dataset exists, can you access it?

• Interoperable
  – Are the files compatible with software you are using?

• Reusable
  – Can you make sense of the dataset if you were not involved in its creation?
Context – FAIR data

• Findable
  – Can you find the data which you didn’t know that existed?

• Accessible
  – Once you know a dataset exists, can you access it?

• Interoperable
  – Are the files compatible with software you are using?

• Reusable
  – Can you make sense of the dataset if you were not involved in its creation?
It is not only about H2020 projects!!!

FAIR data is a core aspect in the scalability of any digitalisation effort.
Solutions for data standardisation and documentation are not adopted if it implies additional work.

Provide tools*
and
Focus on workflows

*Tools should be useful.
Variables dictionary integration with the EOSC

EXAMPLE TOOL
Variables dictionary

Variables dictionary consist of .json files and a python class for handling it.
Variables dictionary – What is it for?

- Library for other tools
- Documentation
- Search engine
- A platform for community collaboration

```python
#fetch the data of a variable
metadata = var_dict.lookup('time')
```
Variables dictionary – What is it for?

- Library for other tools
- Documentation
- Search engine
- A platform for community collaboration

```json
"name": {
    "default": "time",
    "cf": "time",
    "open_oa": "",
    "iec_61400-25": "SecondSinceEpoch",
    "e-WindLidar": "",
    "grib": "",
    "other": [
        "t",
        "timeStamp"
    ]
},
"description": "iec_61400-25 defines time as a complex type consisting of two integers SecondSinceEpoch and
units": "s",
"ref": {
    "nvs": "http://vocab.nerc.ac.uk/collection/P07/current/CFSN0115/"
},
"netcdf": {
    "var_type": "double",
    "other": "units="seconds since 1970-01-01 00:00:00.00 UTC", calendar="gregorian" "
}
```
Variables dictionary – What is it for?

- Library for other tools
- Documentation
- Search engine
- A platform for community collaboration

```json
"name": {
  "default": "northward_wind",
  "cf": "northward_wind",
  "open_oa": "",
  "iec_61400-25": "",
  "e-WindLidar": "",
  "grib": "34 E132",
  "other": ["ws_y", "ws_v", "wind speed", "y_wind", "geostrophic_northward_wind"]
},
"description": "Northward indicates a vector component which is positive when directed northward",
"units": "m s-1",
"ref": {
  "nvs": "http://vocab.nerc.ac.uk/collection/P07/current/CFSN0461/"
},
"netcdf": {
  "var_type": "float",
  "other": ""
}
```
Variables dictionary – What is it for?

- Library for other tools
- Documentation
- Search engine
- A platform for community collaboration
CONCLUSIONS
Conclusions

• Data management is a bottleneck in the digitalisation efforts.
  – Data marketplace without machine readable licenses?
  – Machine learning and big data without machine readable timestamps?

• We should collaborate by:
  – being aware of each others workflows, and tools,
  – and building bridges between the tools.

• We should collaborate in order to:
  – reduce costs of research and data analysis,
  – do more cool science and less data formatting,
  – develop ideas such as data marketplace or big data analysis.
Useful links

• The variables dictionary

• Wind data converter

• Open source Jupyter notebooks
  – tools for compatible data, standardising data analysis, building trust for scientific results)
  – [https://github.com/CENER-EPR/OWAbench](https://github.com/CENER-EPR/OWAbench)

• Taxonomies for WE data
  – Data findability

• Data registry/publishing service
  – Will be operational November 2019
  – [https://sharewind.eu/](https://sharewind.eu/)