CENER

ADItech

Implementing workflows for FAIR data

Variables dictionary

Pawel Gancarski

CENER

WESE 2019, 3 October 2019

MaRINET2

Supported by

GOBIERNO DE ESPAÑA | MINISTERIO DE INDUSTRIA, ENERGÍA Y TURISMO | MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Ciemat

EERA European Energy Research Alliance

The title explained

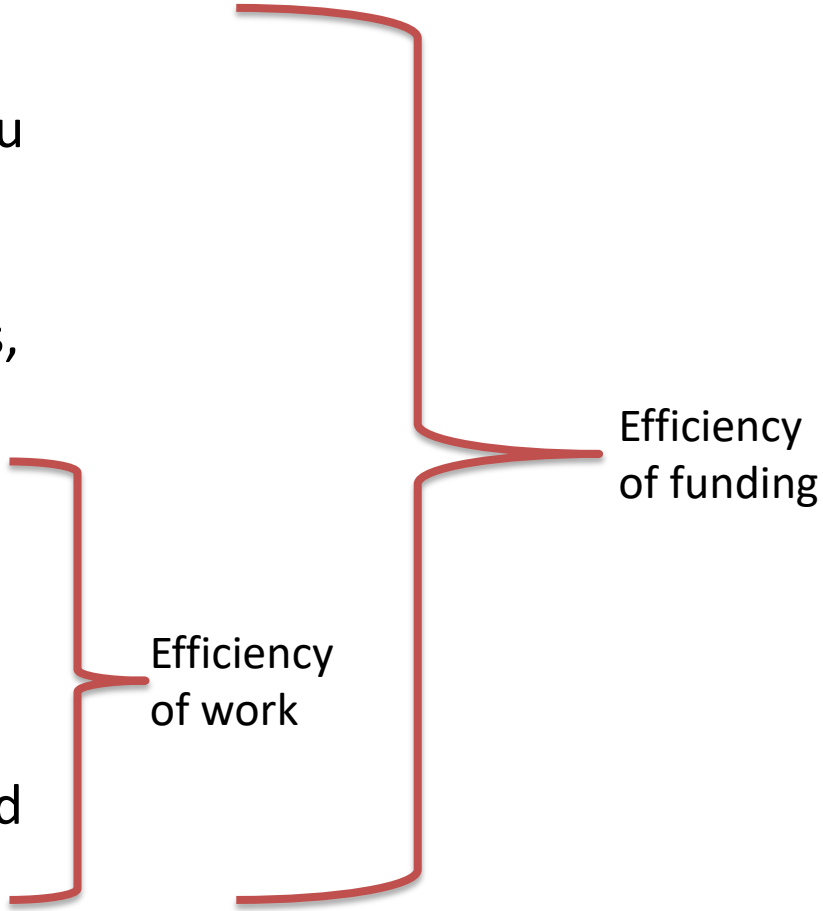
CONTEXT

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?

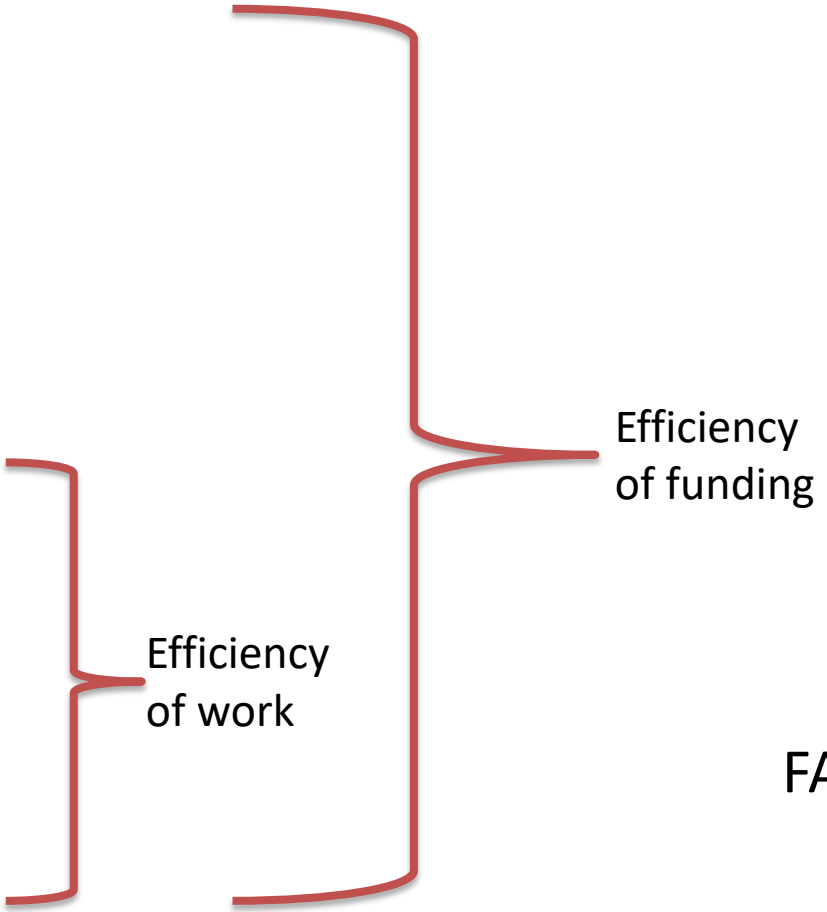
Context – FAIR data

- 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?



Context – FAIR data

- F
- A
- I
- R



It is not only about H2020 projects!!!

FAIR data is a core aspect in the scalability of any digitalisation effort.

Context - Workflows

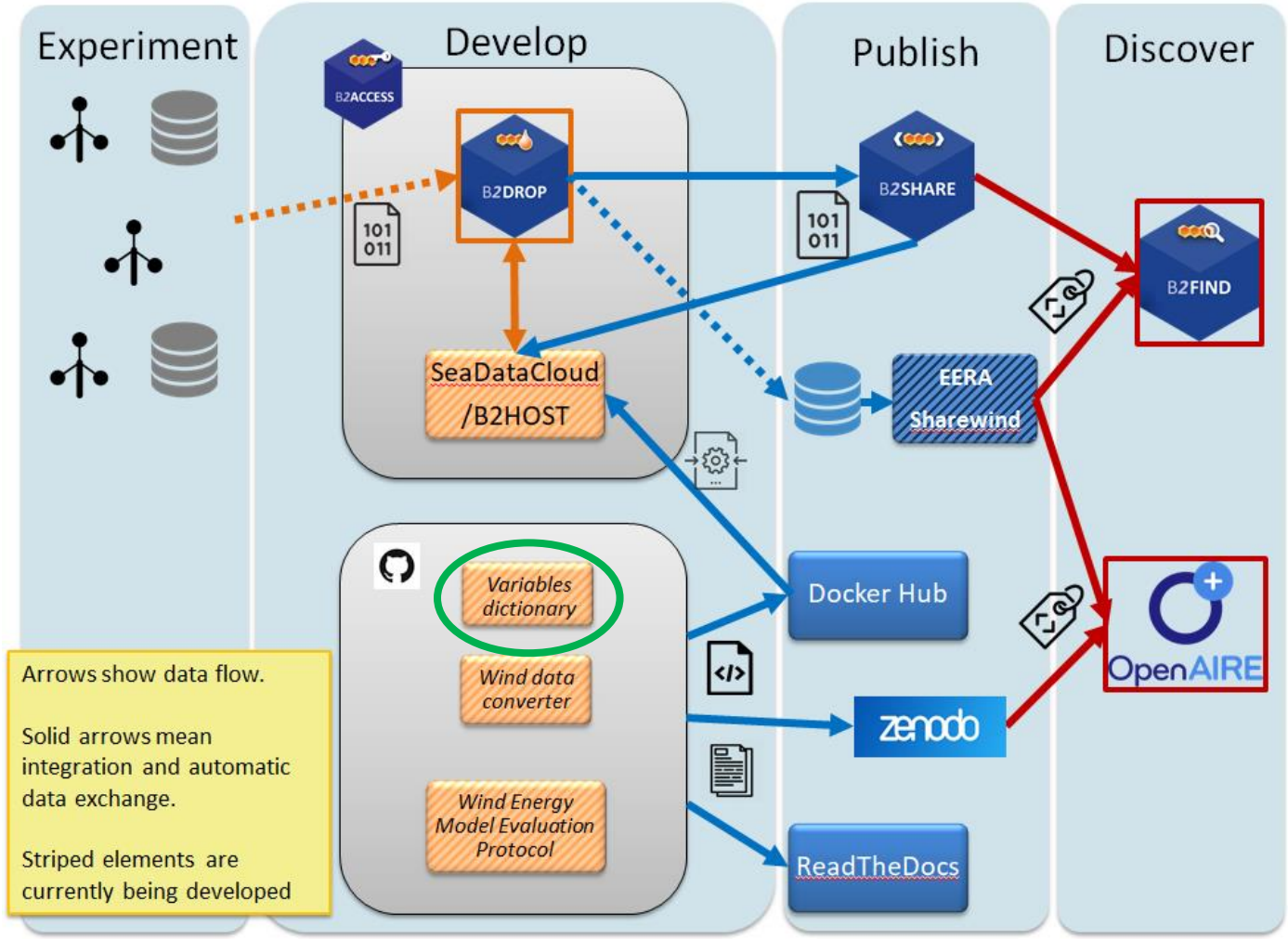
Solutions for data standardisation and documentation are not adopted if it implies additional work.

Provide tools*

and

Focus on **workflows**

*Tools should be useful.



Variables dictionary integration with the EOSC

EXAMPLE TOOL

Variables dictionary

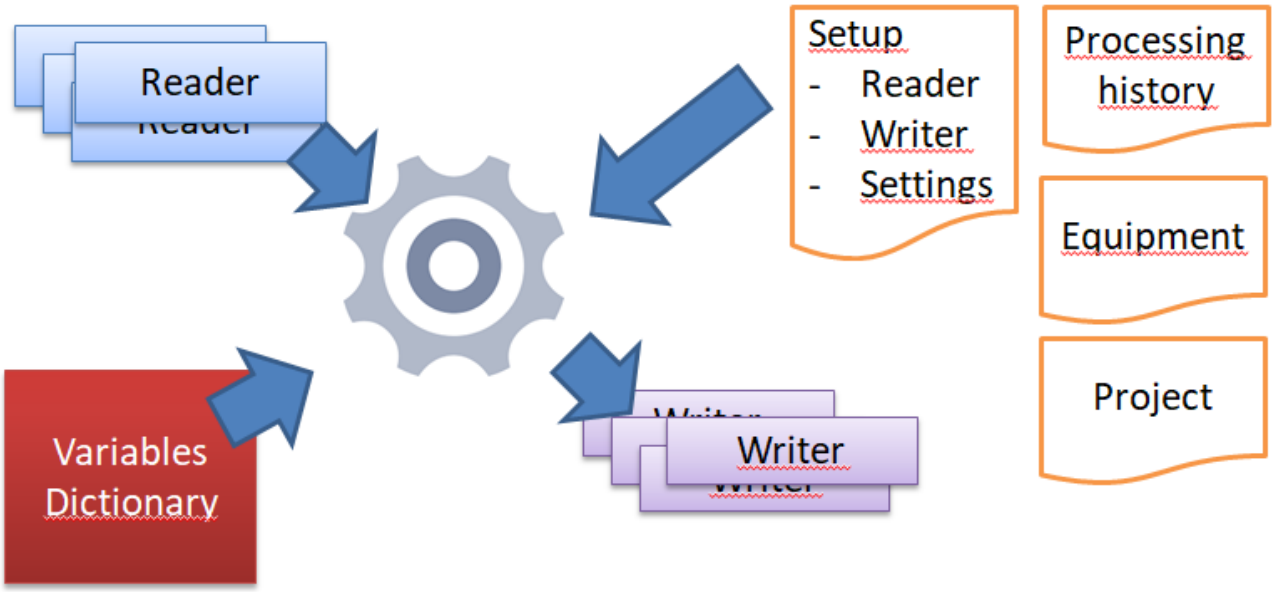
```
{
  "name": {
    "default": "wind_speed",
    "cf": "wind_speed",
    "open_oa": "windspeed_ms",
    "iec_61400-25": "MetAlt1HorWdSpd",
    "e-WindLidar": "",
    "grib": "31",
    "other": [
      "ws",
      "wind speed"
    ]
  },
  "description": "Speed is the magnitude of velocity. Wind is defined as a two-dimensional (horizontal) air velocity vector, w
  "units": "m s-1",
  "ref": {
    "nvs": "http://vocab.nerc.ac.uk/collection/P07/current/CFSN0038/"
  },
  "netcdf": {
    "var_type": "float",
    "other": ""
  }
},
```

Variables dictionary consist of .json files and a python class for handling it.

Variables dictionary – What is it for?

```
#fetch the data of a variable
metadata = var_dict.lookup('time')
```

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



Wind data converter (windaco) structure

Variables dictionary – What is it for?

```
184       "name": {
185         "default": "time",
186         "cf": "time",
187         "open_oa": "",
188         "iec_61400-25": "SecondSinceEpoch",
189         "e-WindLidar": "",
190         "grib": "",
191         "other": [
192           "t",
193           "timeStamp"
194         ]
195       },
196       "description": "iec_61400-25 defines time as a complex type consisting of two integers SecondSinceEpoch an
197       "units": "s",
198       "ref": {
199         "nvs": "http://vocab.nerc.ac.uk/collection/P07/current/CFSN0115/"
200       },
201       "netcdf": {
202         "var_type": "double",
203         "other": "units=\"seconds since 1970-01-01 00:00:00.00 UTC\", calendar=\"gregorian\" "
204       }
```

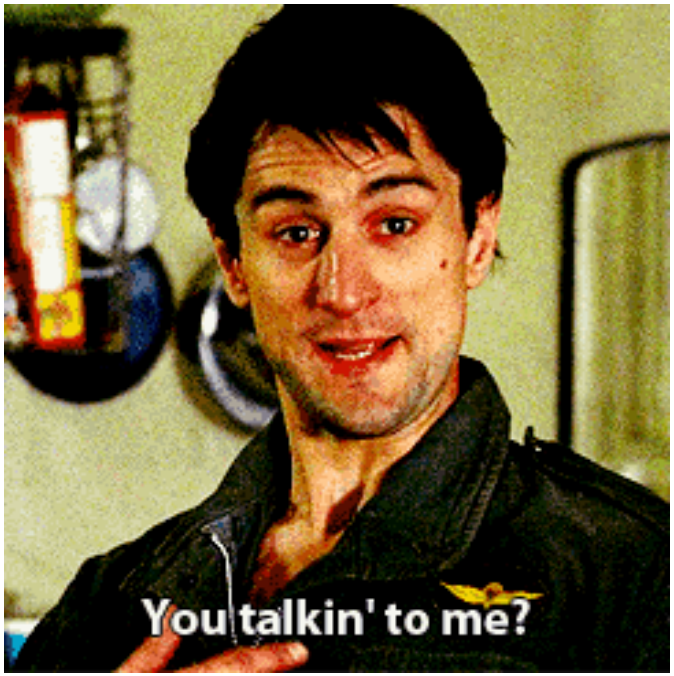
- Library for other tools
- Documentation
- Search engine
- A platform for community collaboration

Variables dictionary – What is it for?

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

```
24      "name": {
25        "default": "northward_wind",
26        "cf": "northward_wind",
27        "open_oa": "",
28        "iec_61400-25": "",
29        "e-WindLidar": "",
30        "grib": "34 E132",
31        "other": ["ws_y", "ws_v", "wind speed", "y_wind", "geostrophic_northward_wind"]
32      },
33      "description": "Northward indicates a vector component which is positive when directed northward
34      "units": "m s-1",
35      "ref": {
36        "nvs": "http://vocab.nerc.ac.uk/collection/P07/current/CFSN0461/"
37      },
38      "netcdf": {
39        "var_type": "float",
40        "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
  - https://github.com/wind-energy/variables-dictionary
- Wind data converter
  - https://github.com/wind-energy/Windaco
- Open source Jupyter notebooks
  - tools for compatible data, standardising data analysis, building trust for scientific results)
  - https://github.com/CENER-EPR/OWAbench
- Taxonomies for WE data
  - Data findability
  - https://github.com/wind-energy/taxonomies-and-vocabularies
- Data registry/publishing service
  - Will be operational November 2019
  - https://sharewind.eu/

MaRINET2

Supported by

GOBIERNO DE ESPAÑA
MINISTERIO DE INDUSTRIA, ENERGÍA Y TURISMO
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Ciemat

EERA
European Energy Research Alliance