

# Making My Data FAIR as a Data Steward

What could go wrong?



**Nami Sunami**  
Data Steward | TU Eindhoven  
[n.sunami@tue.nl](mailto:n.sunami@tue.nl)   



**12:30-12:40**  
**Let's have lunch**

# Join Open Science Retreat



**"I really liked the personal relations you build when attending; It's much easier now to reach out to others in the field."**

OPEN SCIENCE RETREAT  
THE NETHERLANDS

SUNDAY 29 MARCH – THURSDAY 02 APRIL 2026 / SCHOORL

OPEN  
SCIENCE  
RETREAT  
NL

Dutch Edition  
**29 March  
– 2 April 2026**



Global Edition (UK)  
**7–11 April 2026**



Presentation will start at 12:40

# Making My Data FAIR as a Data Steward

What could go wrong?



**Nami Sunami**  
Data Steward | TU Eindhoven  
[n.sunami@tue.nl](mailto:n.sunami@tue.nl)   



Previously  
I made a docker  
container for  
my thesis

The image shows a promotional poster for a meet-up. The background is orange at the top and white at the bottom. At the top left is the OSC/e logo (Open Science Community Eindhoven). At the top right is the MEET-UP logo for Eindhoven University of Technology. In the center, the title 'Writing a reproducible PhD thesis using GitHub, R, and Docker' is displayed in large white font. Below the title is a circular profile picture of a person with glasses. To the right of the profile picture, the name 'Nami Sunami' is written, followed by 'Data Steward | TU Eindhoven' and an email address 'n.sunami@tue.nl'. At the bottom left, the DOI '10.53962/x4b0-jtqj' is shown. At the bottom right, there are icons for a butterfly, a mail envelope, and LinkedIn.

OSC/e  
Open Science Community Eindhoven

MEET-UP  
Eindhoven University of Technology

**Writing a reproducible PhD thesis using GitHub, R, and Docker**

Nami Sunami  
Data Steward | TU Eindhoven  
n.sunami@tue.nl

DOI: 10.53962/x4b0-jtqj

# I shared regrets

## Regret 2

**Not storing data in  
a data repository**

Bloating the git  
repo size

Binary file & version  
control don't go well

Not FAIR

# Reminder



I am a  
**Data Steward**

**Data Steward advises researchers to  
make their data FAIR**

## **Regret 2**

**Not storing data in  
a data repository**

Bloating the git  
repo size

Binary file & version  
control don't go well

Not FAIR

# data in sitory

version  
go well

Not FAIR



I am a  
**Data Steward**

**My data is  
not FAIR**

**I want to fix this!**

**First, let's define terms**

# "Research Data"



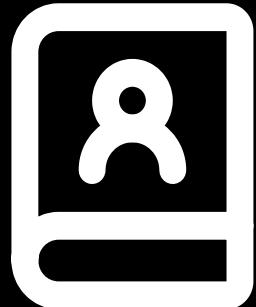
Data itself

<b>Nutrition Facts</b>	
4 servings per container	
<b>Serving size 1 1/2 cup (208g)</b>	
<hr/>	
<b>Amount per serving</b>	
<b>Calories</b>	<b>240</b>
<hr/>	
% Daily Value*	
<b>Total Fat</b> 4g	<b>5%</b>
Saturated Fat 1.5g	<b>8%</b>
Trans Fat 0g	
<b>Cholesterol</b> 5mg	<b>2%</b>
<b>Sodium</b> 430mg	<b>19%</b>
<b>Total Carbohydrate</b> 46g	<b>17%</b>

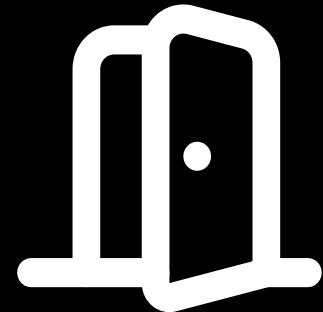
Metadata

# **What are FAIR data principles?**

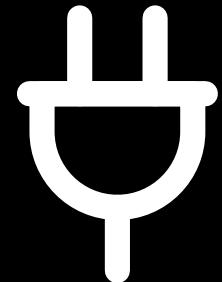
**Findable**



**Accessible**



**Interoperable**



**Reusable**



**Open Data**

**FAIR Data**

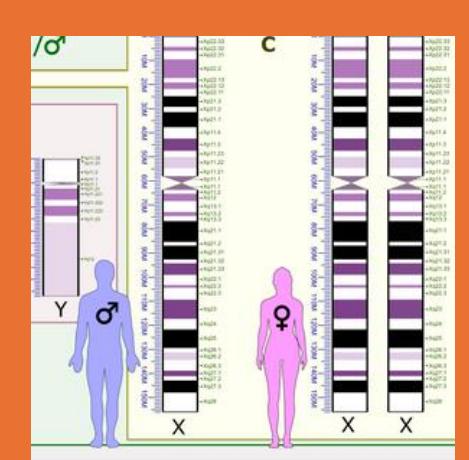
**Open**

kaggle



**FAIR & Open**

**Closed**

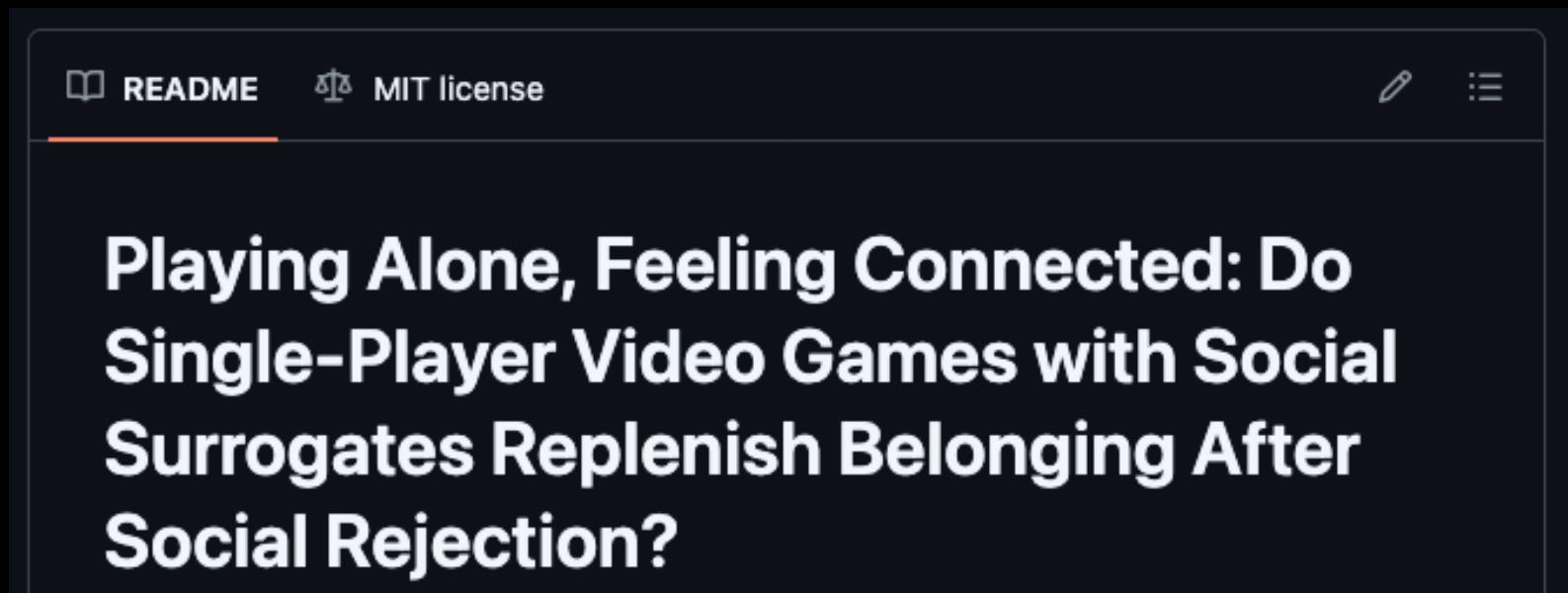


**Not FAIR**

**FAIR**

# **Previous state of my data**

# I had all my code and data in one Git repository



A journey of a thousand miles begins  
with a single step.

A journey to FAIR data begins with



A journey to FAIR data begins with  
posting data on a data repository.

**First goal: Post my data on Zenodo**



# What were my challenges?

- A Having multiple datasets
- B Not knowing metadata schema
- C Variables lacked documentation
- E All of the above

# What were my challenges?

- A Having multiple datasets
- B Not knowing metadata schema
- C Variables lacked documentation
- E All of the above

# Challenge 1

Having multiple datasets

Study 1

8 datasets

Study 2

1 dataset

Study 3

1 dataset

# Which dataset did I choose?

- A. Study 1 datasets
- B. Study 2 dataset
- C. Study 3 dataset
- D. Chose all datasets

# I chose Study 2

Because I thought it was  
the simplest

**Study 1**

**8 datasets**

**Study 2**

**1 dataset**

**Study 3**

**1 dataset**

## Challenge 2

**Not knowing metadata standard**

# Searching for metadata schema on resources

FAIRsharing Standards Search

DCC Metadata Standards List

RDA Metadata Standards Catalog

# I was lost in search

Many standards

Displaying 1 to 20 of 20.

Too specific for  
my project

CogMemo



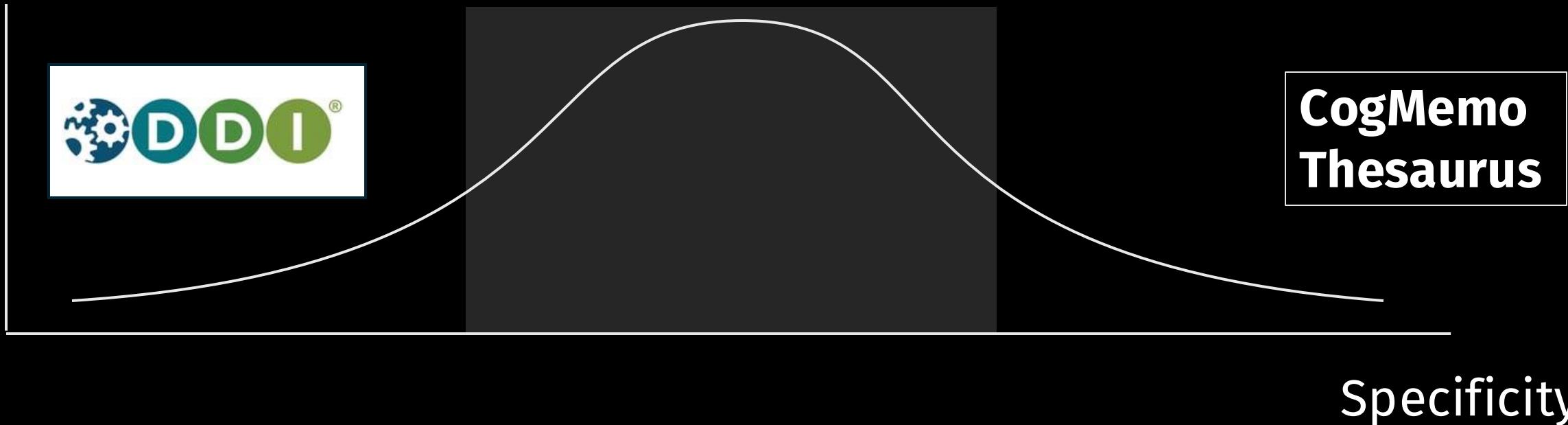
**Thesaurus Cognitive psychology of human memory**

The Thesaurus of the Cognitive Psychology of Human Memory, developed at Inist-CNRS, is a bilingual (French-English) terminological resource coveri...

Dataset    Psychology    Cognition    Homo sapi...

# Most standards were too generic or too specific

Usefulness for me



# What did I do next?

- A. I chose a generic standard
- B. I chose a specific standard
- C. I searched more standards
- D. I searched for happy memories

# What did I do next?

- A. I chose a generic standard
- B. I chose a specific standard
- C. I searched more standards
- D. I searched for happy memories

I remembered my time  
in Italy

Society of Improvement  
of Psychological Science  
(SIPS)  
2023 Conference



# I attended a hackathon for Psych-DS

[Hackathon / Landing Page](#)

rHA22: Can we get Psych-DS datasets indexed on Google Dataset Search? We think so!

---

**Session leader(s):** Melissa Kline Struhl

**Day/Time:** See [Live Schedule](#)

**Zoom Link:** [\[LINK\]](#)

**Abstract:** In this hackathon, we will evaluate a beta validator for Psych-DS by creating datasets that are ready to put their best foot forward! We will focus particularly on the metadata we provide - structured information about the dataset that can help people learn about the nitty-gritty details of your dataset, even if you cannot publicly share the data file itself.

<https://docs.google.com/document/d/1Vf0ylq9KwQhFWp2jKc9jBRjhUqmw5rqUGzD8/edit?tab=t.0>

# Psych-DS is a community data standard

## Project structure

```
my_study/
└── README.md
└── dataset_description.json
└── data/
    └── study-myStudy_data.csv
```

## Simple metadata

```
{  
  "@context": "http://schema.org/",  
  "@type": "Dataset",  
  "name": "Example Study",  
  "description": "It's an example.",  
  "variableMeasured": ["id", "height"]  
}
```

# Psych-DS provides a **step-by-step guide**

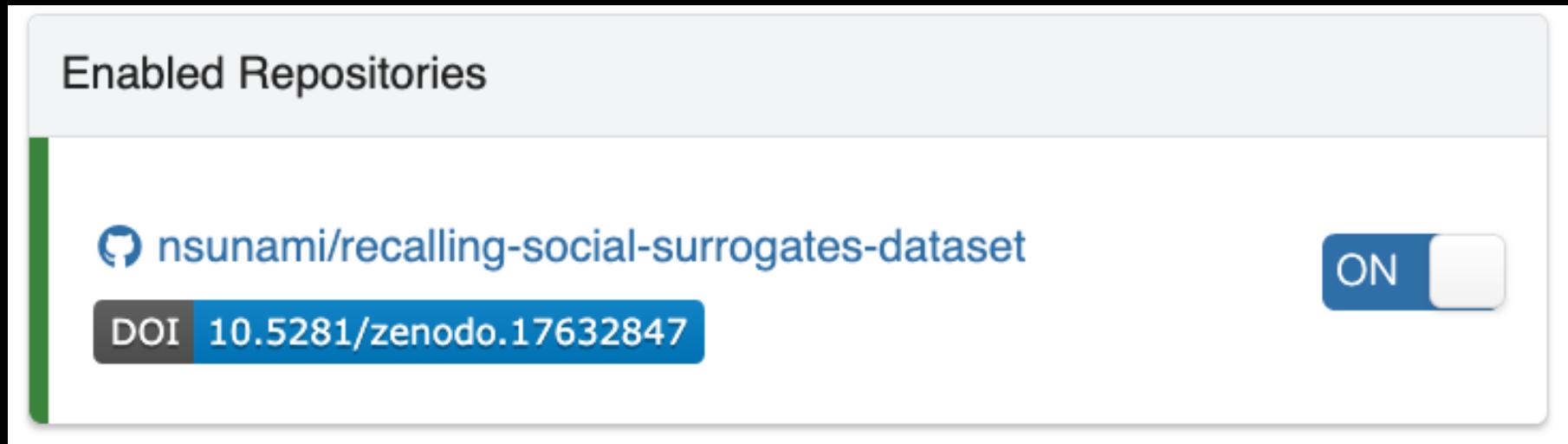
To use Psych-DS, format your data - we'll show you how

# I made a Git repository with Psych-DS metadata

The screenshot shows a GitHub repository page with the following details:

- README** tab is selected.
- CC-BY-4.0 license** is indicated.
- Recalling Social Surrogates Dataset** is the title of the repository.
- Metrics: FAIR assessment (37.50 %), F-UJI (48%), DOI (10.5281/zenodo.17593670).
- Description: The dataset for Study 2 of Nami Sunami's dissertation, "Playing Alone, Feeling Connected: Do Single-Player Video Games with Social Surrogates Replenish Belonging After Social Rejection?".
- Methodology: See the [method section of Study 2](#) for more information about the origin of the data collection.
- Metadata Standards**: I followed [Psych-DS data standard](#).
- License**: This work is licensed under CC BY 4.0.

# I published the repository on Zenodo



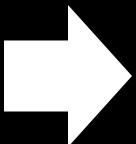
# The metadata about the dataset was looking ok

```
{  
  "@context": "http://schema.org/",  
  "@type": "Dataset",  
  "name": "...",  
  "description": "...",  
  "author": {  
    // Info about Nami  
  },  
  "sameAs": "https://doi.org/10.5281/  
zenodo.17593670",  
  "license": "https://creativecommons.org/  
licenses/by/4.0/",  
  "variableMeasured": [  
    ...  
    "A",  
    "B",  
    "C",  
    ...  
  ]  
}
```

# But how about metadata about variables



```
"variableMeasured": [  
    "reaction_time",  
    "condition"  
]
```



```
"variableMeasured": [  
    {  
        "@type": "PropertyValue",  
        "name": "reaction_time",  
        "description": "Reaction time",  
        "unitText": "milliseconds",  
        "minValue": 0,  
        "maxValue": 2000  
    },  
    {  
        "@type": "PropertyValue",  
        "name": "condition",  
        "description": "Experimental condition",  
    }  
]
```

**So, I looked into my variable-level  
metadata**

# Challenge 3

WTF are these variables?

My variables lacked documentation

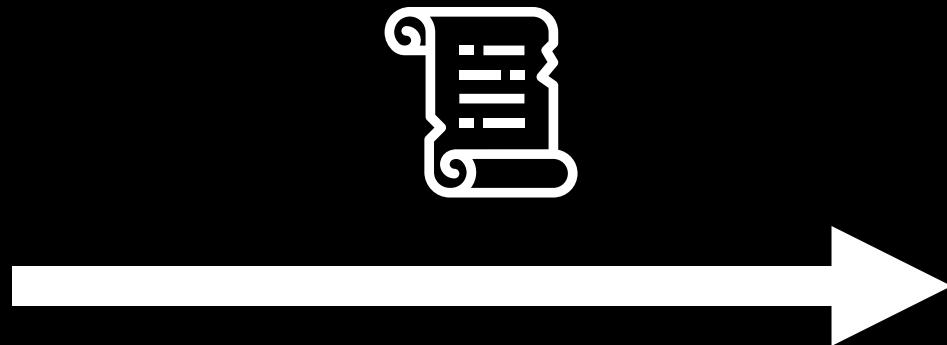
A	B	C
Yes	Yes	Yes
Yes	Yes	No
Yes	Yes	Yes

**Luckily, I embedded variable  
description in my original dataset**

I wrote a **script** to export the variable labels into Psych-DS metadata



My data



Script to export  
variable labels

Updated  
metadata

# I also manually updated variable metadata

1038            {	1056            {
1039                "@type": "PropertyValue",	1057                "@type": "PropertyValue",
1040                "name": "A",	1058                "name": "A",
	1059                +                "description": "Rater A indicating that the participant followed the instructions",
1041                "maxValue": "3",	1060                "maxValue": "3",
1042                "minValue": "2"	1061                "minValue": "2"
1043                },	1062                },

**It was a lot of work**

**Why am I doing this?**

**Do any of these matter?**

**Watch out for FAIR Data Burnout!**



There are tools to **score** FAIR-ness  
of a dataset.

**FAIR-Checker**

**F-UJI**

# Zenodo entry has higher scores

Checker used	<u>GitHub</u>	<u>Zenodo</u>
FAIR-Checker	37.50%	91.67%
F-UJI	42%	84%

Are the scores useful?

# Generic metadata

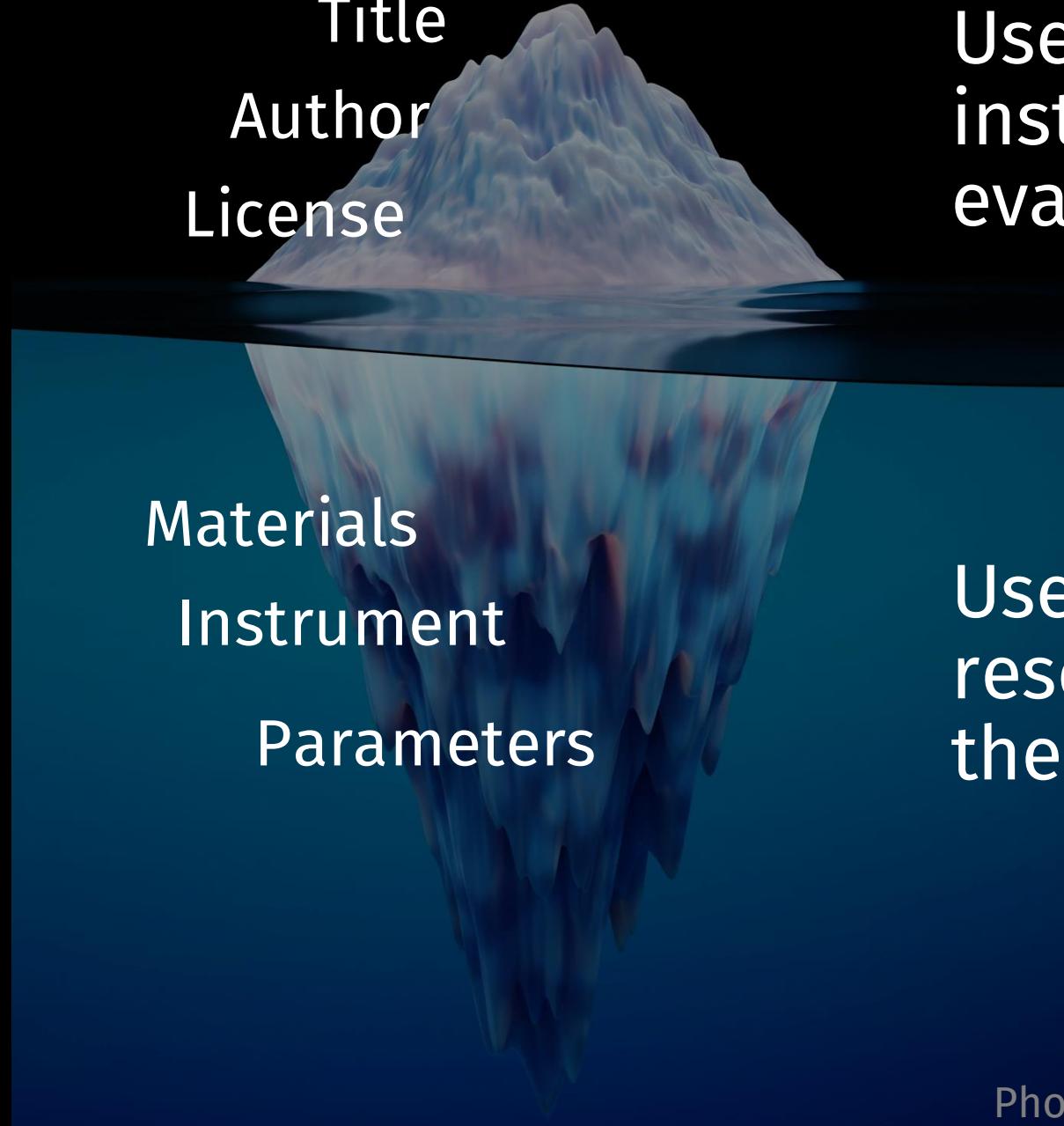
Title  
Author  
License

Useful for  
institutions and  
evaluators

# Specific metadata

Materials  
Instrument  
Parameters

Useful for  
researchers in  
the field





**Why are we making data FAIR?**

We make data FAIR, so that someone  
else can re-use the data.

**Who is the that someone else?**

**A future data user can be me.**

**If I were to do it again, I'd try to make a  
link with another dataset.**

So that I can produce new piece of information, for myself to enjoy

**What are your questions?**

# Questions

What are your challenges?

What are your experiences making  
your data FAIR?

# Join Our Community



✉ [openscience@tue.nl](mailto:openscience@tue.nl)

🌐 [sites.google.com/view/osceindhoven](https://sites.google.com/view/osceindhoven)



# Links

[GitHub | dataset](#)

[GitHub | Nami's dissertation](#)