

Making My Data FAIR as a Data Steward

What could go wrong?



Nami Sunami

Data Steward | TU Eindhoven

n.sunami@tue.nl





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

Join Open Science Retreat



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



Previously

I made a docker container for my thesis



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

🦋 @ in

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 pository

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"



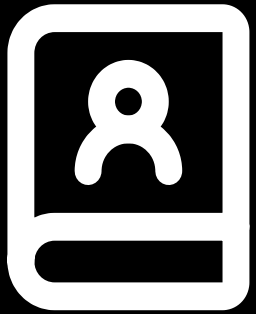
Data itself

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

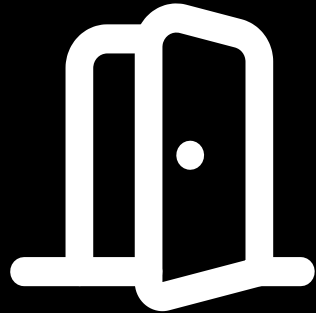
Metadata

What are FAIR data principles?

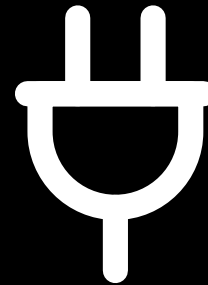
Findable



Accessible



Interoperable



Reusable



Open Data

FAIR Data

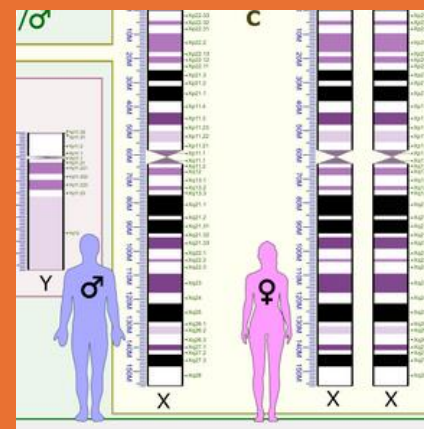
Open

kaggle



FAIR & Open

Closed

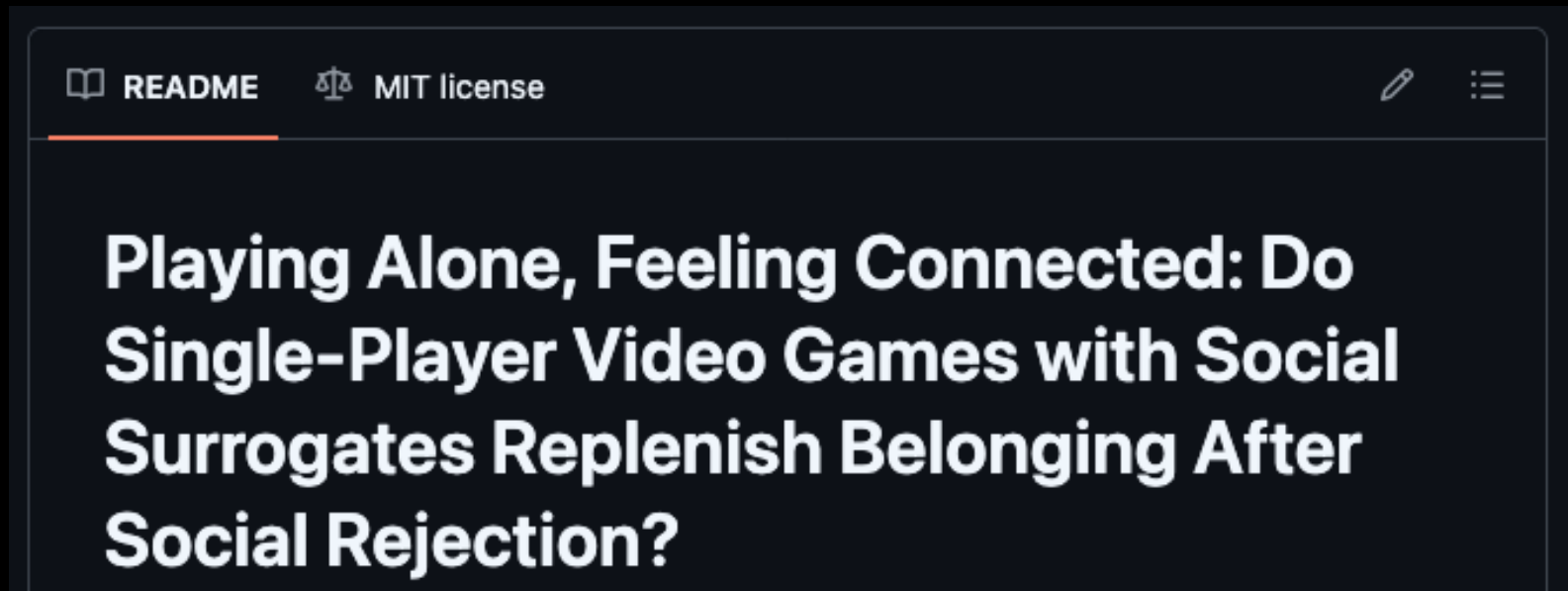


Not FAIR

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

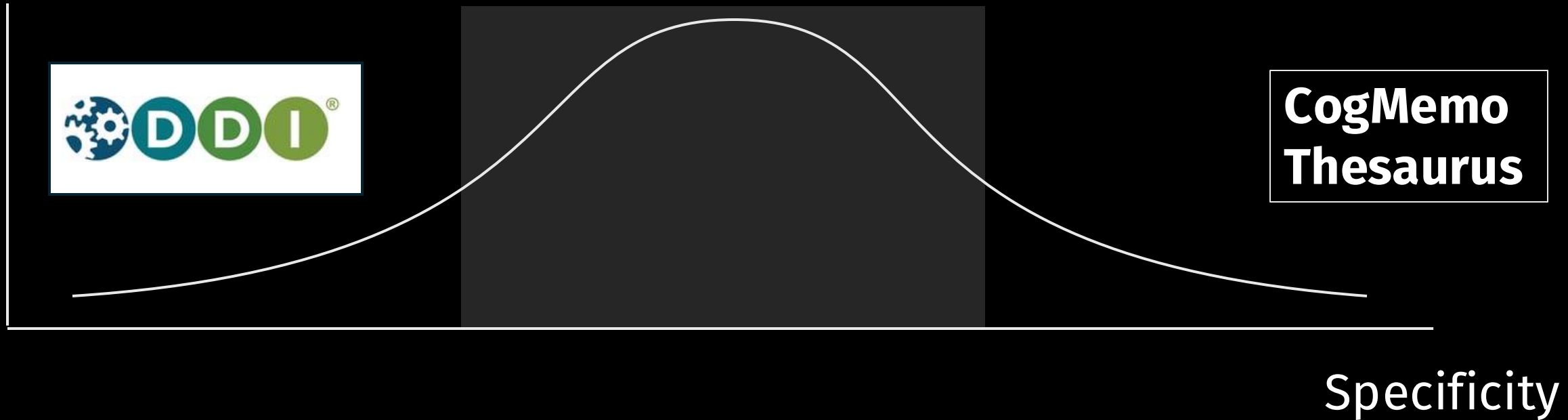


The Thesaurus of the Cognitive Psychology of Human Memory, developed at Inist-CNRS, is a bilingual (French-English) terminological resource coveri...

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/1Vf0ylq9KwQhFWp2jKc9jBRjhu5Gds2HUqmw5rqUGzD8/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

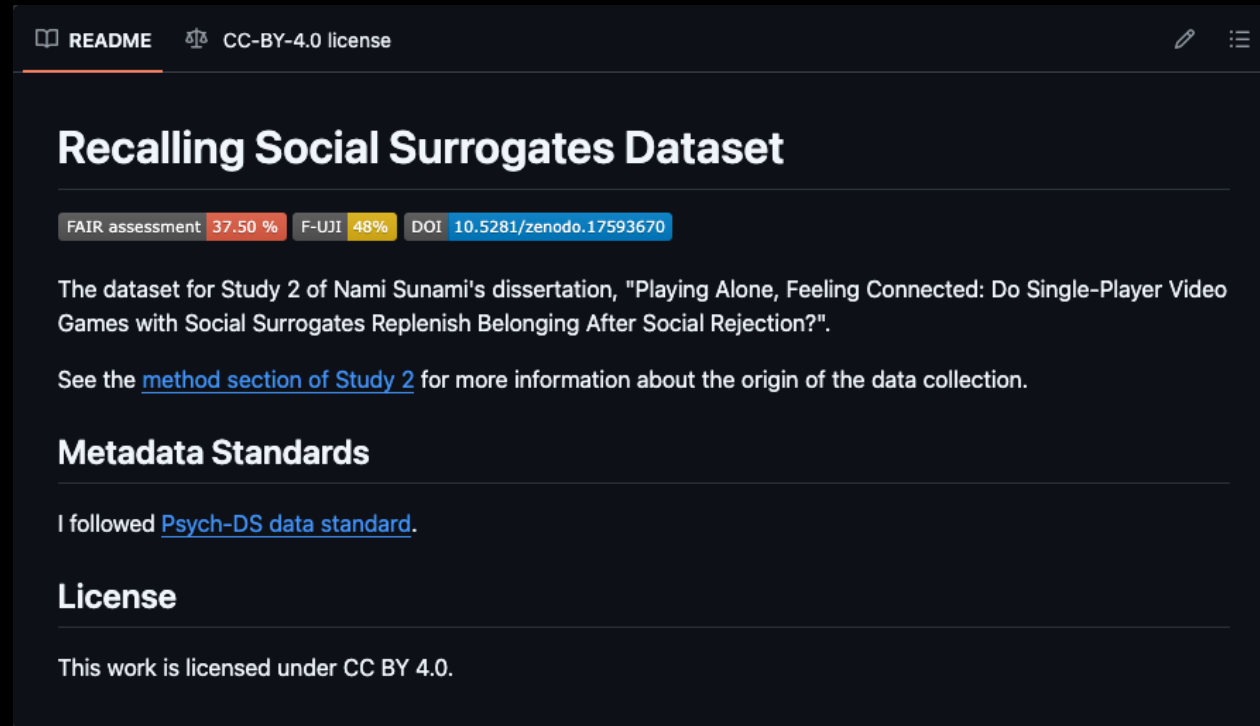


```
{  
  "@context": "http://schema.org/",  
  "@type": "Dataset",  
  "name": "Example Study",  
  "description": "It's an example.",  
  "variableMeasured": ["id", "height"]  
}
```

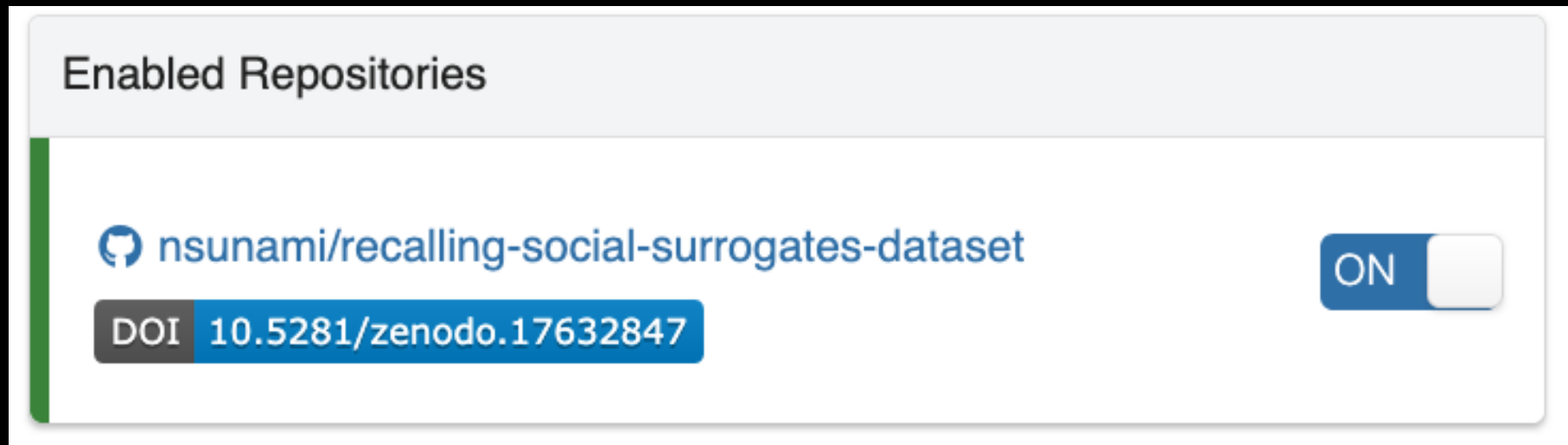
Psych-DS provides a **step-by-step** guide

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

I made a Git repository with Psych-DS metadata



I published the repository on Zenodo

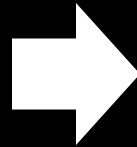


**The metadata about
the dataset was
looking ok**

```
{
  "@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

**My variables
lacked
documentation**

**WTF are these
variables?**

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



I also manually updated variable metadata

```
1038      {
1039          "@type": "PropertyValue",
1040          "name": "A",
```

```
1041          "maxValue": "3",
1042          "minValue": "2"
1043      },
```

```
1056      {
1057          "@type": "PropertyValue",
1058          "name": "A",
```

```
1059      +      "description": "Rater A
            indicating that the participant
            followed the instructions",
```

```
1060          "maxValue": "3",
1061          "minValue": "2"
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

🌐 sites.google.com/view/osceindhoven



Links

[GitHub | dataset](#)

[GitHub | Nami's dissertation](#)