

# Training exercise – Evaluating a dataset using FAIR-Aware

(10.5281/zenodo.8089501)

This exercise was created by Maaïke Verburg  (DANS) – November 2021, and is available under a [CC by 4.0 licence](#)

FAIR-Aware was created by DANS in the FAIRsFAIR project<sup>1</sup> and is currently further developed in the FAIR-IMPACT project<sup>2</sup>.

## Evaluating a dataset using FAIR-Aware

### Goal of the exercise

What does it mean to put FAIR into practice? How can FAIR knowledge be translated into FAIR skills and how can this be applied to a dataset? This is what the FAIR-Aware tool aims to help you with.

The goal of this exercise is to reflect on:

- 1) What does the implementation of FAIR practices look like in a deposited dataset?
- 2) How does the implementation of FAIR practices aid reuse of the data?
- 3) How easy or difficult is it for humans to evaluate the FAIRness of a dataset?

With your group, evaluate the assigned dataset based on the **10 FAIR practices** detailed in the FAIR-Aware tool.

Use the **guidance texts** in the FAIR-Aware tool to discover how FAIR practices can be implemented (see the “How to do this?” sections). Use the dataset and the repository information to determine whether these practices are satisfied.

Gather **evidence**: How can you tell that this FAIR practice is or isn’t satisfied?

---

<sup>1</sup> FAIRsFAIR “Fostering FAIR Data Practices In Europe” has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

<sup>2</sup> FAIR-IMPACT “Expanding FAIR solutions across EOSC” is funded by the European Union.

Discuss **shortcomings**: How could the dataset be improved on this FAIR practice?

Discuss **importance**: How does this FAIR practice facilitate reuse or other important aspects of FAIR? Taking the perspective of a potential reuser of the data, how does the implementation of the FAIR practice help you in making your choice.

FAIR-Aware: <https://fairaware.dans.knaw.nl/>

Group #:

Dataset:

Repository policy:

FAIR Practice	Applied to data?	Evidence:	Room for improvement:	Comments:
1. Persistent identifier	Yes / No			
2. Discovery metadata	Yes / No			
3. Metadata for humans and machines	Yes / No			
4. Access control metadata	Yes / No			

5. Persistence of metadata	Yes / No			
6. Controlled vocabularies	Yes / No			
7. Provenance information	Yes / No			
8. Community-endorsed metadata standards	Yes / No			
9. Preferred file formats	Yes / No			
10. Digital curation and preservation	Yes / No			

## Reflection / highlights

After completing the exercise, please share your reflections and other takeaways from this exercise.

1. What does the implementation of FAIR practices look like in a deposited dataset?
2. How does the implementation of FAIR practices aid reuse of the data?
3. How easy or difficult is it for humans to evaluate the FAIRness of a dataset?

NB: People assessing their own data will have variable results depending on their own choices and practices.

## Worked examples of the datasets, #1

**Article:** Reukers, Daphne F. M., Cornelia H. M. van Jaarsveld, Reinier P. Akkermans, Stephan P. Keijmel, Gabriella Morroy, Adriana S. G. van Dam, Peter C. Wever, et al. 2022. ‘Impact of Q-Fever on Physical and Psychosocial Functioning until 8 Years after Coxiella Burnetii Infection: An Integrative Data Analysis’ . *PLOS ONE* 17 (2): e0263239.

<https://doi.org/10.1371/journal.pone.0263239>.

**Dataset:** Jaarsveld, C.H.M. Van, D.F.M. Reukers, R.P. Akkermans, S.P. Keijmel, G. Morroy, A.S.G. Van Dam, P.C. Wever, et al. 2021. ‘Impact of Q-Fever on Physical and Psychosocial Functioning until 8 Years after Coxiella Burnetii Infection’ .

Application/pdf.csv. Data Archiving and Networked Services (DANS). <https://doi.org/10.17026/DANS-ZPA-FKPH>.

NB: This table is filled in for the **dataset**, not the article that is linked to the dataset.

FAIR Practice	Applied to data?	Evidence:	Room for improvement:	Comments:
1. Persistent identifier	<u>Yes</u> / No	DOI is linked  ‘Identifier’ indicates that other PIDs are also linked	The Dutch Digital Author Identifier (DAI) is also used for some authors but others do not have a PID (such as ORCID) nor does the data collector, in this case Radboud (ROR)	
2. Discovery metadata	<u>Yes</u> / No	Creator, title, date submitted, description, subject keywords PID	The dataset was deposited before the article publication, so it wasn’ t possible to link	Sharing options for social media to increase online presence

FAIR Practice	Applied to data?	Evidence:	Room for improvement:	Comments:
		<p>Data content is described</p> <p>Access rights are included</p>	<p>in ‘Relations’</p> <p>More keywords e.g., about content and type of data collected may be helpful</p>	
3. Metadata for humans and machines	<u>Yes</u> / No	DANS has OAI-PMH protocol in place		This information needs to be found on the repository level, not dataset level
4. Access control metadata	<u>Yes</u> / No	Unrestricted access, includes codebook and documentation		
5. Persistence of metadata	Yes / No	<p>Not applicable since data is still available.</p> <p>In Data Station policy it is stated that ‘DANS will always provide open access to all published metadata’</p>	Explicitly state in preservation plan what the standard data retention period is for the archive	

FAIR Practice	Applied to data?	Evidence:	Room for improvement:	Comments:
6. Controlled vocabularies	Yes / No	Since the Data Station Life Sciences is still being developed at this time, there is no overview of what vocabularies are supported and used yet	Explicit statement about which vocabulary was used for this dataset would help potential re-users	Hard to tell to what extent the (meta)data adheres to the standard
7. Provenance information	<u>Yes</u> / No	Temporal & Spatial coverage and submission dates  Includes contributors  Data generation is described in the dataset description  Codebook and documentation	More information on methods, instruments, protocols would enrich the provenance information	
8. Community-endorsed metadata standards	<u>Yes</u> / No	In DANS preservation policy it is detailed		Documentation on Data Station Life Sciences is not yet



FAIR Practice	Applied to data?	Evidence:	Room for improvement:	Comments:
		that the Dublin Core standard is used, as well as that this information is also mapped to other standards		available, so this is based on what it was like in the previous system: EASY
9. Preferred data format	<u>Yes</u> / No	DANS has an overview of preferred data formats.	For a codebook and documentation, pdf doesn' t seem the most reusable format	It' s hard to tell whether the pdf files are pdf/A
10. Digital curation and preservation	<u>Yes</u> / No	DANS is a CTS certified repository The Data Station Life Sciences is discipline-specific, tailored to the communities specific needs	Since the data station is quite new, not all information and documentation is clearly findable yet at this time	



## Worked examples of the datasets, #2

**Publications:** Kalkdijk-Dijkstra, A.J., J.A.G. van der Heijden, H.L. van Westreenen, P.M.A. Broens, M. Trzpis, J.P.E.N. Pierie, B.R. Klarenbeek, et al. 2020. ‘Pelvic Floor Rehabilitation to Improve Functional Outcome and Quality of Life after Surgery for Rectal Cancer: Study Protocol for a Randomized Controlled Trial (FORCE Trial)’ . *Trials* 21 (1): 112.

<https://doi.org/10.1186/s13063-019-4043-7>.

&

Heijden, J. a. G. van der, A. J. Kalkdijk-Dijkstra, J. P. E. N. Pierie, H. L. van Westreenen, P. M. A. Broens, B. R. Klarenbeek, and On behalf of the FORCE trial Group. 2022. ‘Pelvic Floor Rehabilitation After Rectal Cancer Surgery: A Multicenter Randomized Clinical Trial (FORCE Trial)’ . *Annals of Surgery* 276 (1): 38.

<https://doi.org/10.1097/SLA.0000000000005353>.

**Dataset:** Heijden, BSc J.A.G. van der (Radboud University); Klarenbeek, dr. B.R. (Radboud University); Vries, dr. M. de (Radboud University); Broens, P.M.; Westreenen, H.L. van; Kalkdijk-Dijkstra, A.J. (Radboud University); Pierie, J.P.E.N. (2022): *Pelvic Floor rehabilitation to improve functional Outcome and quality of life after surgery for Rectal CancEr: a randomized controlled trial. FORCE-trial*. DANS. <https://doi.org/10.17026/dans-zhe-rrv2>.

NB: This table is filled in for the **dataset**, not the article that is linked to the dataset.

FAIR Practice	Applied to data?	Evidence:	Room for improvement:	Comments:
1. Persistent identifier	<u>Yes</u> / No	DOI is linked	The Dutch Digital Author Identifier (DAI) is also used for some authors but others do not have a PID	

FAIR Practice	Applied to data?	Evidence:	Room for improvement:	Comments:
			(such as ORCID) nor does the data collector, in this case Radboud (ROR)	
2. Discovery metadata	<u>Yes</u> / No	<p>Creator, title, date submitted, description, subject keywords PID</p> <p>Data content is described</p> <p>Access rights are included</p>	Publication that is linked to this dataset in 'Relation' is the trial summary, so explains the collection of the data, however, the article published 2 years later, that uses the data in this dataset, does not link to the dataset at all	
3. Metadata for humans and machines	<u>Yes</u> / No	DANS has OAI-PMH protocol in place		
4. Access control metadata	Yes / No	Restricted access, but read_me as metadata		Access is via logging in and then requesting access via account

FAIR Practice	Applied to data?	Evidence:	Room for improvement:	Comments:
5. Persistence of metadata	Yes / No	Not applicable to dataset that is still available.		
6. Controlled vocabularies	Yes / No	Since the Data Station Life Sciences is still being developed at this time, there is no overview of what vocabularies are supported and used yet	Explicit statement about which vocabulary was used for this dataset would help potential re-users	Hard to tell to what extent the (meta)data adheres to the standard
7. Provenance information	<u>Yes</u> / No	Temporal & Spatial coverage and submission dates  Includes contributors  Data generation is described in the dataset description  Readme	The readme file is not machine-readable/no metadata standard mentioned	

FAIR Practice	Applied to data?	Evidence:	Room for improvement:	Comments:
8. Community-endorsed metadata standards	<u>Yes</u> / No	In DANS preservation policy it is detailed that the Dublin Core standard is used, as well as that this information is also mapped to other standards		Documentation on Data Station Life Sciences is not yet available, so this is based on what it was like in the previous system: EASY
9. Preferred data format	<u>Yes</u> / No	DANS has an overview of preferred data formats, but accepts that some fields use various file formats.	For a readme, pdf doesn' t seem the most reusable format	It' s hard to tell whether the pdf files are pdf/A
10. Digital curation and preservation	<u>Yes</u> / No	DANS is a CTS certified repository The Data Station Life Sciences is discipline-specific, tailored to the communities specific needs	Since the data station is quite new, not all information and documentation is clearly findable yet at this time	

