D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR

Author(s)	Josefine Nordling, Hannah Mihai, Bert Meerman, Tuomas Alaterä, Mari Kleemola, Ilja Livenson
Status	Final
Version	1.0
Date	2022-06-27

Document identifier:	
Deliverable lead	Josefine Nordling
Related work package	WP4
Author(s)	Josefine Nordling, Hannah Mihai, Bert Meerman, Tuomas Alaterä, Mari Kleemola, Ilja Livenson
Contributor(s)	WP4 team
Due date	2022-06-30
Actual submission date	2022-MM-DD
Reviewed by	Damien Lecarpentier, Adil Hasan
Approved by	
Dissemination level	Public
Website	www.eosc-nordic.eu
Call	H2020-INFRAEOSC-2018-3
Project Number	857652
Start date of Project	2019-09-01
Duration	36 Months

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

License	Creative Commons CC-BY 4.0
Keywords	FAIR, assessment, data repositories, open data, machine-actionable metadata, project outputs

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Abstract

The EOSC-Nordic project has pledged to implement FAIR in the Nordic and Baltic regions and aims to encourage, support and assist the research community to FAIRify their data. This will be achieved by communicating the benefits of going FAIR to a broad scientific community. The project has selected a hundred repositories and evaluated them consistently according to their FAIR maturity.

This deliverable describes the continued assessment of FAIR maturity for Nordic and Baltic scientific digital repositories. The text highlights the results of the FAIR assessment exercise using the F-UJI tool, however, in contrast to the previous deliverable (D4.1) the focus is in this deliverable on raising FAIR awareness by engaging the community, rather than focusing on the numeric results of the assessment and changes seen over time.

The deliverable first sets out the background and methodology of the FAIR assessment study, including a notion of the change of assessment tools. This is followed by a section of the results of the FAIR assessments and the community outreach activities made within the project in an attempt to raise further awareness of FAIR among data repository representatives. Lastly, the document raises attention to the ongoing discussions around contemporary FAIR assessments within the EOSC.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Table of contents

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

1. Introduction

This is the second EOSC-Nordic report on FAIR assessment, following the deliverable D4.1 An assessment of FAIR-uptake among regional digital repositories[1] published in August 2020. The main conclusion of that deliverable was that a majority of the repositories scored relatively low in the FAIR assessment of their metadata (33% passed 7 or less tests out of a total of 22 tests). In these cases, the repositories did provide GUIDs[2], but no other support for machine-actionability in the metadata was detected. A very low number of the repositories assessed scored high (54-77% passed 12-17 tests out of 22). In these repositories, machine-actionability was built into the metadata and data, thus supporting openness and accessibility of data. However, scores measured in the beginning of the project were unreliable due to changes in the FAIR assessment software making it very difficult to evaluate increases in FAIR uptake among the repositories. This was later corrected by using the exact same version of the assessment tool for executing the evaluations.

In this report we will give an overview of the work on assessing the FAIR maturity levels in approximately 100 data repositories in the Nordic and Baltic countries. For the assessment the F-UJI[3] tool was used. F-UJI is a web service to programmatically assess FAIRness of research data objects based on metrics developed by the FAIRsFAIR[4] project. Based on our experience during the first deliverable D4.1, we have found putting community engagement at the centre and reducing focus on the comparison of FAIR scores over time, to be more beneficial to the community. Engaging directly with the research communities and presenting ways in which they can make progress on different aspects essential for the FAIR principles, will over time leverage a change in FAIR research culture among data repositories we want to see in the Nordic and Baltic region. However, cultural changes take time. In this report, we chose to present the overall FAIR scores, as there has only been a slight positive trend over time. We also highlight stories describing how repositories have increased their FAIR scores, which should be helpful for repositories in similar situations. Due to these changed perceptions, the FAIR assessment task force wants to emphasise that by FAIR assessment we rather want to convey the message of providing FAIR assistance.

1.1 Background and objectives

The overall objective of FAIR is to enable research data to be more reusable and science more transparent, efficient, and trustworthy. Information on all the 15 FAIR Principles are described in

[1] Andreas Ortmann Jaunsen, Mari Kleemola, Tuomas J. Alaterä, Heikki Lehvaslaiho, Adil Hasan, Josefine Nordling, & Pauli Assinen. (2020). D4.1 An assessment of FAIR-uptake among regional digital repositories (1.0). Zenodo. https://doi.org/10.5281/zenodo.4045402

[2] GUID stands for Globally Unique Identifier that follows a specific structure defined in RFC 4122 (http://www.ietf.org/rfc/rfc4122.txt)

[3] Anusuriya Devaraju, & Robert Huber. (2020). F-UJI - An Automated FAIR Data Assessment Tool (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.4063720

[4] https://www.fairsfair.eu/

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

https://www.go-fair.org/fair-principles/. The goal of WP4 is to engage with Nordic and Baltic repositories and support them in becoming more compliant with the FAIR principles. A study by Forbes, made in 2016, showed that data scientists spend 80% of their time on data preparation and 57% of them consider data cleaning and organising data to be the part of their work they enjoy doing the least[5]. A more recent study made by Anaconda in 2020[6] showed that 45% of a data scientist's time is spent on data preparation (loading and cleaning data). The drastic decrease in the time spent on data preparation is truly remarkable, but there is still a lot of room for improvement, as data preparation still forms a dominant part of a scientist's work and thus takes up valuable time that could be put into scientific work. To partially address this issue and the mis-alignment between what is expected of the researcher in terms of data management and analysis and the amount of time and resources it actually requires to comply, there is a clear need to facilitate and automate these processes as much as possible. Making sure that the repositories are supporting FAIR practices, so called FAIR-enabling, is one step closer to achieving a research culture where FAIR compliance is made easier for the researcher and becomes part of the everyday work routine.

2. Methodology of the study

The very first step of this task was to define the criteria of the digital scientific repositories to be included in the sample of FAIR assessments. The repositories had to be hosted by a Nordic or Baltic country and store data relevant for research. Repositories mostly hosting publications, documents, preprints etc. were excluded from the sample. Additionally, the repository datasets had to be connected to GUIDs and machine-readable metadata to allow machine-actionable evaluations to take place. The landscaping exercise to locate all relevant data repositories in the region fulfilling all above criteria were performed in a two-phased approach; first using the re3data.org[7] as a source, then performing an internal survey amongst the team members of the WP4 project group. Once the complete sample for data repositories were selected, the WP4 task force manually and randomly selected ten metadata records per repository to be included in the FAIR assessment. D4.1 provides a more detailed description of the selection and criteria approach.

2.1 Choosing the metadata records for evaluation

98 data repositories from the Nordic and Baltic region were evaluated from the start of the study (January 2020). After excluding repositories that did not fit the evaluation criteria, the final sample consisted of 72 repositories. The 10 manually and randomly selected metadata records were evaluated once a month. These metadata records typically describe data that is held by the repository. In our experience, metadata records provide a good approximation of the FAIRness of

[5] Forbes magazine, 2016: "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says" https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=681bedb26f63

[6] Anaconda: "2020 State of Data Science: Moving From Hype Towards Maturity": https://know.anaconda.com/rs/387-XNW-688/images/Anaconda-SODS-Report-2020-Final.pdf

[7] Registry of Research Data Repositories: https://www.re3data.org/

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

all metadata from a given repository. We tested this by selecting six repositories that let us extract all their metadata via OAI-PMH and running one evaluation cycle with the F-UJI tool on all datasets (see Figure 1).

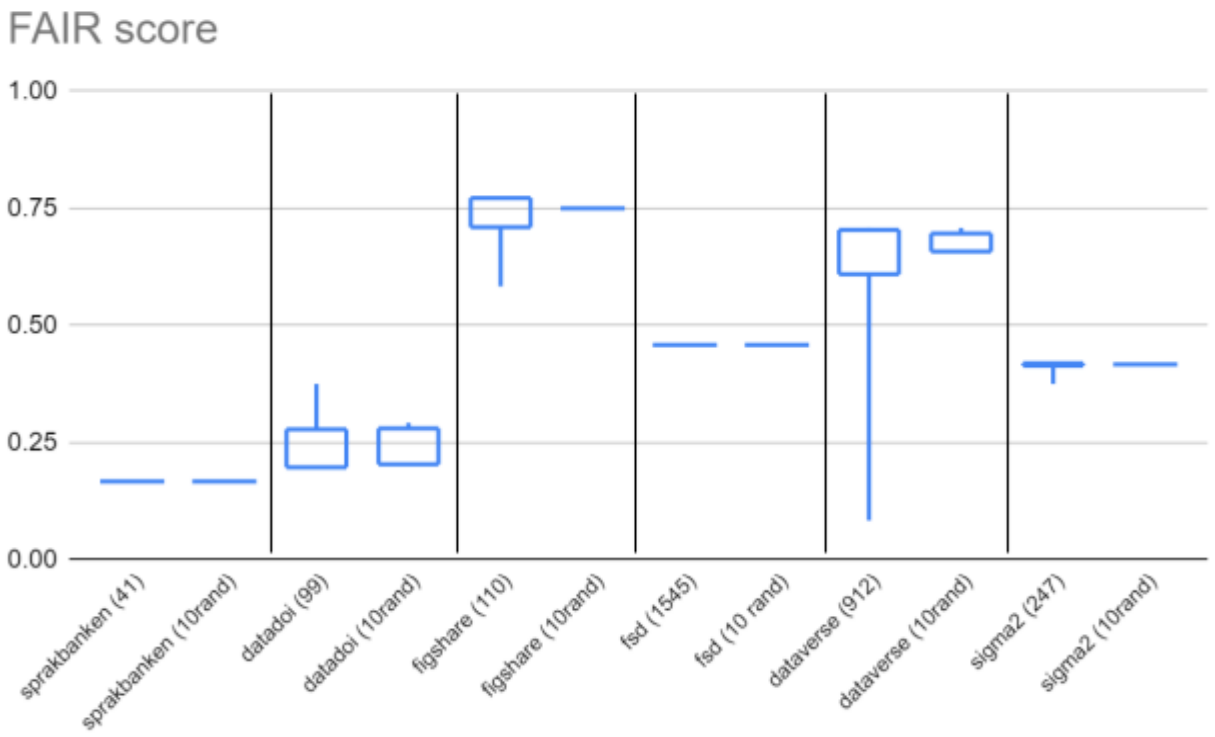


Fig. 1: FAIR score of the 6 repositories that expose their metadata via OAI-PMH visualised with a candlestick chart. The number in brackets after the repository name is the number of analysed metadata records. Tips of the candlestick represent highest/lowest FAIR-score, the boxes represent the average FAIR-score +/- 1 standard deviation. Horizontal bars indicate no variation in the FAIR-score.

Originally, the sample included duplicate entries for those Data Objects (DOs) that had a DOI-type persistent identifier, because we noticed that assessments based on DOIs provided very different results than those based on URIs and we thus wanted to monitor the differences. The differing results are due to a richer set of DataCite metadata included in the DOI-based evaluations. In May 2021, the sample was trimmed to provide only the DOI, as the use of DataCite metadata for DOIs could now be toggled when running the F-UJI tool.

2.2 FAIR evaluation tool

In the beginning of the project, the WP4 team used the FAIR Evaluator developed by Mark Wilkinson et al (2019) to evaluate the repositories. The team invested a substantial amount of time into testing and reporting back on issues with the FAIR Evaluator. By September 2020, we continued to experience inconsistencies in scoring and assessment of certain metrics. In addition, there was a lack of output to allow users to interpret and understand what was failing. At that point we were asked by Anusuriya Devaraju and Robert Huber to test out the F-UJI - Automated FAIR Data Assessment Tool (hereinafter: F-UJI tool) developed by the FAIRsFAIR project. The F-UJI tool gave results that better matched the perceived qualities that we were able to retrieve from

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

random samples and, importantly, the F-UJI tool was quite verbose in its output. We have since collaborated closely with the F-UJI developers with initial test results, iterative testing of reported bugs and suggested improvements and steadily contributed to the github issue tracker[8]. As a result, the last evaluation with the FAIR Evaluator was performed in May 2021.

In summary, the decision to switch from FAIR Evaluator to F-UJI was based on the following:

- Flexibility of the F-UJI evaluator to adopt batch runs and reports according to EOSC-Nordic's needs.
- Technical support received from the F-UJI team.
- Readiness of the F-UJI team to react immediately on EOSC-Nordic feedback and suggestions to improve the tool.
- F-UJI was sponsored by FAIRsFAIR, a European Union's Horizon 2020 project that had several connections and links to the EOSC-Nordic project.

2.3 Streamlining the assessments using Google Scripts

In order to get an overview of the FAIRness of the repository landscape and the possible changes in FAIRness in the Nordics and Baltics, it was necessary to streamline and automate the assessment process.

2.3.1 Automated execution of large sample sizes

To simplify the evaluation of the metadata records, a solution was designed that connected background processing using separate deployment of analytical tools (Wilkinson evaluator and F-UJI) with user interfaces implemented in Google Sheets.

A pool of separate worker scripts were set up to monitor changes in designated Google Sheets - e.g. for new tasks - scheduling parallel processing of records and updating Google Sheets with results of the evaluation. Evaluators as well as supporting services were set up in the ETAIS[9] cloud.

An example of the high level logic for each of the processor scripts is provided in the Appendix A.3.

2.3.2 Executing assessments with and without DataCite metadata

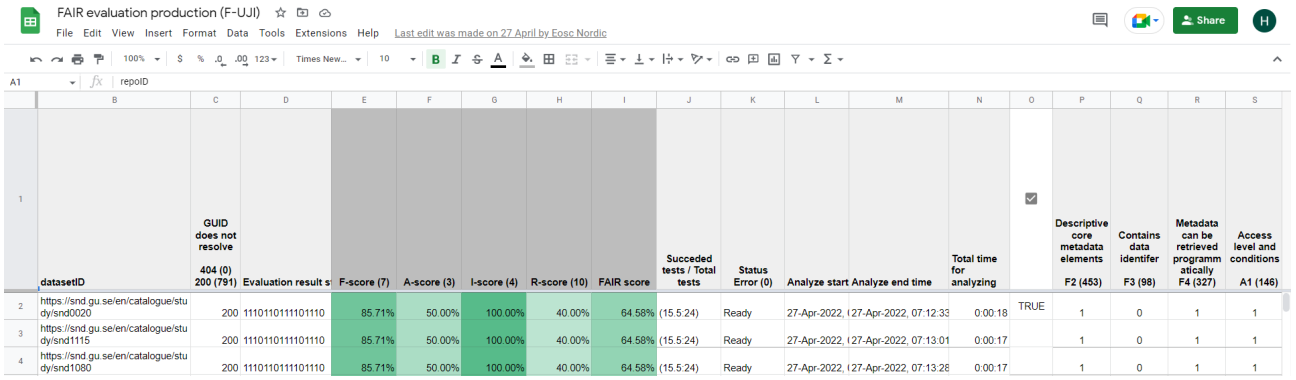
One interesting aspect to look for in an evaluation is to what degree repositories rely on external metadata providers such as DataCite[10] vs. providing FAIRness "locally" through their own data catalogue setup and requirements. In order to enable execution of assessments with and without DataCite metadata (provided via the DOI[11]), we enabled toggling of this F-UJI feature via the Google sheet so that the evaluation of a specific sample could be executed twice, once with DataCite metadata and once without DataCite metadata

[8] https://github.com/pangaea-data-publisher/fuji/issues
[9] https://etais.ee
[10] https://datacite.org/
[11] https://www.doi.org/doi_handbook/1_Introduction.html

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

2.3.3 Execution of assessments

To start an automated assessment, some manual steps have to be taken within the spreadsheet (see figure 2). These manual steps include for example saving the results from the latest evaluation and then clearing the spreadsheet from the current analysis and starting a new analysis.



	datasetID	GUID does not resolve 404 (0) 200 (791)	Evaluation result s	F-score (7)	A-score (3)	I-score (4)	R-score (10)	FAIR score	Succeded tests / Total tests	Status Error (0)	Analyze start	Analyze end time	Total time for analyzing		Descriptive core metadata elements F2 (453)	Contains data identifer F3 (98)	Metadata can be retrieved programmatically F4 (327)	Access level and conditions A1 (146)
2	https://snd.gu.se/en/catalogue/study/snd0020	200	1110110111101110	85.71%	50.00%	100.00%	40.00%	64.58%	(15.5:24)	Ready	27-Apr-2022, (	27-Apr-2022, 07:12:33	0:00:18	TRUE	1	0	1	1
3	https://snd.gu.se/en/catalogue/study/snd1115	200	1110110111101110	85.71%	50.00%	100.00%	40.00%	64.58%	(15.5:24)	Ready	27-Apr-2022, (	27-Apr-2022, 07:13:01	0:00:17		1	0	1	1
4	https://snd.gu.se/en/catalogue/study/snd1080	200	1110110111101110	85.71%	50.00%	100.00%	40.00%	64.58%	(15.5:24)	Ready	27-Apr-2022, (	27-Apr-2022, 07:13:28	0:00:17		1	0	1	1

Fig. 2: Screenshot of the GoogleSheet performing the assessments.

The frequency of our FAIR assessments has varied between 2-4 weeks when using F-UJI. Earlier, with the Wilkinson FAIR evaluator, the frequency was once every 3 months. In our experience, once a month is sufficient to ensure that we are capable of detecting changes of one or several repositories within the sample over a baseline period of 12-18 months.

2.3.4 Fixed version assessments (August 2021 – today)

We observed some fluctuations in FAIR scores as a result of different versions of the assessment tool. To minimise these variations, we fixed the version to v135 of the assessment tool that appeared to be stable. As the tools are evolving continuously, the dilemma will arise when a new, stable version with major updates will be released. Therefore, it is not trivial to compare the results of a FAIR assessment of a different study, that might have used a different FAIR assessment tool, a different version of the same assessment tool or performed the assessment on different repositories.

3. FAIR assessment results

This chapter presents the outcomes of the FAIR evaluations and general conclusions that can be drawn from a FAIR assessment. Based on our experiences, most repositories need specific support to increase their FAIR score in a meaningful way.

3.1 Findings

The FAIR Principles recommend that metadata should be separated from the data; metadata should point to the data and in return, the data should point to the metadata that holds its

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

description. The concept of FAIR Digital Object[12] plays an important role in terms of permanently and intelligently linking the metadata to the related data sets and vice versa. As a result, metadata files and data files can and may be assessed separately for FAIRness, whereby also the assessment criteria between metadata and datafiles may differ.

It is crucial that the metadata is machine-actionable, so that a machine agent can find, interpret and process the (meta)data, based upon the metadata found and harvested on the landing page of the repository. In other words, if we want to make research data reusable, the data needs to be enriched with metadata that can be found, interpreted and processed by machine agents and not only by the human eye. A website loaded with openly available PDF documents is not enough to qualify for FAIR. Machine-actionable metadata is key to implementing FAIR principles and especially to catering for interoperability and reusability.

The following general findings are based on FAIR assessments carried out from 2021 to 2022.

Repositories that run on established platforms (such as Dataverse, Figshare, and others) have noticeably higher FAIR scores and in addition scores are somewhat higher for CoreTrustSeal certified repositories.

During the webinars we found that many Nordic and Baltic repositories were discussing the need to define the right level and content of generic metadata, i.e. metadata related to the overall aspects of the research and the data, independent of the domain. We also noticed that defining the right level and content of the domain specific metadata was often a major challenge. It is essential that domain communities work towards a clearly defined metadata schema that repositories should adopt. Existing and widely used standard structures like DCAT, Dublin Core, DataCite or Schema.org may be helpful when defining metadata-schemas for a community. It is up to the communities to agree upon and define the standards.

With respect to current insufficiencies in machine-actionability, our study found that 30% of the repositories in our sample had no support for machine-actionable metadata at all. A few repositories supported machine-actionable metadata to some extent or had some metadata standards in place. Only a handful of repositories scored more than 50% on machine-actionable metadata. A lot of work remains to be done in this regard.

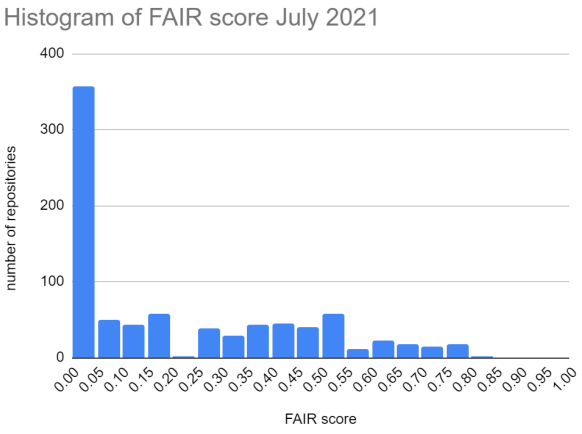
In addition to supporting machine-readability (in RDF or in JSON-LD), it is important that the available information is also Findable, Accessible, Interoperable and Reusable for researchers. FAIR for humans should coexist next to FAIR for machines.

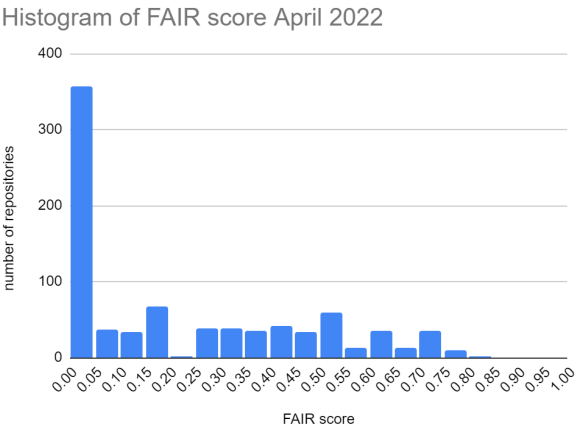
[12] A FAIR Digital Object (FDO) is the smallest machine-actionable data container that adheres fully to the FAIR Principles (definition by Erik Schultes). For more information and specification on FDO's, see https://www.go-fair.org/today/fair-digital-framework/ and/or https://fairdo.org

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

3.2 FAIR score development

To analyse the development of FAIR scores over time, each analysis is represented in a histogram. Figure 3 below shows the left-hand histogram with the fixed version (v.135) from July 2021 and, in comparison, the right-hand histogram contains the most recent assessment (April 2022). Both assessments are run on the exact same 850 metadata records with the same version of the F-UJI tool, both include DataCite metadata.





F	A	I	R	FAIR
0.368	**0.239**	**0.172**	**0.142**	**0.225**

Average FAIR scores split up after F, A, I and R from July 2021

F	A	I	R	FAIR
0.389	**0.261**	**0.194**	**0.156**	**0.243**

Average FAIR scores split up after F, A, I and R from April 2022

Fig. 3: Comparison of FAIR score distribution of all evaluated repositories for July 2021 and April 2022

Corresponding to the two histograms the average FAIR score are 0.225 and 0.243. This means that on average a slight increase can be observed over the timespan of 10 months. As seen in the tables within figure 3, it can be observed that the average F and A score are higher than the average I and R scores. This is due to the fact that the metrics connected to F and A scores are easier to fulfil by repositories.

A direct comparison to the earlier evaluations as presented in D4.1 is not possible because the version of the F-UJI tool has changed a few times, which also had an effect on the overall FAIR score. In addition, approximately 200 metadata records have been discarded since the start of the assessments until July 2021. Among the reasons for removing them were that some were not part of an individual repository but collections within DataverseNO, which means we analysed more than 10 datasets for this repository. Some datasets have been removed, because they had previously been analysed twice, using both their URL and DOI, as described in section 2.1.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

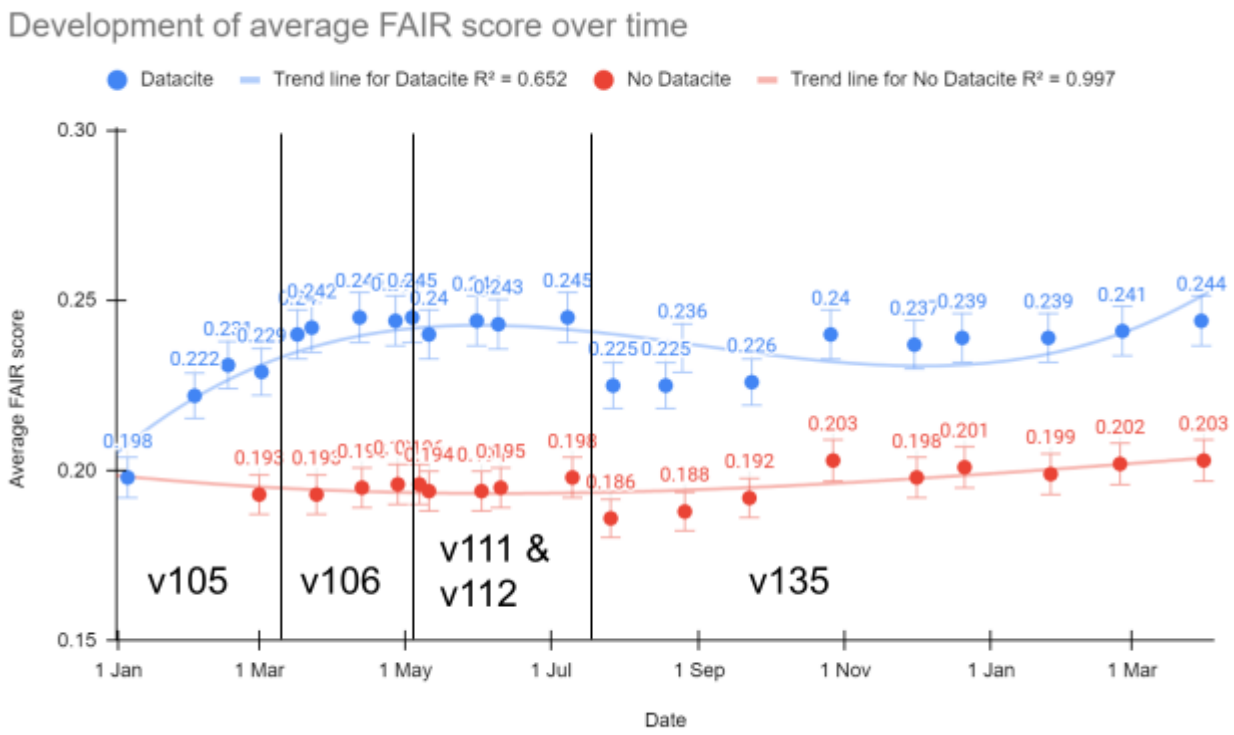


Fig. 4: Development of average FAIR score over time, including markings for change of version in F-UJI

Figure 4 shows the development of overall FAIRness of all repositories with DataCite metadata (blue) and without DataCite metadata (red). It is obvious that DataCite metadata significantly increases the FAIR score. It is important to note that the changes of versions of the F-UJI can also have an effect. It can be seen that a slight decrease in FAIR scores occurred due to the change from F-UJI version v106 to v111 and an even more significant change resulted from a change from v112 to v135. Afterwards, the FAIR score slowly increased again. This can be assumed to reflect the FAIRification changes in the metadata itself. It can also be observed that especially in the beginning of the study, the increase in FAIR score was higher in datasets that have been assigned DataCite metadata. This can be explained by a larger number of repositories being positively affected by changes in the DataCite metadata scheme than just 10 datasets, if a single repository decides to implement changes.

More detailed results for the evaluated repositories can be found in Appendix A.1.

3.3 Repositories with notable changes

Further analysis showed that the FAIR scores of most repositories did not change much during the project. Therefore, we selected a few repositories that had a notable change in their FAIR score for closer examination. These are SND (Swedish National Data service), Bolin Center Database, QsarDB, ICOS and DATICE.

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

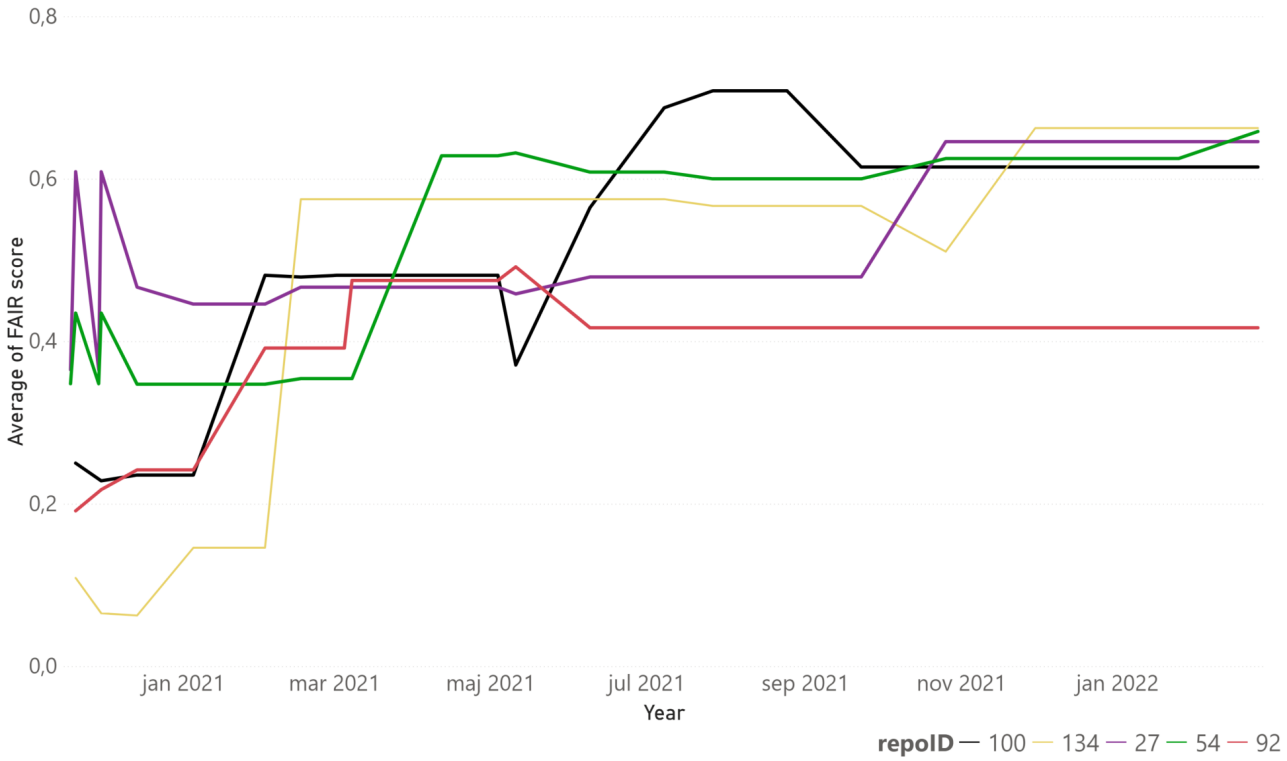


Fig. 5: Evolution of FAIR score of selected repositories over time. Black: QsarDB (100), yellow: ICOS (134), purple: SND (27), green: Bolin Centre Database (54), red: DATICE (92).

These four repositories were contacted and asked about the changes they made to their metadata records. Two of the named repositories have received personalised support from WP4. The material this support process was based on can be found in Appendix A.2.

SND (contact: Olof Olsson)
SND used the results of the F-UJI evaluation to improve the repository's machine readable metadata. The improvement of their FAIR score can be correlated with changes they made that included a fix in content negotiation on the landing page, so that F-UJI could access and use more metadata for evaluation, changing the access conditions in schema.org to machine readable, and changes to the metadata itself. The latter include a combination of English and Swedish metadata in the same structure, minor changes in the metadata getting sent to DataCite, and meta-links in the html header. They have published a simple example of their improvements on GitHub[13]

Bolin (contact: Ander Moberg)
Bolin Centre Database was part of the WP4 support process and shows a significant increase in their FAIR score between March and April 2021. Based on the guidance and support by WP4, they improved their machine-readable metadata files with additional data types (identifier, licence, is accessible for free, size, is based on and distribution).

[13] https://github.com/borsna/fair-examples

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Qsar (contact: Uko Maran, Sulev Sild)
QsarDB received personalised FAIRification support from WP4[14] and have participated in the FAIRification webinars (Step 1-5, see chapter 4). They were also part of the support process for achieving CoreTrustSeal certification.

One of the issues with the initial evaluation was that the datasets had both DOIs and handles assigned which F-UJI could not recognize. This error was resolved by an update to the F-UJI tool. QsarDB themselves also solved a couple of issues in regards to their platform and metadata fields. The repository is based on the DSpace platform, where they could change the configuration to include the missing metadata fields and add some generic Dublin Core metadata (DC.rights, DC.subject, DC.relation). Additionally, they updated the html header to include the link and type of data.

ICOS (contact: Oleg Mirzov)
ICOS showed a significant increase in F, A, I and R score in the spring of 2021 and then again in September 2021. ICOS was not part of WP4's support process. They achieved increased FAIRness by adding more metadata, including schema.org metadata, to the data object landing pages. The changes are documented in detail on the repository's GitHub page[15].

DATICE (contact: Guðbjörg Andrea Jónsdóttir)
DATICE used the F-UJI evaluation tool to monitor and improve their machine-readable metadata. After installing Dataverse, their FAIR metrics improved substantially, especially the F scores, which have reached an "advanced" level. The increase in F scores is attributable to the structured metadata that is now embedded in their landing page codes and can be retrieved programmatically. They have also managed to raise the I and R scores of their datasets somewhat by making improvements to DATICE's records in the DataCite metadata registry, e.g. by providing a valid (SPDX) machine readable licence for the datasets and information about their community specific metadata standards (i.e., through DATICE's re3data record).

3.4 Link rot

Since the beginning of the automated FAIR assessments and until the summer of 2022, approximately 6% (54/845) of the evaluated (meta)data can no longer be found under the link that was collected for evaluation in the beginning of the project. For some repositories, it affects all 10 metadata records, sometimes only individual data are affected. There are various reasons for link rot. In one case, it was a migration of all datasets to a new repository platform without a redirect from the old location. A few other repositories experience API connectivity issues that have not been fixed in months, and in some cases the F-UJI tool does not work as expected. All these issues are unfortunate, but cannot be avoided entirely. However, from the FAIR point of view this means that also the relations built on these links would eventually fail or become inaccurate. It is

[14] https://qsardb.org/blog/the-qsardb-repository-and-fair-principles
[15] https://github.com/ICOS-Carbon-Portal/meta/commits/master?after=9ba53b84891214ac401c06d13589c3772666b03f+314&branch=master&qualified_name=refs%2Fheads%2Fmaster

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

known that link rot increases over time[16] and therefore the use of well-maintained persistent identifiers is of utmost importance. It is worth mentioning that datasets with a DOI are less often affected by link rot than datasets without a DOI.

In this project we did not consider the option of content drift, meaning the link still leads to a testable resource, but entirely different content. The dataset in this study was far too large to keep track of the content behind the links.

3.5 From FAIR assessment to FAIR assistance

All FAIR evaluator tools should be used primarily as a conversation starter to help research communities to figure out where they can start to improve the FAIRness, rather than a certification tool. This became very clear through the support measures, both during the one-on-one discussions with the repositories as well as during the discussions in the support webinars. It is not expected, nor in many cases purposeful, for (meta)data to score 100% to be 'FAIR enough'. All metrics are not equally important for every research community. For these reasons, we would like to underline, similarly to 'Community-driven Governance of FAIRness assessment: an Open Issue, an Open Discussion (Wilkinson, M.D, Sansone, S-A. et al., 2022)[17], that FAIR evaluations of data repositories should be perceived more as FAIR assisting efforts, rather than FAIR assessing efforts. An observation made from the FAIR assessments in work package 4 is that communities or repositories are more likely to seek for FAIR assistance when they already have a considerably high score in the FAIR assessment. Reaching a broader range of data repositories, which were not necessarily all scoring high in the FAIR assessments, is one of the reasons why we decided to increase our efforts in community engagement activities.

4. Community engagement

In the EOSC-Nordic FAIRification initiative, the aim was to guide and coach the repositories to reach a higher level of FAIRness over time. This was done via a number of webinars, where we selected specific topics that contributed to a lower FAIRness score for a substantial part of repositories in our sample and where we felt improvements could be achieved. In this process, we have drawn several concrete conclusions that could be helpful for similar projects where data sharing and FAIR are relevant.

[16] Zittrain, Bowers & Stanton (2021). The Paper of Record Meets an Ephemeral Web: An Examination of Linkrot and Content Drift within The New York Times: https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37367405

[17] Wilkinson. M. D, Sansone. S-A, Mendéz. E, David. R, Dennis. R, Kleemola. M, Lacacgnina. C, Hecker. D, Nikiforova. A, Castro. L. J. (2022). Community-driven Governance of FAIRness assessment: an Open Issue, an Open Discussion [not yet published]

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Webinars are an excellent tool to provide information to a broad audience and with a maximum of interactions and engagement from the participants. Running several two hour webinars on a regular basis was an efficient way to guide the progress of the targeted repositories over time. The webinars consistently attracted around 100 participants. A typical format for the webinars were a short intro and project updates, two to three keynote speakers, and at least 30 minutes for a lively discussion based upon questions from the audience.

4.1 FAIR support webinars

The topics of the webinars were chosen based on the results of the FAIR assessment exercise. More specifically, the parts of the assessment where data repositories generally scored rather low formed the topics of the so called FAIRification webinars. The topics addressed were:

1. April 2020 - First assessment hackathon - Initial exercise
2. November 2020 – Webinar Step 1 – Focus on PIDs (global, unique, persistent, and resolvable identifiers)
3. February 2021 – Webinar Step 2 – Focus on split between data and metadata
4. April 2021 – Webinar Step 3 – Focus on generic metadata
5. October 2021 – Webinar Step 4 – Focus on domain-specific metadata
6. February 2022 – Webinar Step 5 – Focus on the value and limitation of FAIR assessment tools

The EOSC-Nordic project also offers support for repositories interested in achieving a CoreTrustSeal (CTS) certification or completing self-assessments against the CTS requirements and with added FAIR elements. The webinar 'From Self-Assessment to Certification with FAIR Results' addressed these topics on 3 June 2021.

4.2 Value and limitations of FAIR assessment tools

On 8 February 2022, the EOSC-Nordic project team organised a webinar on the subject of FAIR assessment tools aimed at repositories in the Nordic and Baltic countries, with the intention to guide and assist them in increasing the "FAIRness" of their metadata and datasets. The event attracted about 120 participants.

As this webinar on FAIR assessment tools brought a lot of important concepts together, we provide a number of highlights of this successful and very well attended event.

The webinar was the fifth in a series of multiple steps and focused on The Value and Limitations of FAIR Evaluators. While in earlier steps the FAIRification team was focused mainly on steps

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

towards improving the published metadata of the repositories, this webinar Step 5 concentrated on the use of FAIR evaluators that have the ability to “score” the FAIRness of the metadata from a given repository by running a number of tests against the FAIR Principles. Based upon these tests, the FAIR assessment tool is capable of showing a score and of providing clear guidance to the repositories so that improvements in FAIRness can be made.

For proper evaluations, it is crucial that the metadata of the repository is machine-actionable so that a machine agent can find, interpret and process the metadata found, for instance, on the landing page of the repository. The FAIR principles set out the guideline for FAIRness of data by indicating the relevance and importance of enriching datasets with clear machine-actionable metadata.

The FAIRification Webinar Step 5 on 8 February 2022 gave an interesting view on the value of these evaluators, showing several examples how repositories could improve their FAIRness level. The webinar also made clear that multiple evaluators exist and that they do not necessarily give the same outcomes. It also became clear that testing against concepts like “community standards” is not an easy task. The challenge now is to work towards sharper criteria-setting and towards convergence in defining, articulating and measuring the different FAIR components / metrics, so that multiple evaluators will give more or less similar scores. The webinar also demonstrated that the a-priori use of community standards, templates, vocabularies and ontologies, collectively defined as the “FAIR at the source” process, is seen as a much easier route than trying to work on curation of existing (meta)data sets.

The webinar Step 5 also demonstrated that organising a Metadata for Machine workshop (M4M**)** and/or defining a FAIR Implementation Profile (FIP) for a community or domain could be a very good exercise to define and publish the implementation choices & enabling resources of a particular community. For more information on these M4M’s and FIP’s, see:
https://www.gofairfoundation.org/m4m/

Main takeaways of the webinar**:**

- FAIR data is not the same as “open and free data”.
- Data needs to stay fully under the control of the data-owners.
- Communities play an important role in defining, publishing and sharing metadata schemas and FAIR implementation choices.
- Machine readability / Machine actionability is crucial. The machine has to interpret and understand what is meant.
- Community Implementation choices - published in a FAIR Implementation Profile (PIF).
- Encourage Metadata for Machine workshops (M4M’s) for communities in order to define the metadata schemas and templates.
- Funders are to demand “FAIR at the source” and to thus drive convergence and offer domain relevant maDMPs (machine-actionable Data Management Plans).
- Convergence needs to take place among the different evaluators currently in use.

EOSC-Nordic project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857652

- Repositories may play an important role in the provision of FAIR data by automatic inclusion of for instance a proper PID, relevant metadata that points to the data and/or the adherence to a domain specific vocabulary.
- Focus on making data FAIR at the source, rather than on curating legacy data.

All webinar material (agenda, presentations, chats, recording) is available: https://www.eosc-nordic.eu/fairification-step-5-webinar-value-and-limitations-of-fair-assessment-tools/

5. Discussion

Working towards an increase in FAIRness is a process. Clear improvements in FAIR scores were seen when repositories actively worked on a step-by-step approach towards understanding and implementing the FAIR principles to suit their needs. Raising awareness on FAIR at multiple levels was one of the greatest achievements of this work-package.

Convergence of FAIR evaluation tools is also important and the work package has been able to contribute to it. The forthcoming whitepaper 'Community-driven Governance of FAIRness: an Open Issue, an Open Discussion' by Mark Wilkinson et. al.[18] calls for a convergence on how FAIR research is assessed and how the FAIR principles are understood and interpreted. The paper also outlines the need for an open discussion on FAIR governance, the mechanisms involved for collaboratively achieving that, and the actors that play important roles in driving this change.

Community-agreed standards were another welcome side effect of the webinars that we have been running (see example of the Climate Control Community NICEST2[19]).

Currently, there are several tools available for assessing FAIRness of (meta)data. The EOSC Task Force FAIR Metrics & Data Quality[20] identified as many as 13 independent FAIR evaluation platforms, all listed in an online registry of resources for FAIR assessment of digital objects.[21] Most of them are questionnaire-based and only a few are automated. There are, however, ambiguities in the assessment methods and scoring, as none of the outputs can be compared with each other. This ambiguity also became very clear during the webinar "Values and limitations of FAIR assessment tools" (see chapter 4.2).

The EOSC Task Force FAIR Metrics & Data Quality and an Apples-to-Apples group (engagement with the assessment tool developers to tease out the metadata gathering workflow) are trying to address these ambiguities, where the evaluators are using just about the same metrics, but they are testing them in different ways and are using different approaches to gathering the metadata to

[18] Wilkinson. M. D, Sansone. S-A, Mendéz. E, David. R, Dennis. R, Kleemola. M, Lacacgnina. C, Hecker. D, Nikiforova. A, Castro. L. J. (2022). Community-driven Governance of FAIRness assessment: an Open Issue, an Open Discussion [not yet published]
[19] https://www.eosc-nordic.eu/events/nicest2-hackathon-on-fair-climate-data/
[20] https://www.eosc.eu/advisory-groups/fair-metrics-and-data-quality
[21] https://www.Fairassist.org

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

be tested. There is a clear call for e.g. agreed controlled vocabularies and agreed data formats and models to help solving the puzzle and to reach a consensus on the level of detail for applying standards. Some of the conclusions made by the Apples-to-Apples group is that the solution should also be universally applicable and undependable of technical options. The group agreed that there is a call for signposting[22] and use of link headers instead of DOIs resolving to landing pages. This will be taken forward by two sub-groups concentrating on first defining metrics and tests, and second developing a series of benchmark environments.

6. Conclusions and Next Steps

While the project team saw quantitative improvement in *descriptive core metadata elements* and in *automatically retrieved metadata* over the last 3-6 months, it is risky to make any hard conclusions as the improvements could also be partly due to the fact that updates of the evaluation software had some impact on the evaluation scores.

Rich metadata is important for FAIR. The project facilitated Metadata Templates and Metadata for Machine Workshops (M4Ms) as a structure to create consensus within a community. We noticed that *controlled vocabularies* are crucial in reducing the amount of free text and strongly increases the interoperability and reusability of data. Combining domain expertise and FAIR metadata-expertise in an M4M Workshop can result in a (domain specific) metadata schema that can be used, reused and shared by a particular community.

The following steps could help to speed up the process:

1. Build a community metadata schema (for example, as a result from a M4M workshop)
2. Store the metadata schema (for example, as a (CEDAR[23]) template).
3. Publish the agreed metadata templates (e.g. Bioportal[24])
4. Present the template to researchers as a web-form.

This above mentioned process was successfully demonstrated and used for the NICEST 2 Climate Control community, as part of WP 5.3.2. However, it is worth noting that several communities already have well-established metadata schemas (see. e.g. Goble and Juty 2021[25]) and generally, it is advisable to build on already existing and working schemas and standards instead of creating totally new ones.

[22] https://signposting.org/
[23] https://metadatacenter.org/
[24] https://bioportal.bioontology.org/
[25] Carole Goble, & Nick Juty. (2021). Analysis of existing research data cataloguing efforts towards integrated discovery. https://doi.org/10.5281/zenodo.4693217

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Another conclusion from the project was that a large percentage of the data repositories struggled to provide the right object types to datasets. e.g. whether the file is a graph, a report, a picture, or an Excel file. Controlled vocabularies with mandatory input could help to improve the situation.

The different webinars have made us aware that participants are eager to get recommendations and are particularly interested in a number of topics:

- How to get more practical examples of FAIR implementations
- How to organise access control related to sensitive data
- How to maintain a balance between human and machine-actionability
- How to receive tangible benefits from the presented developments

In order to address the sustainability of the project outcomes, it has to be kept in mind that the results of a FAIR assessment cannot be considered from a purely numeric point of view. Therefore, it is probably not very helpful to develop the current automated assessments further into finished products. One should question the need for extensive mass evaluation of metadata or data objects. On the other hand, building on the materials developed so far, a service for research communities to test selected data objects could be useful. Automated evaluations provide quick feedback on the FAIRness and help to highlight the areas of improvement. If a repository is interested in their own FAIRness, it is possible for them to manually evaluate randomly selected datasets in F-UJI web interface, since we have demonstrated this to be a good approximation (see chapter 2.1). The results should be carefully evaluated afterwards to decide where improvements make the most sense for a specific community. However, full versions of scripts created for the assessment are available on-demand.

At the time of writing, the project team is also comparing results with an EC project[26] and analysing the reasons for differences in FAIR scores. It seems that, for example, inclusion of many small repositories may be a large contributor to lower scores. This reinforces our view that the FAIR scores themselves do not provide simple nor true insights into the state of FAIR data in Europe, at least not ones that concrete actions can be derived from, and that careful interpretation of scores and underlying factors is needed.

To summarise our conclusions, we have in this project tried to keep the information sharing and the guidance process on a very practical level. All material is available on the EOSC-Nordic project's Knowledge Hub[27]. The mass-evaluation of the repositories was certainly a good starting point in order to figure out what the biggest challenges were in order to start assisting the repositories. As a result, we have been able to make repositories in the Nordic and Baltic countries more aware of the necessity of sharing data and we have been able to guide the repositories through the technical challenges of making their (meta)data more FAIR, through a step-by-step process. We have witnessed an overall increased awareness of the FAIR principles and a significant increase in applying the FAIR principles for data sharing and data visiting purposes. Increased community

[26] Report coming soon, information comes from personal communication with Maaike Verburg <maaike.verburg@dans.knaw.nl>
[27] More information available on EOSC-Nordic's Knowledge Hub: https://eosc-nordic.eu/knowledge-hub/

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

engagement efforts was a conscious decision made by the WP4 team, as we noticed that community engagement was the right format for getting the discussions going, reaching the repositories representing different stages of FAIR maturity, and getting the repositories aware of the ‘why’ and the ‘how’ of reaching higher FAIR scores. We also noticed that the repositories with the highest FAIR scores were generally the most active ones in the webinars and were most keen on improving their FAIR scores even more.

EOSC-Nordic project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857652

Bibliography

Anaconda - Data science platform and the foundation of modern machine learning (2020): “2020 State of Data Science: Moving From Hype Towards Maturity”: https://know.anaconda.com/rs/387-XNW-688/images/Anaconda-SODS-Report-2020-Final.pdf

Andreas Ortmann Jaunsen, Mari Kleemola, Tuomas J. Alaterä, Heikki Lehvaslaiho, Adil Hasan, Josefine Nordling, & Pauli Assinen. (2020). D4.1 An assessment of FAIR-uptake among regional digital repositories (1.0). Zenodo. https://doi.org/10.5281/zenodo.4045402

Anusuriya Devaraju, & Robert Huber. (2020). F-UJI - An Automated FAIR Data Assessment Tool (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.4063720

Carole Goble, & Nick Juty. (2021). Analysis of existing research data cataloguing efforts towards integrated discovery. Zenodo. https://doi.org/10.5281/zenodo.4693217

CEDAR: https://metadatacenter.org/

DataCite: https://datacite.org/

Estonian Scientific Computing Infrastructure (ETAIS): https://etais.ee

FAIRassist, a FAIRsharing resource: https://www.Fairassist.org

FAIR metrics and Data Quality Task Force, EOSC: https://www.eosc.eu/advisory-groups/fair-metrics-and-data-quality

FAIRsFAIR: https://www.fairsfair.eu/

FDO - FAIR Digital Object: https://www.go-fair.org/today/fair-digital-framework/ and https://fairdo.org

Forbes magazine, 2016: “Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says” https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=681bedb26f63

F-UJI Github issue tracker: https://github.com/pangaea-data-publisher/fuji/issues

F-UJI - an automated FAIR data assessment tool: https://www.f-uji.net/

ICOS repository Github page with documented changes: https://github.com/ICOS-Carbon-Portal/meta/commits/master?after=9ba53b84891214ac401c06d13589c3772666b03f+314&branch=master&qualified_name=refs%2Fheads%2Fmaster

Knowledge Hub of EOSC-Nordic: https://eosc-nordic.eu/knowledge-hub/

EOSC-Nordic project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857652

Leach. P, Microsoft, Mealling. M, Refactored Networks, LLC, Salz. R, DataPower Technology, Inc. (2005). Globally Unique Identifier: http://www.ietf.org/rfc/rfc4122.txt
NCBO BioPortal: https://bioportal.bioontology.org/

NICEST2 hackathon on FAIR climate data:
https://www.eosc-nordic.eu/events/nicest2-hackathon-on-fair-climate-data/

QSAR DataBank - platform for organisation and long term preservation of QSAR data and models:
https://qsardb.org/blog/the-qsardb-repository-and-fair-principles

Registry of Research Data Repositories: https://www.re3data.org/

Signposting the scholarly web: https://signposting.org/

SND - Swedish National Data Service, repository Github page with documented changes:
https://github.com/borsna/fair-examples

The International DOI Foundation Handbook (2015):
https://www.doi.org/doi_handbook/1_Introduction.html

Wilkinson. M. D, Sansone. S-A, Mendéz. E, David. R, Dennis. R, Kleemola. M, Lacacgnina. C, Hecker. D, Nikiforova. A, Castro. L. J. (2022). Community-driven Governance of FAIRness assessment: an Open Issue, an Open Discussion [not yet published]

Zittrain, Bowers & Stanton (2021). The Paper of Record Meets an Ephemeral Web: An Examination of Linkrot and Content Drift within The New York Times:
https://dash.harvard.edu/handle/1/37367405

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

Appendices

A.1: Results of the latest F-UJI evaluation (April 2022)

Colour code:

F - A - I - R and FAIR score:

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

A high score is good, it means a high compliance with the FAIR principles (according to the F-UJI tool)

Sigma(F - A - I - R and FAIR)

0	0.1	0.2	0.3	0.4	0.5

A low sigma score is good, it means the results of the analyses have a low variation.

F-UJI evaluator (27 April 2022)

repo ID	Name	Data-sets	Platform	F-score	A-score	I-score	R-score	FAIR	Sigma	Sigma (F)	Sigma (A)	Sigma (I)	Sigma (R)	CTS	DSA	WDS	CLARIN
2	CLARIN-DK	10	Dspace	57.14 %	50.00 %	12.50 %	17.78 %	32.4 %	0.04	0.00	0.00	0.13	0.10	X			X
3	DDA	10		85.71 %	33.33 %	75.00 %	10.00 %	45.83 %	0.00	0.00	0.00	0.00	0.00				
4	Det Kgl. bibliotek	20		31.79 %	13.33 %	11.25 %	9.00%	16.56 %	0.13	0.33	0.27	0.24	0.18				
6	Kielipankki	10	META-SHARE	14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00	X			X
7	Data Service Portal Aila	10		71.43 %	66.67 %	25.00 %	20.00 %	41.67 %	0.00	0.00	0.00	0.00	0.00				
8	Fairdata IDA	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
10	NSD	20	NESSTAR	43.22 %	13.33 %	10.00 %	8.00%	19.27 %	0.11	0.36	0.17	0.13	0.10	X			
11	HUNT Databank	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
13	CLARINO Bergen Center repository	10	Dspace	57.14 %	50.00 %	20.00 %	20.00 %	34.58 %	0.04	0.00	0.00	0.11	0.11	X			X
16	Språkbanken	10		21.43 %	33.33 %	0.00%	20.00 %	18.75 %	0.00	0.00	0.00	0.00	0.00				
17	ESS Data	9		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
18	TROLLing	10	Dataverse	95.71 %	66.67 %	70.00 %	46.00 %	67.08 %	0.03	0.07	0.00	0.11	0.05	X			

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

19	EED	10	Nestar	14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
24	Språkbanken	13		15.94 %	0.00%	0.00%	0.00%	4.65 %	0.01	0.03	0.00	0.00	0.00		X		X
25	Lund University Humanities Lab corpus server	20		20.72 %	0.00%	0.00%	0.00%	6.04 %	0.02	0.07	0.00	0.00	0.00				
26	su.figshare.com	10	Figshare	87.14 %	56.67 %	60.00 %	46.00 %	61.67 %	0.13	0.28	0.23	0.32	0.25				
27	SND	20		85.71 %	50.00 %	100.0 %	40.00 %	64.58 %	0.00	0.00	0.00	0.00	0.00	X			
28	ICES data portals	8		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
29	JASPAR	10		35.71 %	33.33 %	25.00 %	30.00 %	31.25 %	0.00	0.00	0.00	0.00	0.00				
30	STRING	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
32	GBIF	11	IPT	79.22 %	33.33 %	50.00 %	50.00 %	56.44 %	0.02	0.08	0.00	0.00	0.00			X	
39	HPA	10		20.72 %	10.00 %	15.00 %	6.00%	12.29 %	0.08	0.10	0.16	0.24	0.10				
41	Fishbase	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
45	ISIG	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
47	GERDA	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
49	ACTRIS	8		50.00 %	25.00 %	12.50 %	15.00 %	26.04 %	0.13	0.38	0.27	0.13	0.16				
52	NPDC	20		44.29 %	20.00 %	26.25 %	21.50 %	28.75 %	0.13	0.31	0.25	0.28	0.23				
54	Bolin Centre Database	10		94.29 %	96.67 %	70.00 %	38.00 %	67.08 %	0.06	0.12	0.11	0.11	0.18				
55	SMHI open data	10		35.71 %	33.33 %	25.00 %	20.00 %	27.08 %	0.00	0.00	0.00	0.00	0.00				
57	NIRD Archive	20		50.71 %	20.00 %	12.50 %	10.00 %	23.54 %	0.11	0.36	0.17	0.13	0.10				
60	GTN-P Database	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
62	UNITE	20		42.86 %	19.17 %	18.75 %	18.50 %	25.73 %	0.11	0.29	0.20	0.21	0.19				

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

63	Estonian Biocentre Public Data	10		21.43 %	33.33 %	0.00%	20.00 %	18.75 %	0.00	0.00	0.00	0.00	0.00				
64	DataDOI	10		70.00 %	50.00 %	22.50 %	21.00 %	39.17 %	0.11	0.26	0.28	0.14	0.12				
65	CELR META-SHARE	20		42.86 %	33.33 %	12.50 %	10.00 %	22.92 %	0.08	0.29	0.00	0.13	0.10	X			X
66	AHEAD	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
68	USN RDA	10	Figshare	98.57 %	63.34 %	72.50 %	47.00 %	68.33 %	0.04	0.05	0.11	0.08	0.05				
71	LOAR	10		78.57 %	58.34 %	25.00 %	27.00 %	45.62 %	0.09	0.23	0.21	0.12	0.10				
72	AIDA Data Hub	10		85.71 %	50.00 %	65.00 %	30.00 %	54.58 %	0.03	0.00	0.00	0.13	0.00				
73	QoG Institute's data	10		36.43 %	33.33 %	7.50%	20.00 %	24.38 %	0.07	0.24	0.00	0.12	0.00				
76	JYX	9		50.00 %	61.11 %	0.00%	10.00 %	26.39 %	0.02	0.00	0.08	0.00	0.00				
78	B2SHARE	12	Invenio	44.05 %	27.78 %	10.42 %	12.50 %	23.26 %	0.13	0.35	0.34	0.13	0.15				
79	DH	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
80	NLL	10		31.43 %	53.34 %	40.00 %	35.00 %	37.08 %	0.10	0.09	0.28	0.21	0.20				
84	RTU RIS	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
85	FinBIF	23		20.50 %	33.33 %	64.13 %	15.65 %	27.36 %	0.06	0.03	0.00	0.21	0.08				
87	SARV	10		21.43 %	33.33 %	0.00%	20.00 %	18.75 %	0.00	0.00	0.00	0.00	0.00				
92	SSRI	9		66.67 %	22.22 %	50.00 %	26.67 %	41.67 %	0.15	0.39	0.17	0.38	0.20				
94	IINH	9		14.29 %	0.00%	0.00%	10.00 %	8.33 %	0.00	0.00	0.00	0.00	0.00				
100	QsarDB	20		82.15 %	100.00%	37.50 %	45.00 %	61.46 %	0.06	0.18	0.00	0.13	0.05				
104	Bird	20		35.71 %	36.66 %	5.00%	6.00%	18.33 %	0.07	0.19	0.07	0.10	0.12				
106	Migration Institute of Finland	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

108	Musiikkiarkisto	4	CKAN	35.71 %	66.67 %	50.00 %	50.00 %	47.92 %	0.03	0.00	0.00	0.00	0.12				
109	SLS	14		19.39 %	23.81 %	0.00%	7.14%	11.61 %	0.04	0.03	0.16	0.00	0.05				
113	SweFreq	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
114	Metabolic Atlas	10		35.71 %	33.33 %	25.00 %	20.00 %	27.08 %	0.00	0.00	0.00	0.00	0.00				
115	SEAD	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
116	NOW	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
117	SNM Digital Assets	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
120	GEUS	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
123	LARM	10		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
127	Garamantas	9		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
129	MMB	10		21.43 %	33.33 %	0.00%	20.00 %	18.75 %	0.00	0.00	0.00	0.00	0.00				
130	PlutoF	20		42.86 %	19.17 %	16.25 %	17.50 %	24.90 %	0.11	0.29	0.20	0.19	0.19				
131	MIDAS	9		14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				
133	IINH BIOTA	10		21.43 %	33.33 %	0.00%	20.00 %	18.75 %	0.00	0.00	0.00	0.00	0.00				
134	ICOS	10		82.85 %	60.00 %	75.00 %	53.00 %	66.25 %	0.04	0.06	0.14	0.00	0.08				
135	CESSDA DC	14		34.70 %	14.29 %	28.57 %	11.43 %	21.43 %	0.16	0.34	0.23	0.47	0.19				
136	DTU data	10	figshare	95.71 %	56.67 %	70.00 %	43.00 %	64.58 %	0.05	0.07	0.16	0.11	0.07				
137	CLARIN IS	10	CLARIN	57.14 %	48.33 %	22.50 %	26.00 %	37.29 %	0.03	0.00	0.05	0.08	0.08				
138	LIDA	11	Nesstar	14.29 %	0.00%	0.00%	0.00%	4.17 %	0.00	0.00	0.00	0.00	0.00				

The spreadsheet includes the results as presented here for the latest evaluation, as well as additional evaluations from August 2021, April 2021 and December 2020: https://docs.google.com/spreadsheets/d/13U8pF4vNnH92hpBQTCIvBvOwkuA1RyAoZIztibKDu6Q/edit#gid=1381189767

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

A.2: Recommendations for FAIR Evaluation Services, based upon the F-UJI Automated FAIR Data Assessment Tool [28,29]

The FAIR Maturity Evaluation Service deploys 17 automated tests confirming FAIR properties for the majority of the FAIR Principles. See FAIRsFAIR Deliverable 4.5 for more details on each of the metric tests (eg. appendix A). An overview of the 17 tests:

METRIC	METRIC NAME	FAIR Principle	METRIC DESCRIPTION	Recommendation on TEST- FAILURE	Detailed information relevant to recommended actions in case of TEST-FAILURE
1	FsF-F1-01D	F1	**Data** is assigned a globally unique identifier.	Provide an identifier for the dataset, e.g. by assigning a Global Unique Identifier[30](s) – such as URL or IRI.	For a digital object (metadata and data) the assumption is that the object is identified using a Globally Unique Identifier (GUID). The GUID should resolve to a 'landing page' that contains two elements; metadata and data.
2	FsF-F1-02D	F1	**Data** is assigned a persistent identifier.	Provide a persistent identifier (PID). Make sure the persistent identifier is also given in the metadata you provide on the landing page.	This indicator tests if the unique identifier of the data resource is likely to be persistent. PID schemes-based assessment supported by the assessment service: ● ark ● arxiv ● bioproject ● biosample ● doi ● ensembl ● genome ● gnd

[28] https://www.f-uji.net/

[29] Anusuriya Devaraju, & Robert Huber. (2020). F-UJI - An Automated FAIR Data Assessment Tool (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.4063720

[30] GUID stands for Globally Unique Identifier that follows a specific structure defined in RFC 4122 (http://www.ietf.org/rfc/rfc4122.txt)

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

					• handle • lsid • pmid • pmcid • purl • refseq • sra • uniprot • urn See https://identifiers.org/ for additional types.
3	FsF-F2-01M	F2	**Metadata** includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.	Provide the minimum (6) core descriptive elements. Use a common metadata standard (e.g. schema.org/Dataset, Dublin Core, DCAT-2 etc.). Make sure this metadata is either embedded in the landing page, linked via typed links (see: signposting.org) or delivered via content negotiation. See also FsF-F4-01M	The 6 core descriptive elements are: • creator • title • object_identifier • publication_date • publisher • object_type
4	FsF-F3-01M	F3	**Metadata** includes the identifier of the data it describes.	The metadata should explicitly specify the identifier of the data (e.g. download link) such that users can discover and access the data through the metadata. If the identifier specified is persistent and	For a digital object (PID, metadata and data) the assumption is that the data has its own identifier and that this is provided by standardised keys/predicates. See below for some examples: http://schema.org/distribution http://www.w3.org/ns/dcat#distribution For software the following example illustrates how this can be achieved:

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

				points to a landing page, the data identifier and links to download the data content should be taken into account in the assessment.	schema:codeRepository
5	FsF-F4-01M	F4	**Metadata** is offered in such a way that it can be retrieved programmatically.	Metadata is given in a way major search engines can ingest it for their catalogues (JSON-LD, Dublin Core, RDFa) or registered in major research data registries (DataCite, Mendeley Data, Google Dataset Search(partly)). This means structured content could be provided as: ● 'structured' metadata embedded in landing page ● content negotiation ● typed or signposting links	Supported tests of metadata retrieval/extraction: ● Embedded DublinCore ● Embedded OpenGraph ● Schema.org JSON-LD (Embedded) ● Schema.org JSON-LD (Negotiated) ● Datacite Search ● Typed Links ● Signposting Typed Links ● RDF-based Typed Links ● Linked Data (RDF) ● Guessed XML Link ● Generic XML (Negotiated) ● Embedded RDFa ● Embedded Microdata ● OAI-ORE
6	FsF-A1-03D	A1	**Data** is accessible through a standardized communication protocol.	Metadata includes a resolvable link to data based on standardized web communication protocols.	List of standardized communication protocols: ● http ● https ● ftp
7	FsF-A1-02M	A1	**Metadata** is accessible through a	Assure your metadata resources can be resolved	List of standardized communication protocols: ● http ● https

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

			standardized communication protocol.	(f.i. InChi keys, DOIs , Handles and URLs)	• ftp
8	FsF-A1-01M	A1	**Metadata** contains access level and access conditions of the data.	The access rights is not to be confused with licenses.	Notes: Avoid links for licenses (access does not equal license) add predicates for “access rights” from JSON-LD, Datacite, EC, Core (controlled vocabulary). For example: https://guidelines.openaire.eu/en/latest/data/field_rights.html http://publications.europa.eu/resource/authority/access-right https://www.dublincore.org/specifications/dublin-core/dcmi-terms/terms/accessRights/
9	FsF-I1-01M	I1	**Metadata** is represented using a formal knowledge representation language.	Use a machine-actionable language to structure your metadata or use an existing metadata template.	Parsable, structured metadata (JSON-LD, RDFa) is embedded in the landing page XHTML/HTML code. OR Parsable, graph data (RDF, JSON-LD) is accessible through content negotiation, typed links or sparql endpoint.
10	FsF-I1-02M	I1	**Metadata** uses semantic resources.	Use vocabulary namespace URIs in your metadata.	Known semantic resources are: • LOV (https://lov.linkeddata.es/dataset/lov/) • LinkedOpenDataCloud (https://lod-cloud.net) • refer to Zenodo doc (p.19)

11	FsF-I3-01M	I3	**Metadata** includes links between the data and its related entities.	Assure that your dataset can be represented as Linked Data and has machine-actionable references to other metadata.	Example:
					`"related_resource": "snd.gu.se",`
					`"relation_type": "isPartOf"`
					
					Examples from DataCite
					https://support.datacite.org/docs/schema-optional-properties-v43#12-relatedidentifier

12	FsF-R1-01MD	R1	**Metadata** specifies the	The metadata contains information about the data it describes.	This information includes: • resource type

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

			content of the **data**.		• verifiable data descriptors (file info, measured variables or observation type) • file size and type
13	FsF-R1.1-01M	R1.1	**Metadata** includes license information under which **data** can be reused.	Assure an explicit pointer to the license or use existing schemas that include license terms.	License information has to be given as one of the metadata elements. The recognized license should be given as a link and be registered at SPDX.
14	FsF-R1.2-01M	R1.2	**Metadata** includes provenance information about data creation or generation.	Metadata contains elements which hold provenance information and can be mapped to PROV. Preferably, use formal provenance ontologies.	
15	FsF-R1.3-01M	R1.3	**Metadata** follows a standard recommended by the target research community of the data.	Community-specific metadata standards are detected using namespaces or schemas found in provided metadata or metadata services outputs.	In the re3data.org record of the responsible repository, the community specific metadata standard can be found. Typed links OAI-PMH Sparql
16	FsF-R1.3-02D	R1.3	**Data** is available in a file format recommended by the target research community.	The format of a data file given in the metadata is listed in the long term file formats, open file formats or scientific file formats controlled list	Examples for long term file formats: https://schema.org/encodingFormat

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

A.3: An example of the high level logic for each of the processor script

```
def main():
    while True:
        try:
            googlesheet_url = config['GOOGLE']['googlesheet_url']
            evaluator_url = config['EVALUATOR']['evaluator_url']
            script_start_check(googlesheet_url, evaluator_url)
        except Exception as e:
            traceback.print_exc()
            print('Failed. Cause of the error: ' + str(e))
            time.sleep(60)
            continue
        else:
            time.sleep(60)
```


EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652

List of Figures and Appendices

EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652