

# Blueprint for automated assessment of FAIRness following the EOSC-Nordic approach

Hannah Mihai, Ahti Saar, Mari Kleemola (November 2022)

## Introduction

This blueprint describes the basic steps to perform the assessment of the FAIRness of multiple repositories' metadata following the EOSC-Nordic approach. The process uses Google scripts and spreadsheet, separate worker scripts, and the F-UJI evaluator. For each step, we provide additional hints and further material (e.g. examples and scripts from EOSC-Nordic) that will allow interested parties to perform a similar process <sup>1</sup>.

The process steps are:

1. Collect dataset identifiers
2. Execute the assessments
3. Analyze results

## Process steps

### 1. Collect dataset identifiers

Define a criteria for repositories to be included in the assessment and select a representative number of datasets (or metadata records) from each repository. For all datasets, collect the GUID (global unique identifier). In the EOSC-Nordic sample, we found out 10 randomly selected datasets from each repository to be sufficiently representative. Repositories that do not provide a unique identifier for each dataset can not be tested using the evaluator tools. The EOSC-Nordic sample selection is explained in detail in *D4.1 An assessment of FAIR-uptake among regional digital repositories* (<https://doi.org/10.5281/zenodo.4045402>).

Store the repository ID, the name of the repository and the GUID of each dataset in a Google spreadsheet.

### 2. Execute the assessments

You need to install the F-UJI tool locally to run the assessments. The instructions are available at <https://github.com/pangaea-data-publisher/fuji>

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 we used in the EOSC-Nordic project to v135 of F-UJI.

---

<sup>1</sup>The process and scripts were developed for project purposes and actively maintained during the project end (November 2022). The process was designed at the beginning of the project in 2020. Since then, the assessment metrics and tools have developed further. Thus if we started the project now, we would most likely make different choices and especially add automation. However, the process was a good working solution for the project. Developing this into a service was out of scope of EOSC-Nordic project.

## Automation part

The automation part is built on microservice, which can be run locally on the user's computer or as a cloud service in a virtual machine. This microservice takes the identifier from the Google sheet and pushes it to the F-UJI evaluator, pulls after analyzing the results and pushes it to the same Google sheet.

Since the result pulled from F-UJI evaluator contains a lot of detailed information, the necessary information must be extracted from the result before pushing it to the Google sheet. Information which is pushed to Google sheet is:

- Evaluation result string
- Each FAIR component results separately
- Average FAIR score
- Number of succeeded tests compared to the number of overall tests
- Status
- Analyze start and end time; the total time for analyzing

Evaluation result string length equals the number of tests across the different FAIR components. 1 means that a particular test is passed and 0 means failed.

Each FAIR component result is calculated by dividing the earned score by the maximum score. For example, if the F component maximum score is 7 points and a particular identifier gets 5, then the F score is 71,4%. The average FAIR score is calculated by dividing earned points by the total available points. Two different statuses are available: Ready and Error. Ready means the evaluation is completed, and Error means there is some issue.

During the initial run, you must validate yourself through a web interface. This process saves Google authorization token to your local environment and enables you to use the microservice in the future. This token has a certain lifetime, so it must be renewed after some time.

The evaluation script is located in GitHub and can be accessed [here](#).

## Manual steps

To start an automated assessment, some manual steps have to be taken within the Google spreadsheet.

- Decide if you want to use DataCite metadata or not → for that write TRUE (with DataCite-metadata) or FALSE (without DataCite-metadata) in cell O2
- The columns C-N and Z are containing the data from a previous assessment. In order to start a new evaluation, these columns need to be empty.
- To start the evaluation, set a checkmark in cell O1.
- Before deleting the data from the columns make sure to save it.
- Optional: make plots that show the evolution of your data over time

