

MS25. Existing tools available for FAIRization identified and classified

Lead Partner:	CNR
Version:	V1.1
Status:	
Dissemination Level:	
Document Link:	10.5281/zenodo.5547594

Deliverable Abstract

This document, in its first version being associated to M25 aims now at identifying and keep updated a list of tools already available and consider of some help within the EOSC-Pillar project in the process to make/create FAIR data.

The list of tools presented and discussed come from a survey of several materials (online document/ deliverables from other project / web resources/ recent papers, suggestions from WP6 use-cases and community).

Classification here proposed is based on several important aspects discussed in some detail in the introduction of this document.





COPYRIGHT NOTICE



This work by Parties of the EOSC-Pillar is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). The EOSC-Pillar project is co-funded by the European Union Horizon 2020 programme under grant number 857650.

DELIVERY SLIP

	Name	Partner/Activity	Date
From:	Stefano Cozzini	CNR/T6.1	06/05/2020
Reviewed by:			
Approved by:			

DOCUMENT LOG

Issue	Date	Comment	Author
V0.1	26/04/2020	First draft	S.Cozzini
V0.2	01/05/2020	Updated version	S.Cozzini
V0.3	05/05/2020	Update version 2	S.Cozzini
V0.4	20/05/2020	Final draft	S.Cozzini
V1.0	22/06/2020	Final	S.Cozzini
V1.1	14/12/2020	Updated version	S.Cozzini,E.Osmenaj, Adeline Joffres, Benedicte Kuntziger,Marko Kulule

www.eosc-pillar.eu



TERMINOLOGY

https://eosc-portal.eu/glossary

Terminology/Acronym	Definition
CAHIER	Corpus d' Auteurs pour les Humanités: Informatisation,
	Edition, Recherche
EOSC	European Open Science Cloud
ERA	European Research Area
FAIR	Findable Accessible Interoperable Reusable
IFDS	Internet of FAIR Data and Services
IN	Implementation Network
SSH	Social Sciences and Humanities
VLRI	Very Large Research Infrastructure



Contents

1	Introduction: Identification and classification methodology	. 6
2	List of identified and classified tools	. 8
2.	1 Updated list of tools	. 9
3	Conclusions	26
App	endix I. List of resources consulted2	27



Executive summary

This deliverable aims at identifying a list of tools already available to help in the process of making/creating FAIR data.

The list of tools presented and discussed come from a survey of several materials (on-line document/ deliverables from other project / web resources).

Classification here proposed is based on several important aspects discussed in the introduction of this document.

This is an updated version of the already submitted accompanying document of milestone 25. It has been improved in several aspects thanks to the contribution of WP6 use cases. Several other suggestions have been considered to update the list of tools and the classification schema. Being the Milestone reached, now we consider this document a sort of "living document" where the entire landscape of tools supporting FAIRness is recorded and kept healthy and updated during the project lifetime. The development of such document should be considered a joint effort of WP6 use-cases with contribution of WP5 partners.



1 Introduction: Identification and classification methodology

In this introduction we briefly discuss the method we decided to adopt in the present document to firstly identify FAIRization tools available when this document is assembled and then to classify them.

It is useful, as a first step, to define the FAIRization process to better identify the tools we are interested in. A simple possible definition, adapted from GO FAIR Initiative (https://www.go-fair.org/fair-principles/fairification-process/), where actually the process is identified as FAIRification, involves several steps:

- Original data retrieval
- Data identification and analysis
- Definition and application of the semantic model
- Data transformation
- License assignment
- Metadata definition
- Deploy/ Publish FAIR Data resource (data, metadata, license)

This process is still performed manually by most of data owner/producers and this limit its scalability and can be error prone and several differences/approaches can be generated. It is therefore important to have tools that automates as much as possible such process and make it stable and reproducible across different research communities.

In this new release of the document, we significantly enlarge the list of tools and technologies that are currently available in each step of the FAIRification process identified above. We also included in the analysis tools for FAIR Data management planning, considering DMP a fundamental aspect to start an appropriate FAIRification process.

In the first release of this document, we took into consideration the following source to identify FAIRification tools:

- Already published tools on the web, found by means of web searches and/or because referred within different documents.
 An interesting starting point is this site, even if focused on life-science community: https://www.dtls.nl/fair-data/research-data-management/data-management-knowledge-tools/.
- 2. Deliverables of active projects we are aware of.
- 3. Online documents again identified and found by means of searches and/or because referred within different documents.

This list, far from been completed is now complemented by other sources indicated by WP6 community. We are therefore including new references and sources as reported in appendix 1 and we will keep doing that in further releases and updates of this document.

Feedback received by the community suggested that original classification criteria, are not the most appropriate ones being too much generic and difficult to properly evaluated for each tool. For this reason, in this release of the document we are proposing new classification criteria, taking into account the steps reported above in the FAIRification process and creating the following classes which comprises more than one macro areas:



Criterium	Macro-areas involved
1	Data Management Plan/Policy data
2	Original data retrieval and dataset identification and analysis
3	Definition of the semantic model
4	Data transformation
6	License assignment//Data Privacy GDPR
7	Metadata definition
8	Deploy/ Publish FAIR Data resource

We dedicated a specific class to the Data Management Tools: even if they are not, technically speaking FAIRification tools, nevertheless we consider them the correct starting point of all FAIRification procedures.

Classification results are then reported in a summary table at the beginning of the next session where each single tool is classified according to the above criteria. We underline the fact that in future releases of this document this classification method can be further enhanced and completed.



2 List of identified and classified tools

This section reports the tools we identified, and we considered of interest for our future work. For each of them we provide a short description, the source where we get this information and some more information.

In this new release of the document, we inserted here at the beginning the following summary table that reports all the tools we included and their classification with respect to the new classification criteria we discussed in the introduction.

We also inserted a column indicating from which WP6 uses case the suggestion come from.

Tool	Main class	Direct experience with the tool within WP6?	Suggested by UC
Data Stewardship Wizard	1	YES	UC1
DMPonline	1	NO	UC1
FAIRPort	8	NO	UC1
OpenRefine	4	NO	UC1
FAIRifier	2	Yes	UC1
FDP	8	Yes	UC1
Argos	1		UC5
Opidor	1		UC5
ChooseLicence.com	6		UC5
License Selector	6		UC5
Amnesia	4		UC5
The wizard consent form	4		UC5
d'Eldah			
The Standardization Survival Kit (SSK)	6		UC5
RDMO (Research Data Management Organizer)	1		UC 6.2/6.3
F-UJI Automated FAIR Data Assessment Tool	2		UC 6.2/6.3
IOOS compliance checker	2		UC 6.2/6.3
DMPTool	1	No	UC1
easyDMP	1	Yes	UC1
The Metadata Editor (MDE)	7	No	UC1
ORKA		No	UC1
Rightfield	3	No	UC1
FAIR Search Engine	2	No	UC1
CKAN	4	No	UC1
The Materials Data Facility (MDF)	8	No	UC1
The DataCite Metadata Store (MDS) API	7	No	UC1
ISA Software Suite	7	No	UC1
wwPDB	3	Yes	UC1
NAKALA	8	Yes	UC5



HAL reposi	itory		8		UC5
NOMAD	Repository	and	8	NO	UC1
Archive					

2.1 Updated list of tools

Name of the tool	Data Stewardship Wizard
Link	https://ds-wizard.org/about.html
Description	Data Stewards capture and combine their knowledge and expertise with respect to the specific needs of a domain or an organisation. Researchers are truly guided through composing a DMP which can be then exported using selected template and format, including machine-actionable. The benefit lies not only in having a nowadays often obligatory DMP for funders but mainly learning how to handle data correctly, make them FAIR, maintain them well during the project, and curate them long-term.
Additional Note	A starting point tool to help setup a correct data management plan
Which Class	1

Name of the tool	ARGOS
Link	https://argos.openaire.eu/splash/
Description	Argos (argos.openaire.eu) is the online machine-actionable tool developed by OpenAIRE to facilitate Research Data Management (RDM) activities concerning the implementation of Data Management Plans (DMPs). It is an open, extensible and collaborative tool which follows global



	standards as endorsed by the Research Data Alliance (RDA). Argos uses OpenAIRE guides created by its RDM Task Force to familiarize users with basic RDM concepts and guide them throughout the process of describing their data. It also utilizes the OpenAIRE pool of services and inferred sources to make DMPs more dynamic in use and easier to be completed and published. Argos is based on the OpenDMP open-source software, developed in collaboration with EUDAT CDI. ARGOS applies the RDA DMP Common Standard for interoperability which enables users to exchange DMPs with other machine-actionable platforms without losing vital information.
Additional Note	Template for Ariadne-Plus: https://ariadne-infrastructure.eu/wp-content/uploads/2020/07/D3.1%20Initial%20report%2 0n%20policies%20and%20strategies.pdf
Which class	1

Name of the tool	OPIDOR
Link	https://dmp.opidor.fr/
Description	DMP OPIDoR is based on the open source DMPRoadmap codebase, which is jointly developed by the <u>Digital Curation</u> Centre (DCC) and the <u>University of California Curation</u> Center(UC3). DMP OPIDoR source code is available on <u>GitHub</u> . DMP OPIDoR has been customised for French researchers.
Additional Note	DMP OPIDoR includes a number of templates that represent the requirements of different funders and institutions https://dmp.opidor.fr/public_templates
Which class	1



Name of the tool	DMPonline
Link	https://doi.org/10.1007/978-3-642-15464-5_74
Description	DMPonline has recently seen rapid adoption from researchers and organizations as the go-to tool to produce funder-compliant DMPs. It provides an online, collaborative environment with (mostly) open text forms divided into sections following a configurable funder's DMP template. For each section, DMPonline embeds explanatory text from a configurable set of sources, which may be DMP guidelines from funding organizations or academic institutions and may (or may not) contain FAIR-specific guidance.
Additional Note	It is not clear if the system is already available. At a first glance it seems not so friendly.
Which class	1

Name of the tool	Data FAIRport
Link	https://www.datafairport.org/
Description	The Data FAIRport is an interoperability platform that allows data owners to publish their (meta)data and allows data users to search for and access data (if licenses allow).
Additional Note	It Involves different tools touching different aspects of FAIR process. For more information, please check the web page of the project.
Which class	8



Name of the tool	OpenRefine
Link	http://openrefine.org/
Description	OpenRefine is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data. OpenRefine always keeps your data private on your own computer until YOU want to share or collaborate. Your private data never leaves your computer unless you want it to. (It works by running a small server on your computer and you use your web browser to interact with it)
Additional Note	Born as a Google product, in October 2012, it was renamed OpenRefine as it transitioned to a community-supported product. It is indicated more for people which do not have programming skills.
Which class	4

Name of the tool	FAIRifier
Link	Not available
Description	This tool enables a post-hoc FAIRification workflow: load an existing data set (from a wide range of formats), (optionally) perform data wrangling tasks, add FAIR metadata) attributes to the data, generate a linked data version of the data and, finally, push the result to an online FAIR data infrastructure to make it accessible and discoverable.
Additional Note	It is not clear the level of maturity: it is based on other tools and still in development. Not easy to use for non-experienced users.



Which class	2
-------------	---

Name of the tool	FAIR Data Point (FDP)
Link	https://github.com/FAIRDataTeam/FAIRDataPoint-Spec
Description	The FAIR Data Point's metadata layered approach is composed of five layers, namely, the FAIR Data Point itself, the collection of datasets, each one of the offered datasets, the dataset's distribution and the structure and semantics of each of dataset.
Additional Note	Not easy to be used for well-structured communities
Which class	8

Name of the tool	RDMO (Research Data Management Organizer)
Link	https://rdmorganiser.github.io/en/
Description	RDMO is a web-based software which allows scientists researchers and data management staff to easily compile data management plans (DMPs) to either satisfy funder requirements or for own use in organizing research projects. RDMO is meant to be installed and run as decentralized local instance at any institution (open source, open access) wishing to offer it as a service. DMPs can be configured for any scientific discipline and the look and feel can be freely adopted by the hosting institution. Documenting the FAIRness of research data in the DMP is part of the standard RDMO question set.



	DMPs compiled in RDMO are meant to be treated as real "living documents" – snapshots can be exported and saved at any time during a project. Further, collaboration on DMPs is also possible for a research team, which is useful for larger projects. Please see the documentation for the full suite of RDMO capabilities.
Additional Note	RDMO is maintained by a dedicated team of roughly 5 developers. The developers receive input from the rather large RDMO user community. New versions of RDMO are released on a quasi-monthly basis. New functionalities are constantly added. The long-term perspective of RDMO maintenance is a main goal of the RDMO community. By now, RDMO runs operationally at 16 institutions across Germany and 24 more have a test instance running: https://rdmorganiser.github.io/en/cooperations/ Configuring the RDMO question catalogs and the output formats is manageable with overseeable overhead – even for non-IT experts.
Which class	1

Name of the tool	IOOS compliance checker
Link	https://github.com/ioos/compliance-checker or https://compliance.ioos.us/about
Description	The IOOS Compliance Checker is a python based tool for data providers to check for completeness and community standard compliance of local or remote netCDF files against CF and ACDD file standards. The python module can be used as a command-line tool or as a library that can be integrated into other software.
Additional Note	The IOOS compliance checker is maintained by several developers and regularly updated.



	A good tool to write plugins for the IOOS compliance checker is the compliance-check-lib (https://github.com/cedadev/compliance-check-lib).
Which class	2

Name of the tool	F-UJI Automated FAIR Data Assessment Tool
Link	https://github.com/pangaea-data-publisher/fuji or https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool
Description	F-UJI is a web service to programatically assess FAIRness of research data objects based on metrics developed by the FAIRsFAIR project. In total the tool checks for the compliance to 17 metrics including for example a check wether the data is assigned a globally unique identifier or a persistent identifier.
Additional Note	The source code I well maintained and continuously updated. Datasets from five CoreTrustSeal certified repositories have been tested with the automated FAIR assessment tool (F-UJI).
Which class	2

Name of the tool	License Selector
Link	http://ufal.github.io/public-license-selector/
Description	The license selector tool allows a user to attach a specific license to their data set or software package in order to protect it from certain usage or abuse. This document is targeted at all users that want to add a license to their data set, but don't know which license suits best to their data set distribution and rights requirements. The selector tool is an addition to the B2SHARE service.



	The license selector tool is provided to ease the selection of the correct license to attach to your dataset or software package without requiring expert knowledge of every available license. By answering a few questions in a wizard-like interface, a user will be guided to a single or set of licenses which can be selected for inclusion in their metadata upon publishing. The tool is by no means mandatory to use, but purely aimed as a support for the publishing user.
Additional Note	The license selector tool developed by a team led by Pawel Kamocki (researcher IT Law at IDS Mannheim, Germany), made available as open-source software: https://github.com/ufal/public-license-selector
Which class	3

Name of the tool	Choose a license
Link	Choose an open source license
Description	A web site that guides you in choosing the correct open- source license for your software.
Additional Note	None
Which class	3

Name of the tool	Amnesia



Link	https://amnesia.openaire.eu/index.html#features
Description	Amnesia is an application written in java and JavaScript and should be used locally for anonymizing a dataset. The basic idea behind anonymization is that a file containing personal data is loaded to Amnesia, we refer to this as the original dataset, and Amnesia transforms it to an anonymous dataset, which can then be stored locally. The transformation is guided by user choices and provides an anonymization guaranty for the result.
Additional Note	None
Which class	4

Name of the tool	DARIAH ELDAH Consent Form Wizard
Link	Welcome to the DARIAH ELDAH Consent Form Wizard (CFW)! CFW
Description	The aim of the CFW is to support humanities researchers within the European Union in obtaining valid consent for data processing in the context of their specific professional activity.
Additional Note	None
Which class	6

Name of the tool	Standardization Survival Kit (SSK)
Link	http://ssk.huma-num.fr/#/
Description	The Standardization Survival Kit (SSK) is designed to support researchers in selecting and using the appropriate standards for their particular disciplines and workflows.



Additional Note	None
Which class	6

Name of the tool	DMPTool
Link	https://dmptool.org/
Description	The DMPTool is a free, open-source, online application that helps researchers create data management plans. The DMPTool provides a click-through wizard for creating a DMP that complies with funder requirements. It also has direct links to funder websites, help text for answering questions, and resources for best practices surrounding data management.
Additional note	management.
Which class	1

Name of the tool	easyDMP
Link	https://easydmp.eudat.eu/



Description	The goal of easyDMP is to provide a tool that enables researchers with minimal experience in data management a simple way of creating a data management plan. This is achieved by transforming the funding agency's data management guidelines into a series of easy to answer questions, many containing canned answers. The resulting plan can be used as a blueprint for researchers to put in place the necessary elements that ensure their data are adequately managed. EasyDMP has been developed by Sigma2 in collaboration with EUDAT2020. EasyDMP is free of charge and accessible to every researcher in Norway and in Europe.
Additional note	The electronic DMP created is machine readable
Which class	1

Name of the tool	The Metadata Editor (MDE)
Link	https://docs.google.com/document/d/1eBJKSg1u5gep6- 2cbjGpWNulMvC7F9BSmnKYq7FFSGA/edit
Description	Is a software tool that makes it easy for non-technical users to define and publish the metadata required by a FAIR Data Point (FDP). The MDE helps you create the metadata for the FDP. Its web interface allows you to fill or edit a simple form. As you fill in the form, the MDE will build and display the RDF representation of your metadata.
Additional note	None
Which class	7



Name of the tool	Reusable Knowledge graph Annotator (ORKA)
Link	https://dtl- fair.atlassian.net/wiki/spaces/TA/pages/1409029/ORKA+an d+NanoPub+Store+Software+Specification#ORKAandNano PubStoreSoftwareSpecification-SystemArchitecture
Description	ORKA supports easy human curation of knowledge graphs by offering graph annotation as a service and capturing the provenance of the annotator and the original statement. ORKA is responsible for providing support to users who want to annotate a triple originated from a graph analysis tool by changing the triple's predicate. The edited triple is then transformed into a nanopublication adding the provenance of the author and keeping the relationship with the original triple.
Additional note	None
Which class	7

Name of the tool	Rightfield
Link	https://rightfield.org.uk/
Description	Rightfield is an open-source tool for adding ontology term selection to Excel spreadsheets. Rightfield is used by a 'Template Creator' to create semantically aware Excel spreadsheet templates. The Excel templates are then reused by Scientists to collect and annotate their data; without any need to understand, or even be aware of, Rightfield or the



	ontologies used. Rightfield embedded templates are used within the <u>Samples</u> framework of the <u>SEEK</u> .
Additional note	For the scientist, the main advantages of RightField are that it enables them to consistently annotate their data without the need to explore and understand the numerous standards and ontologies available to them, and it does not require them to change normal practice. Everything is embedded in the Excel spreadsheet.
Which class	3

Name of the tool	FairSearch
Link	??
Description	The FAIR Data Search Engine harvests the metadata available on FAIR Data Points or compatible data repositories, indexes them, and provides a search interface.
Additional note	None
Which class	7



Name of the tool	CKAN
Link	https://ckan.org/
Description	CKAN is a powerful data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data.
Additional note	CKAN is built with Python on the backend and Javascript on the frontend, and uses The Pylons web framework and SQLAlchemy as its ORM. Its database engine is PostgreSQL and its search is powered by SOLR. It has a modular architecture that allows extensions to be developed to provide additional features such as harvesting or data upload.
Which class	4

Name of the tool	The Materials Data Facility (MDF)
Link	https://materialsdatafacility.org/
Description	The Materials Data Facility (MDF) is a service that allows to publish, discover, and access materials datasets. After publishing the dataset, you receive a permanent identifier (e.g., DOI).
Additional note	None
Which class	8



Name of the tool	The DataCite Metadata Store (MDS) API
Link	https://support.datacite.org/docs/mds-api-guide
Description	The <u>DataCite Metadata Store (MDS) API</u> allows users to register DataCite DOIs and associated metadata. The API requires authentication. To retrieve DOI metadata records users should use our <u>DataCite REST API</u> .
Additional note	All requests to the MDS API require authentication. For this reason, only traffic via a secure connection (HTTPS) is supported. The DataCite Metadata Store (MDS) uses Basic authentication">https://example.com/html/> Basic authentication .
Which class	7

Name of the tool	ISA Software Suite
Link	https://isa-tools.org/
Description	The open-source ISA metadata tracking tools facilitate ISA-TAB-compliant collection, curation, local management and reuse of datasets in an increasingly diverse set of life science domains.
Additional note	None
Which class?	7

Name of the tool	wwPDB

MS25. Existing tools available for FAIRization identified and classified



Link	http://www.wwpdb.org/
Description	Sustain freely accessible, interoperating Core Archives of structure data and metadata for biological macromolecules as an enduring public good to promote basic and applied research and education across the sciences. It provides expert deposition, validation, bio-curation, and remediation services at no charge to Data Depositors worldwide.
Additional note	
Which class	3

Name of the tool	Nakala
Link	https://www.nakala.fr/index.html.en
Description	NAKALA is a Huma-Num service which enables research teams to deposit all types of documented digital data (text files, sounds, images, videos, 3D objects, etc.) in a secure repository in order to share them.
Additional note	Used within UC5
Which class	8

Name of the tool	HAL



Link	Home - Archive ouverte HAL (archives-ouvertes.fr)
Description	HAL is an open archive where authors can deposit scholarly documents from all academic fields.
Additional note	Used in Use case 5
Which class	8

Name of the tool	NOMAD Repository and Archive
Link	https://nomad-lab.eu/prod/rae/gui/search
Description	NOMAD Repository and Archive provide search, access, and download of all NOMAD data in its raw (Repository) and processed (Archive) form. It makes possible uploading and managing raw materials science data. Published data can accessed without an account.
Additional note	Of some interest for UC1
Which class	8



3 Conclusions

We searched among several sources to identify tools in order to provide an overview of what is available and avoid re-inventing the wheel in the activities within the WP6 uses cases. In particular, we noticed a problem of fragmentation, in sense that tools are scattered and was really difficult to found them.

The revision of the document and the feedback we received still confirm the impression, already reported that does not yet exist a mature ecosystem of tools, platforms and standards to support human and machine agents to manage, produce, publish and consume FAIR data in a user-friendly and efficient (i.e., "easy") way.

An update version of this document will be released when additional tools and methods are suggested and considered useful for some WP6 FAIRification activities.



Appendix I. List of resources consulted

In this appendix we list the most important resources used in the compilation of this milestone: The list will be updated in new releases of the present document.

- https://www.dtls.nl/fair-data/find-fair-data-tools/
- https://www.rd-alliance.org/system/files/Results%20Analysis%20of%20FAIR%20assessment%20tools%20v3.pdf
- M. Thompson, K. Burger, R. Kaliyaperumal, M. Roos & L.O. Bonino da Silva Santos. Making FAIR easy with FAIR tools: From creolization to convergence. Data Intelligence 2(2020), 87 - 95. doi: 10.1162/dint_a_00031
- www.openaire.eu
- http://www.parthenos-project.eu
- www.dariah.eu
- https://ariadne-infrastructure.eu
- https://www.fairsfair.eu/
- public-EOSC-SYNERGY-WP3-D3.3-FINAL.pdf (eosc-synergy.eu)
- https://rightfield.org.uk/about.html
- https://fair-dom.org/platform/
- https://www.dcc.ac.uk/guidance/standards/metadata/tools
- https://easydmp.eudat.eu/
- https://dmptool.org/
- https://docs.google.com/document/d/1eBJKSg1u5gep6-2cbjGpWNulMvC7F9BSmnKYq7FFSGA/edit
- https://ckan.org/
- https://materialsdatafacility.org/
- https://json-ld.org/
- http://www.wwpdb.org/
- https://nomad-lab.eu/prod/rae/gui/search