

IMI2 Project 802750 - FAIRplus
FAIRification of IMI and EFPIA data

WP3 – Identification of project data sources for FAIRification

D3.2 IMI FAIR Metrics Publication

Lead contributor	Tony Burdett (1 – EMBL (EBI)) and Fuqi Xu (1 – EMBL (EBI)) tburdett@ebi.ac.uk , fuqi@ebi.ac.uk
Other contributors	Mélanie Courtot (1 – EMBL (EBI)) Nick Juty (4 – University of Manchester)

Due date	31 Dec 2020
Delivery date	08 Jan 2021
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V1.0	08 Jan 2021

Document History

Version	Date	Description
V0.1	14 Dec 2020	First Draft
V0.2	18 Dec 2020	Comments

V0.3	22 Dec 2020	Final review
v1.0	08 Jan 2021	Final version

Table of Contents

Document History	1
Executive Summary	3
Background	3
Methods	3
Indicators	3
RDA data maturity indicators	3
FAIRplus indicators	4
Assessment process	6
Automated assessment	6
Manual assessment	6
FAIR assessments results	7
eTOX	7
ND4BB	7
ReSOLUTE	7
OncoTrack	7
EBiSC and EBiSC II	8
IMIDIA	8
RHAPSODY	8
Discussion	8
Dependencies between Assessment Process and Indicators	8
Automated vs manual assessment	8
FAIR enough for specific competency questions	9
Conclusion	10
Appendix A - Comparison between the FAIR evaluator indicators and the RDA indicators	11

1. Executive Summary

FAIRification of datasets is an ongoing area of research by different consortia. In the context of FAIRplus, we reviewed and selected from existing metrics, which were then applied to pilot datasets before and after our fairification process to evaluate its effectiveness. Based on the results in this report and our experiences of the process, we recommend a combination of manual and automated assessments against strong use cases and competency questions. This supports the attainment of the right level of FAIR, which we call “FAIR enough”. Reaching the FAIR enough level ensures we achieve the right balance of providing maximal returns at minimum costs for data owners. The metrics used by FAIRplus have been published on our website and the process for evaluation using these metrics is described in this report.

2. Background

Measuring the effectiveness of the FAIRification process, and validating its advantages and improvements to datasets provided by IMI and EFPIA partners, requires a set of shared metrics to evaluate FAIRplus outputs quantitatively and qualitatively. To this aim, FAIRplus work package 2 is working to provide a set of FAIR indicators. These FAIRplus indicators are derived from reviewing and contributing to existing efforts in the domain, as well as adding and expanding them as required through Task 2.3.1. This report describes which indicators have been used to date (see section 3.1), how the assessments have been performed (3.2), current results on pilot datasets (4) and next steps (5).

3. Methods

This section describes the different types of FAIR indicators that were defined by T2.3.1. We describe an assessment process to assess the FAIRness of IMI projects datasets before and after FAIRification using those indicators, including both automated and manual components. The current version of the FAIRplus indicators does not support automated assessments and therefore we do not describe an application offering a complete automated solution for FAIR evaluation.

3.1. Indicators

When reviewing existing FAIR indicators and tools, we selected different sets of indicators provided by different groups. We concentrated on the ability to reuse existing suitable metrics, provide feedback to the community to evolve the indicators, and fill in gaps where we could not find appropriate metrics for our own evaluation. Two different sets of indicators have been selected by WP2 and tested by WP3 against pilot datasets.

RDA data maturity indicators

To provide a more comprehensive assessment of the FAIR level of the project, the Research

Data Alliance (RDA) data maturity indicators¹ were used. The RDA indicators are under active development by the Alliance working groups, and the indicators used evolved significantly during the FAIRplus FAIRification process. Feedback was provided to the RDA community and contributed to the evolution of the RDA indicators. Versions v0.02² and v0.03³ have been used by FAIRplus.

FAIR data maturity model Indicators

Version 0.03

Mandatory: indicator MUST be satisfied for FAIRness
 Recommended: indicator SHOULD be satisfied, if at all possible, to increase FAIRness
 Optional: indicator MAY be satisfied, but not necessarily so
<https://www.ietf.org/rfc/rfc2119.txt>

	PRINCIPLE	INDICATOR_ID	INDICATORS	PRIORITY
F	F1	F1-01M	Metadata is identified by a persistent identifier	Mandatory
	F1	F1-01D	Data is identified by a persistent identifier	Mandatory
	F1	F1-02M	Metadata is identified by a universally unique identifier	Mandatory
	F1	F1-02D	Data is identified by a universally unique identifier	Mandatory
	F2	F2-01M	Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard	Recommended
	F3	F3-01M	Metadata includes the identifier for the data	Recommended
	F4	F4-01M	Metadata is offered/published/exposed in such a way that it can be harvested and indexed	Mandatory
A	A1	A1-01M	Metadata includes information about access conditions	Mandatory
	A1	A1-01D	Data can be accessed manually (i.e. with human intervention)	Mandatory
	A1	A1-02D	Data can be accessed automatically (i.e. by a computer program)	Recommended
	A1	A1-02M	Metadata identifier resolves to a metadata record	Mandatory
	A1	A1-03D	Data identifier resolves to a digital object	Mandatory
	A1	A1-03M	Metadata is accessed through standardised protocol	Recommended
	A1	A1-04D	Data is accessible through standardised protocol	Recommended
	A1.1	A1.1-01M	Metadata is accessible through a free access protocol	Mandatory
	A1.1	A1.1-01D	Data is accessible through a free access protocol	Recommended
	A1.1	A1.1-02M	Metadata is accessible through an open-source access protocol	Recommended
	A1.1	A1.1-02D	Data is accessible through an open-source access protocol	Recommended
	A1.1	A1.1-03D	Actions to be taken by a reuser to get access to the data are well documented	Recommended
	A1.2	A1.2-01M	Metadata includes information relevant for access control	Mandatory
	A1.2	A1.2-01D	Data is accessible through an access protocol that supports authentication	Optional
	A1.2	A1.2-02D	Data is accessible through an access protocol that supports authorisation	Optional

Figure 1: An example selection of the RDA FAIR indicators (v0.3).

FAIRplus indicators

FAIRplus indicators are derived from and aligned to the set of RDA data maturity indicators, generated by community agreement. The details necessary to run the domain-specific parts of the RDA metrics are delegated to be refined by the specific community of use, through the application of relevant community standards (e.g. minimum information checklists, metadata standards, etc.). For the FAIRplus project, the most relevant indicators are around infrastructure and process maturity; these were selected for use and aligned with the FAIRplus Dataset Maturity Model Framework. The first version of the FAIRplus Dataset Maturity Model Framework includes four primary data usage areas⁴ and 18 FAIRplus indicators⁵ (Figure 2), which are designed for measuring dataset compliance to Data Usage Areas as shown in Figure 3. One indicator can support more than one Data Usage Area. Indicators are grouped according to the ISA framework. Additional details of the development

¹ <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>

²

<https://docs.google.com/spreadsheets/d/1mkjEIFrTBPBH0QViODexNur0xNGhJqau0zkL4w8RRAw/edit#gid=1325892715>

³

<https://docs.google.com/spreadsheets/d/1mkjEIFrTBPBH0QViODexNur0xNGhJqau0zkL4w8RRAw/edit#gid=1558529297>

⁴ <https://fairplus.github.io/fairification-results/2020-10-11-FAIRplus-data-usage-areas-v0.1/>

⁵ <https://fairplus.github.io/fairification-results/2020-10-11-FAIRplus-indicators-v0.1/>

of the FAIRplus indicators can be found as part of Milestone 2.3 (M24).

ID	Indicators
F+S01	Study level documentation is available in a human readable format.
F+S02	Data is reported by following community specific minimum information guidelines
F+S03	Metadata documents and provides references about all data biological data types and formats in data is expressed.
F+S04	Relationships between different data sets in a study is well defined.
F+S05	A versioning policy is applied to uniquely identify a particular form of a dataset from an earlier form or other forms of itself.
F+S06	Share not only derived and publication related data but data generated in early phases of research data workflow such as primary data and analyzed data.
F+S07	Negative results are shared.
F+S08	The study is described with metadata including context, samples and data acquisition, methods for analyzing and processing data, quality control, and restriction for reuse.
F+S08a	Metadata includes information about the study design, protocols and data collection methods.
F+S08b	Metadata includes explicit references to research resources such as samples, cell lines
F+S08c	Metadata contains information about data processing methods, data analysis and quality assurance metrics.
F+S08d	Metadata includes information about data ownership, license and reuse constraints for sensitive data.
F+A01	Data is organized and documented in a human understandable way
F+A02	Data is encoded in a community specific exchange standard.
F+A03	A machine and human readable formal description of the structure of data is available including types, properties.
F+A04	Data is structured by following a life sciences domain model, core classes and their semantic relations refers to a common data model.
F+A05	Data is described with terminology standards.
F+A06	Core data classes (important data elements) follows a common master and reference data entity.

Figure 2: FAIRplus indicators (V0.1)

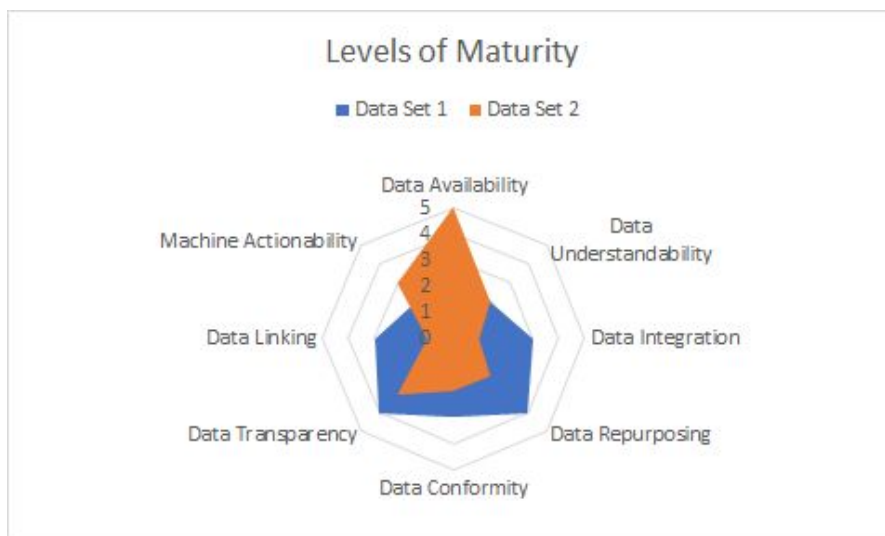


Figure 3. Example output from FAIRplus assessment, highlighting the FAIRness improvement of each dataset against axes of maturity.

3.2. Assessment process

Automated assessment

The FAIR evaluator⁶ online tool aims at providing a fully automated assessment of FAIRness using “Maturity Indicators” (MIs). Each MI corresponds to an aspect of the FAIRification level that can be encoded in a specific automated test, which returns an evaluation outcome against this specific MI. For example, an automated test inspects whether the dataset being evaluated has information about the license it is distributed under as shown in Figure 4. The FAIRvaluator repository⁷ provides the full list of the 30 MIs for automated FAIR assessment⁸. 22 of those indicators have been converted into automated tests⁹. Appendix A shows the mapping between the FAIR Evaluator maturity indicators and the FAIRplus indicators.

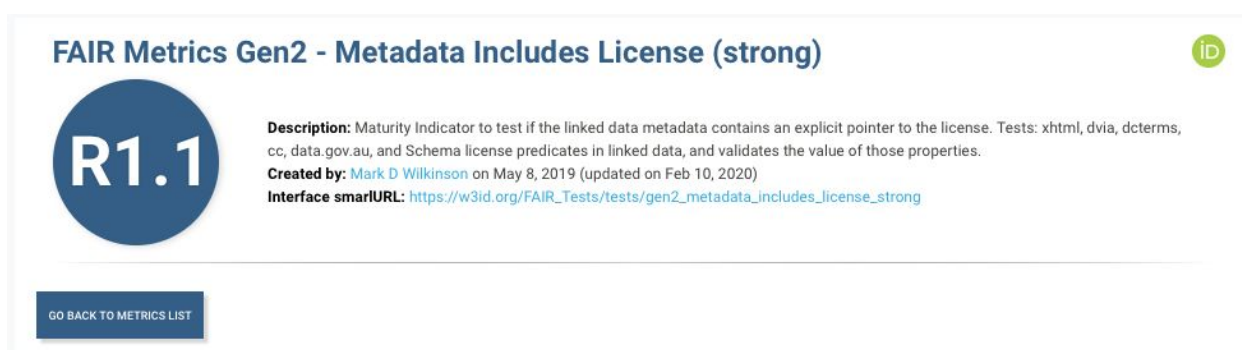


Figure 4. FAIR evaluator test “FAIR Metrics Gen2 - Metadata Includes License (strong)” from <https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/metrics/21>.

Manual assessment

Manual assessments were performed by assigning a compliance value for each indicator. For the RDA indicators, true or false values for compliance were assigned. An overall FAIR score was generated by calculating the percentage of compliant indicators. The RDA indicators were used to assess the four pilot projects (ND4BB, eTOX, OncoTrack, ReSOLUTE). For the FAIRplus indicators, a compliance value of *Fully Complies*, *Partially Complies* or *No compliance* was assigned. The assessment results per indicator are mapped to data usage areas and summarized to reflect the level of maturity in different data usage areas. The FAIRplus indicators have been used to assess the FAIR levels of EBiSC I and II, IMIDIA and Rhapsody and will be applied to upcoming datasets. An overall FAIR score is not calculated using the FAIRplus indicators.

To improve the reproducibility and reliability of manual assessment, each dataset was evaluated by an assessment team made up of project personnel with different backgrounds. This team included experts with knowledge of the FAIR indicators, experts with knowledge of the dataset, and FAIRification technical experts. The assessment team discussed the score of each indicator to produce the final assessment results. The detailed assessment process

⁶ <https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/>

⁷ <https://github.com/FAIRMetrics/Metrics/tree/master/MaturityIndicators>

⁸ https://fairsharing.org/standards/?q=&selected_facets=type_exact:metric

⁹ <https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/collections/6>

varies depending on the availability of the IMI project owners and squad members. For example, in the eTox FAIR assessment, members of the assessment team performed individual assessments and met following those individual assessments to a) discuss the meaning of the question (as some very viewed as being subjective), b) discuss whether the scoring was correct based on the collective understanding of the metric, c) assign an agreed score for each metric. In contrast, for the assessments of IMIDIA and Rhapsody datasets, two members of the assessment team discussed the results for each indicator together before the data owner was invited to validate their assessment results.

4. FAIR assessments results

FAIRification has been completed, and pre- and post-FAIRification assessments are available for the following four datasets:

eTOX

[Initial FAIR assessment results](#)¹⁰ for eTOX.

[FAIR assessment results](#)¹¹ for eTOX after FAIRplus FAIRification processes.

ND4BB

[Initial FAIR assessment results](#)¹² for ND4BB.

[FAIR assessment results](#)¹³ for ND4BB after FAIRplus FAIRification processes.

ReSOLUTE

[Initial FAIR assessment results](#)¹⁴ for ReSOLUTE.

[FAIR assessment results](#)¹⁵ for ReSOLUTE after FAIRplus FAIRification processes.

OncoTrack

[Initial FAIR assessment results](#)¹⁶ for OncoTrack.

[FAIR assessment results](#)¹⁷ for OncoTrack after FAIRplus FAIRification processes.

For three datasets, FAIRification is ongoing, and only pre-FAIRification assessment is available. Post-FAIRification assessment results will be provided once FAIRplus FAIRification is complete.

¹⁰ <https://fairplus.github.io/fairification-results/2019-12-17-eTOX-pre-assessment/>

¹¹ <https://fairplus.github.io/fairification-results/2019-12-17-eTOX-post-assessment/>

¹² <https://fairplus.github.io/fairification-results/2019-12-17-ND4BB-Pre-assessment/>

¹³ <https://fairplus.github.io/fairification-results/2019-12-17-ND4BB-post-assessment/>

¹⁴ <https://fairplus.github.io/fairification-results/2019-12-17-RESOLUTE-Pre-assessment/>

¹⁵ <https://fairplus.github.io/fairification-results/2019-12-17-RESOLUTE-post-assessment/>

¹⁶ <https://fairplus.github.io/fairification-results/2019-12-17-OncoTrack-Pre-assessment/>

¹⁷ <https://fairplus.github.io/fairification-results/2019-12-17-OncoTrack-post-assessment/>

EBiSC and EBiSC II

[Initial FAIR assessment results for EBiSC and EBiSC II](#)¹⁸

IMIDIA

[Initial FAIR assessment results for IMIDIA](#)¹⁹

RHAPSODY

[Initial FAIR assessment results for RHAPSODY](#)²⁰

5. Discussion

5.1. Dependencies between Assessment Process and Indicators

The RDA indicators are very comprehensive and - barring ambiguity in understanding their definitions - very easy to score against. Each indicator is scored true or false, which supports the calculation of an overall FAIRification score, as shown in the results for the four pilot projects. However, this evaluation is done entirely manually by groups of experts as described in the assessment section 4. All indicators have the same 'rank' - they represent a flat list of true or false elements, and changes to these indicators do not necessarily reflect actual FAIRness improvements of the dataset. We see two main limitations of the RDA indicators: firstly, that definitions are subjective and difficult to automate, and secondly that they do not well reflect FAIRness improvements. The FAIRplus metrics group (T.2.3.1) have designed a set of FAIRplus indicators to move beyond these limitations, but this work is not yet complete and introduces its own set of limitations, discussed further below. As FAIRplus indicators stabilise and move towards increased clarity and objectivity, we expect the possibilities for producing applications for automated FAIR assessment applications will improve.

5.2. Automated vs manual assessment

Trialling different manual assessment approaches (3.2.) allows us to derive general guidelines for best evaluation:

1. **Manual assessment should be performed by more than one assessor.** Establishing inter evaluator agreement and scores mitigates the subjectivity of the assessment results.
2. **A blend of expertise is required for the assessment process.** Data owners or squad members that are familiar with the dataset are needed to provide sufficient information about its content and characteristics. Experts familiar with the FAIR indicators are required to clarify the definition of indicators. Finally, FAIRification experts are required to support the consistent use of the indicators across datasets.

¹⁸ <https://fairplus.github.io/fairification-results/2020-12-13-EBiSC-pre-assessment/>

¹⁹ <https://fairplus.github.io/fairification-results/2020-12-13-IMIDIA-pre-assessment/>

²⁰ <https://fairplus.github.io/fairification-results/2020-12-13-Rhapsody-pre-assessment/>

The FAIR evaluator is attractive thanks to its fully automated assessment capabilities, but may not be sufficient as a solution for FAIRplus. Some aspects of FAIR assessments will be hard, if not impossible, to fully automate due to the nature of some of the FAIR indicators. For example, RDA indicator R1-01M (v0.03) states: “sufficient metadata is provided to allow reuse, following domain/discipline-specific metadata standard”. Whether enough metadata has been provided, and its reusability, is impossible to quantify in the absolute: this requires expert domain knowledge as well as specific use cases to validate “reuse” against. Consequently, these types of evaluation remain manual, and we anticipate a small proportion of all assessments will always require manual review. We suggest a multi-layered approach to the construction of indicators - those that can provide clear, objective assessments through automation (but may not be well tailored to specific scenarios), and those that provide an additional, customised layer for assessment of domain- or use case-specific scenarios (but that are not recommended for automated use).

For automated tests, manual review is still currently recommended; as described in 3.2.1, automated tests may have dependencies which, if absent, do not negatively impact the level of FAIR. This can lead to inaccurate results that may present an artificially low judgement on the level of FAIR. For example, the test of FAIR evaluator indicator FAIR “Unique identifier”²¹ extracts GUIDs using regular expression patterns and compares the extracted GUIDs against predefined unique identifier schemas on FAIRsharing²². Resources that contain unique identifiers which are not registered in FAIRsharing would fail this assessment, but this would be a false negative. Automated tests clearly require a canonical source of identifier patterns for this type of check to be possible and we recognise the challenges with creating a fully comprehensive resource. FAIRsharing provides the best opportunity that is currently available to deliver this resource, but still requires greater use, uptake and feedback from consumer applications like the FAIR evaluator before it can be relied on for complete automation without false negatives. Nonetheless, automated evaluation improves the reproducibility of the assessment and is the only way to scale evaluation up to hundreds if not thousands of datasets.

Finally, the FAIR evaluator requires the datasets to be available publicly via a FAIRsharing DOI to be assessed. This is not always possible for IMI datasets. For example, IMIDIA collects sensitive clinical metadata which needs to remain controlled-access and therefore cannot be assessed without first performing a data access request and sharing access credentials with the FAIR evaluator.

Despite these limitations, we recognise that the FAIR evaluator provides an excellent service. It is the FAIRplus choice for performing appropriate automated assessments, but we recognise that more guidance is needed over what assessments should be manual, what can be automated, and how to automate suitable assessments. A recipe²³ has been developed to provide guidance on using the FAIR evaluation service, and we hope to incorporate automated checks for FAIRplus indicators into the FAIR evaluator in future.

5.3. FAIR enough for specific competency questions

FAIR scores are calculated in both the FAIR evaluator assessments and the RDA indicator FAIR

²¹ <https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/metrics/1>

²² https://fairsharing.org/standards/?q=&selected_facets=type_exact:identifier%20schema

²³ <https://fairplus.github.io/cookbook-dev/recipes/assessing-fairness/fair-assessment-recipe.html>

assessments. Having a score helps to evaluate and validate the FAIRification technical solutions (D3.3) of a specific dataset. For example, the FAIRification on the ReSOLUTE project improves the general FAIR score from 58.7% before FAIRification to 82% after FAIRification. However, the FAIR score alone is not enough to compare FAIRness across different datasets: complying with different indicators affect the dataset FAIRness differently depending on the use case considered. Consequently, we aim to make the dataset FAIR enough for specific competency questions rather than simply achieving a higher FAIR score. For example, EBISC aims to streamline deposition of cell line data in the EBISC cell line catalogue and improve the metadata quality to support their discovery by users. Currently, answering positively to either RDA indicator A1.1-01D (v0.03): *“Metadata is accessible through a free access protocol”* or RDA indicator R1.3-0.1M(v0.03): *“Metadata complies with a community standard”* increments equally the FAIR score. However, making the dataset compliant to R1.3-0.1M has a more practical impact on the data reusability and findability for EBISC, as using community standards and FAIR vocabularies directly enables better search by users.

6. Conclusion

Performing an automated assessment, in addition to manual review and further use-case based evaluation of the dataset FAIRness will provide the most scalable result. The rate of false negatives can be decreased by improving consistency in standards used to represent types of information - such as a limited set of predicates to represent licensing information. Deploying automated evaluation over more datasets will also highlight current gaps and should help alleviate some of the limitations in the early, current version of the tool. Work is ongoing to supplement atomic indicators with new capabilities and data usage areas, which will help weigh the indicators according to specific use cases and competency questions.

Scalable automated evaluation coupled to manual review and assessments will ensure data sets are “FAIR-enough” against users requirements, and that the cost benefit of the fairification process can be maximised.

Appendix A - Comparison between the FAIR evaluator indicators and the RDA indicators

19 out of 41 RDA indicators are similar to the FAIR evaluator maturity indicators, and easy to automate.

FAIR evaluator Maturity		RDA FAIR Indicator		Similarity
Indicator ID	Indicator	Indicator ID	Indicator	
Indicator_1	Unique identifier	RDA-F1-01M	Metadata is identified by a persistent identifier	Similar
Indicator_3	data identifier persistence	RDA-F1-01D	Data is identified by a persistent identifier	Same
Indicator_1	Unique identifier	RDA-F1-02M	Metadata is identified by a globally unique identifier	Same
Indicator_1	Unique identifier	RDA-F1-02D	Data is identified by a globally unique identifier	Same
Indicator_6	data identifier explicit in metadata	RDA-F3-01M	Metadata includes the identifier for the data	Similar
Indicator_8	searchable in major search engine	RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	Similar
Indicator_10	uses open free protocol for metadata retrieval	RDA-A1-04M	Metadata is accessed through standardised protocol	Similar
Indicator_9	uses open free protocol for data retrieval	RDA-A1-04D	Data is accessible through standardised protocol	Similar
Indicator_10	uses open free protocol for metadata retrieval	RDA-A1.1-01M	Metadata is accessible through a free access protocol	Similar
Indicator_9	uses open free protocol for data retrieval	RDA-A1.1-01D	Data is accessible through a free access protocol	Similar
Indicator_11	data authentication and authorization	RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation	Same
Indicator_14,15	metata knowledge representation language (strong)	RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	Similar
Indicator_16,17	data knowledge representation language (strong)	RDA-I1-01D	Data uses knowledge representation expressed in standardised format	Similar
Indicator_18,19	metadata uses FAIR vocabularies (strong)	RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	Same
Indicator_20	metadata contains qualified outward references	RDA-I3-01M	Metadata includes references to other metadata	Similar
Indicator_20	metadata contains qualified outward references	RDA-I3-03M	Metadata includes qualified references to other metadata	Similar
Indicator_21,22	metadata includes license(strong)	RDA-R1.1-01M	Metadata includes information about the licence under which the	Similar

			data can be reused	
Indicator_2 1,22	metadata includes license(strong)	RDA-R1.1-02 M	Metadata refers to a standard reuse licence	Similar
Indicator_2 1,22	metadata includes license(strong)	RDA-R1.1-03 M	Metadata refers to a machine-understandable reuse licence	Similar