Do I-PASS for FAIR?
Measuring the FAIR-ness of Research Organizations

J. RINGERSMA (ID)

M. MIEDEMA (ID)

Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

Given the increased use of the FAIR acronym as adjective for other contexts than data or data sets, the Dutch National Coordination Point for Research Data Management initiated a Task Group to work out the concept of a FAIR research organization. The results of this Task Groups are a definition of a FAIR enabling organization and a method to measure the FAIR-ness of a research organization (The Do-I-PASS for FAIR method). The method can also aid in developing FAIR-enabling Road Maps for individual research institutions and at a national level. This practice paper describes the development of the method and provides a couple of use cases for the application of the method in daily research data management practices in research organizations.

CORRESPONDING AUTHOR:
J. Ringersma

Wageningen Data Competence Center, Wageningen University & Research, NL

jacquelijn.ringersma@wur.nl

INTRODUCTION

"Our research institute will be FAIR within five years"

The acronym FAIR encompasses 15 principles which equip digital objects with the properties required to make research data both reproducible and machine actionable. The acronym and principles were first introduced in 2016 [Wilkinson et al, 2016] and have been embraced by both researchers and research support staff hoping to advance data-driven science.

Over the last few years FAIR has been used as an adjective in many other contexts than digital data, and increasingly so: FAIR data stewardship, FAIR infrastructure, FAIR Research Software, etc. On the road towards Open Science (OS), several Dutch universities proclaimed their desire to be a FAIR university within a 5-year period [WUR, 2021][UM, 2021]. Triggered by the use of the acronym FAIR in relation to organizations, we initiated a Task Group (TG) of Research Data Management (RDM) experts in the Netherlands, with the goal of defining the concept of a FAIR research organization and a set of principles on the basis of which an organization could be assessed on the degree of FAIR enabling.

As FAIR has become mainstream, other interesting initiatives have also been active. A Research Data Alliance (RDA) Working Group FAIR Data Maturity [RDA (1), 2021] has developed a common set of core assessment criteria for FAIRness and a generic and expandable self-assessment model for measuring the maturity level of a dataset. Other RDA Working Groups address FAIR for Research Software [RDA (2), 2021] and FAIRSharing Registries [RDA (3), 2021]. In addition, the RDA Special Interest Group on Data Stewardship addresses how Data Stewardship can contribute to FAIR data [RDA (4), 2021]. The FAIRsFAIR project (*https://www.fairsfair.eu/*). aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle. That project offers detailed tools on assessing one's knowledge on FAIR datasets, offers assistance in FAIRifying repositories and hands-on expertise. The work by semantic science (*http://www.semantic-science.org/*) too has focuses on the digital object data, contributing to the knowledge on machine readability of data. Having studied the existing initiatives, we concluded that additional work on a definition and tooling for **FAIR Enabling Organizations** could contribute to the advancement of RDM support and FAIR data policy & services developments within Dutch research organizations.

This article provides a definition of a FAIR Enabling Organization and describes the **Do I-PASS for FAIR** method to measure the FAIR-ness of research organizations [De Bruin et al, 2020] .

The acronym **I-PASS** is derived from the five subject categories on which an organization can measure or self-assess the degree to which it enables the researchers to be FAIR: **P**olicy, **S**ervices, **S**kills, **I**ncentives and **A**doption. By mixing the order of the categories we created the acronym **I-PASS** which we found to be well applicable to the method, since the method is meant to assess whether an organization 'passes' in providing sufficient support for FAIR data management.

In the process of developing the method for organizations, it also appeared to be useful for anonymous national benchmarking [see Miedema, 2021]. Comparing the outcome of the individual assessments to the aggregated results served as a starting point (both institutional and national), in developing a FAIR-enabling Road Map.

We believe that **Do I-PASS for FAIR** is of interest to our peer supporters in research data management, since it provides an opportunity to assess the level of Data Management support offered and also defines a Road Map for becoming more FAIR enabling to the research community. Finally, it is an opportunity to observe institutional and national challenges to improving the support provided to researchers to achieve compliance with the 15 FAIR principles.

MATERIALS AND METHODS

The National Coordination Point Research Data Management (LCRDM) is a network of RDM experts in the Netherlands (the data support collective) [LCRDM, 2021] [Verheul et al., 2019]. In this network, 250 members of 60 Dutch research institutes share knowledge and expertise, and work together to address RDM topics that require a joint and national approach. The LCRDM

has a coordinator (facilitated by the collaborative IT organization for Dutch education and research (SURF)) and has an advisory group of 12 RDM representatives from Dutch universities, universities of applied sciences, medical research institutes and data repositories providers. Members of the network can initiate new Task Groups (TGs) to address new and relevant Data Management developments jointly. A request for participation explaining the objectives of the Task Group is distributed within the network. The motivation for members to participate in a Task Group is that its result may contribute to better knowledge and insights on a given topic and easier adoption and implementation of these insights within their local research organization.

The initiators of the Task Group of which the results are described in this article, called upon the network members to **"define a number of principles on the basis of which an organization could be assessed on the degree of FAIR enabling and to deliver practical recommendations derived from these principles."** The Task Group consisted of 13 members who met four times in plenary discussions. Due to the COVID-19 crisis these meetings were held online. Sub-groups were set to work out different aspects. We created a prototype of a tooling, which we then tested through 15 personal interviews with senior RDM support staff at both universities and research institutes. Based on the interviews we made adjustments to the questions. In the process of the improvement, the 15 respondents requested that it be expanded to allow for anonymous benchmarking of research organizations. Once the final version was complete we invited 24 senior RDM staff members to answer the questions, which allowed us to establish the current state of FAIR-enabling maturity amongst Dutch research organizations.

The Task Group yielded two outcomes: (1) the **Do I-PASS for FAIR** method to evaluate the FAIR-ness of a research organization and to perform national surveys [De Bruin et al, 2020], and (2) a definition of a FAIR enabling organization.

RESULTS
DO I-PASS FOR FAIR FOR MEASURING THE FAIRNESS OF ORGANIZATIONS

The **Do I-PASS for FAIR** method provides 19 multiple-choice questions in five categories: **P**olicy, **S**ervices, **S**kills, **I**ncentives and **A**doption (**I-PASS**) (see *Figure 1*). The answers to the questions are

Figure 1 Summary of Categories of **DO I-PASS for FAIR.**

predefined, in three levels of maturity: beginner, intermediate and advanced. The 19 questions guide the user through an evaluation of the organization's level of enabling researchers to comply with FAIR. In addition, there are three open questions on barriers, achievements and plans for action for the coming year. The method is presented in an editable PDF file and is easily transferable to survey tooling like e.g. Google Forms or a Teams poll. The file with the questions and answers is available under a CC-BY through Zenodo [De Bruins et al, 2020].

DEFINITION OF A FAIR ENABLING ORGANIZATION

Applying the five categories, we arrived at the following definition of a FAIR enabling organization:

> *A FAIR Enabling Organization is a research organization with a dedicated staff of data professionals that has implemented policies on technical, infrastructural and organizational facilities and services to enable the researchers to create and publish FAIR data as a result of their research.*

APPLICATIONS OF DO I-PASS FOR FAIR
SELF-ASSESSMENT AND FAIR ENABLING ROADMAP DEVELOPMENT

With users going through the 22 questions and answering them on the basis of the current situation of the organization, the total set of answers provide information on the extent to which the organization enables researchers to handle their data in a FAIR manner. Preferably, each individual member of the organization's data team answers the questions in order to establish a commonly shared view. In addition, the answers provide information on the goals to set and activities to undertake in order to grow.

Table 1 shows how it works, using the example of question 5 of the category 'Services' of the tooling (the other questions work in a similar manner). The category 'Services' is defined as: *Does your organization have a (virtual) Digital Competence Center which provides services, including infrastructure, to allow researchers to comply with FAIR?*

SERVICES	BEGINNER	INTERMEDIATE	ADVANCED
Q5: Which services does you organization provide in order for researchers to comply with the F principles	We provide or refer to a service to deliver a PID for a data set	We provide or refer to a service to deliver a PID and adding metadata (including the reference to the dataset)	On top of PID and metadata we provide or refer to a service to make the data and metadata findable through an indexed resource

Table 1 Predefined answers to Question 5 in Category 'Services' of **Do I-PASS for FAIR**.

Imagine, for example, an organization which supports its researchers with publishing data with a persistent identifier (PID). The organization would answer the question with "We provide (or refer to) a service that delivers a PID for a data set". So, for this specific question the evaluation is that the organization is a Beginner in FAIR enabling.

From the answers in the Intermediate level "We provide (or refer to) a service for PID and adding metadata" it becomes clear that the next step this organization could make to become more FAIR enabling is to not just to provide a PID service, but in addition provide a metadata service. To become advanced, both the PID and the metadata should be indexed and part of findable resources.

Each of the 19 questions results in an evaluation of the current level and identifies the challenges that need to be mitigated in order to improve. The sum of the 19 challenges could be used as the starting point of a Road Map towards FAIR maturity of the organization. Thus, the method provides an instrument for self-assessing the current state of FAIR enabling of an individual organization and an instrument for structuring the discussion within the organization on how to increase the level of FAIR enabling and create Road Maps.

CROSS INSTITUTIONAL EVALUATION

Another application of the method is to use it in a national evaluation of research organizations on how they assist their researchers in making research data FAIR. The role of the LCRDM was crucial here. The LCRDM asked the Data Management Coordinators of different Dutch

universities and academic medical centers to fill in the form once per institute (anonymously). The aggregated data provided a means to evaluate, at a national level, how well organizations support their researchers in FAIR data handling and in which aspect of policy, services, skills, incentives or adoption, national action should be taken.

We approached 24 research organizations to fill in the survey, of which 20 responded within a month. The 20 institutes consisted of 13 universities, five UMCs, one university of applied sciences and one national research institute. The experience and knowledge of the individual filling in the survey has an influence on the outcome of the assessment. We approached individuals centrally placed within Research Support Services; these individuals were identified via the LCRDM Network.

Most Dutch research organizations assess themselves as intermediate or advanced in the policy category. The Services that support the F and the A aspects of data management are better developed than services for the I and the R. On research software sustainability almost half of the institutes assess themselves as real beginners.

The majority indicate that they are in the process of recognizing different roles within data support: the first step on the road to professionalization of data support. As discussion continues in the Netherlands on the support staff/scientific staff ratio (we do not know what ratio would be optimal), for now it is interesting to see that 40 % of the institutes have a ratio of less than 1 FTE data support professional to 500 scientific staff. Whatever ratio ultimately appears to be optimal, it is clear that one person will not be able to support 500 scientists with FAIR enabling support. Another notable outcome is that the majority indicate that their institute only has a partial overview of the data generated in the institute.

At the end, the participating research organizations were offered the accumulated results, and thus were able to compare their own assessment with the aggregated data of 19 anonymous others. In addition to this anonymous benchmarking, we were able to identify the challenges common to FAIR enabling organizations, and for taking national steps to address these challenges.

DISCUSSION

Our definition of a FAIR enabling organization is a first step to disconnecting the acronym FAIR from other contexts than data, without losing sight of the intention of research organizations in becoming FAIR.

In developing the method **Do I-PASS for FAIR**, we had to decide on the categories and how to define the levels of beginner, intermediate and advanced. For the three categories Policy, Services and Skills, it was relatively easy to set the initial levels. For example, for the questions and definition for the levels in the category Services, the 15 FAIR principles provided the four questions (one question on each main principle) and the levels (inspired by the principles for each letter of the acronym). However, for the categories Incentives and Adoption, defining the questions and levels in the answers provides a challenge for the TG. For the category Adoption, we decided to focus on the visible adoption of FAIR data practices, being how much of the organization's data has been made available in a FAIR manner and how many of the peer-reviewed publications contain a reference (PID) to the data underlying the publication. For the category Incentives, we decided to focus on if and how well FAIR practices are included in the contracts and recruitment processes of new researchers. We realize that especially for this category (Incentives) the questions will evolve over time as incentives for FAIR data become more explicit within academic rewarding systems.

The initial questions and definition of the levels in the categories were tested by means of individual interviews with 15 senior RDM staff members of universities, repositories and academic medical centers. We believe that this testing phase was crucial to the proper working of the method. The questions can be answered in one hour. The experience of our RDM colleagues contributed to the refinement of the definition of the different levels. We believe that being a TG of the LCRDM network played a crucial role in allowing us to contact our RDM colleagues in a spirit of openness. The strength of the network played an essential role in the development of the method. LCRDM members were eager to test, which led to high response rates and an easy and wide outreach. We also believe that FAIR data practices and FAIR data support is still

a dynamic field, and therefore we recommend that this first version of the method is to be regarded as merely a first step, which we hope will be used by many and developed further.

The questions and even the answers can easily be adjusted for other national contexts or languages. For instance, in the follow-up of the first version of **Do I-PASS for FAIR** we are now working together with RDM experts from universities of applied sciences to make adjustments for this stakeholder group. In the Netherlands, the universities and academic medical centres first started support activities for RDM around 2010. In the last few years, universities of applied sciences have also initiated RDM support. We argue that in time more versions become available depending on the needs and characteristics of the stakeholder group. Similarly, given the KISS principle [Wikipedia, 2021] on which we built the method, local versions for different countries can easily be created.

We have already explained that the application of the method comes in two forms: (1) an instrument for self-assessment and roadmap development for individual organizations and (2) an instrument for monitoring to identify the national level of FAIR data support and national challenges. We advise that **Do I-PASS for FAIR** will be used as anonymous benchmarking with which individual organizations can match their results to the aggregated levels.

The results of the first national survey with **Do I-PASS for FAIR** indicated that in the Netherlands most universities and academic medical centres perform at least at an intermediate level in Policy, Services and Skills. However, in Incentives and Adoption many perform at or even below the Beginner level. We also identified that many support units are not aware of how much data is used, created or published within their organizations. This is why a follow-up Task Group was initiated on 'Measuring the adoption of FAIR data practice'.

Do I-PASS for FAIR is intentionally kept as simple and robust as possible (KISS), and its simplicity promotes its use for different purposes. By answering 22 questions, organizations can get a good overview of how they are doing in terms of enabling their researchers in becoming FAIR, and establish a roadmap for actions in the years ahead.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **First NL Survey Do I-pass for FAIR? jan 2021 (Version 1).** Report and survey data Margriet Miedema (2021). *https://doi.org/10.5281/zenodo.5172012*. DOI: *https://doi.org/10.5334/dsj-2021-030.s1*

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

J. Ringersma 🄳 *orcid.org/0000-0001-7465-9172*
Wageningen Data Competence Center, Wageningen University & Research, NL
M. Miedema 🄳 *orcid.org/0000-0002-4664-9492*
National Coordination Point for Research Data Management (LCRDM), SURF, NL

REFERENCES

LCRDM. RDM@LCRDM_National Coordination Point Research Data Management: Output and products. *https://zenodo.org/communities/lcrdm-output-and-products/* (visited online April 2021).

Miedema, M. First NL Survey Do I-pass for FAIR? Jan 2021 (Version 1). DOI: *https://doi.org/10.5281/zenodo.5172012*

RDA (1). FAIR Data Maturity Model WG. *https://www.rd-alliance.org/groups/fair-data-maturity-model-wg* (visited online April 2021).

RDA (2). FAIR for Research Software (FAIR4RS) WG. *https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg* (visited online April 2021).

RDA (3). FAIRsharing Registry: Connecting data policies, standards and databases RDA WG. *https://www.rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html* (visited online April 2021).

RDA (4). Professionalising Data Stewardship IG. *https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig* (visited online April 2021).

Taco de Bruin, SC, de Jong, J, Haslinger, I, van den Hoogen, H, Huigen, F, … Ringersma, J. 2020, November 3. Do I-PASS for FAIR. A self-assessment tool to measure the FAIR-ness of an organization. DOI: *https://doi.org/10.5281/zenodo.4080867*

UM. Maastricht University, Open Science, *https://www.maastrichtuniversity.nl/research/open-science* (visited online May 2021).

Verheul, I, Mordant, A, Ringersma, J, Sesink, L, Smeele, T, Boiten, J-W, … Verheij, M. 2019, July 15. National Coordination Point for Research Data Management (LCRDM) Positioning paper for 2019 and beyond (Version final - online version). *Zenodo*. DOI: *https://doi.org/10.5281/zenodo.3336040*

Wilkinson, M, Dumontier, M, Aalbersberg, I. et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3: 160018. DOI: *https://doi.org/10.1038/sdata.2016.18*

Wikipedia. KISS principle, *https://en.wikipedia.org/wiki/KISS_principle* (visited online April 2021).

WUR. Wageningen University & Research, Wageningen University & Research to make important steps to FAIR data science, *https://www.wur.nl/en/show/WUR-steps-to-FAIR-data-science.htm* (visited online May 2021).