

Introducing FAIR Scores in a Global Agricultural Science Reporting Service: An Analysis of the First Reporting Period

Sebastian S. Feger, ICARDA & LMU Munich, S.Feger@cgiar.org; Valentina De Col, ICARDA, V.DeCol@cgiar.org; Moayad AL-Najdawi, CodeObia, moayad@codeobia.com; Juan Carlos Cadavid, CIAT, J.Cadavid@cgiar.org; Hector Tobon, CIAT, H.F.Tobon@cgiar.org; German Martinez, CIAT, G.Martinez@cgiar.org; Enrico Bonaiuti, CIP-ICARDA, E.Bonaiuti@cgiar.org;

Abstract

Communicating the meaning and value of the FAIR (Findable, Accessible, Interoperable, and Reusable) principles and suitable implementation strategies to research communities and their broader ecosystem is a crucial challenge. Automated FAIR scoring algorithms are being developed to provide immediate machine-driven insight into (meta)data compliance. They are also expected to return instructions to improve FAIR compliance where applicable. However, we still lack a systematic understanding of how automated FAIR scoring impacts adoption in repositories that implement them. Accordingly, we are excited to share findings from the first large-scale machine-driven FAIR scoring of global agricultural research at One CGIAR, a global partnership of international organizations dedicated to sustainable food production. Characteristics that make this analysis particularly interesting include the organization-wide visibility of the FAIR scoring and the mandatory reporting of all scientific resources in the preceding calendar year. We find that 39% of the 418 records received updates. Out of those, 7.3% resulted in improved FAIR scores. Those improved records outperformed the mean FAIR scores of the complete 2022 reporting data, while falling behind them before the update. We further show differences according to the knowledge product type and outline how we gather additional data on users' perceptions for the OR2023 presentation.

Keywords

Automated FAIR scoring; Adoption and perception across scientists and data managers; Usage analysis.

Audience

We expect repository managers and developers to profit from this first-hand analysis of newly introduced machine-calculated FAIR scores in a large-scale agricultural science repository. We further assume that librarians and data managers will gain valuable insights into the collaborative practices with researchers to create the most suitable repository entries that achieve high FAIR scores.

Proposal

Our analysis of FAIR score adoption and perception strongly aligns with the conference theme (i.e., "Repositories Unlocked for Discovery and Interoperability"). Ultimately, *discoverability* closely relates to findability and accessibility of scientific resources and, with *interoperability*, they represent three of the four pillars of the FAIR data principles. In recent years, an increasingly data-intensive science revealed the need to calculate and present automated FAIR scores of research materials for immediate compliance feedback and as a support tool for the broader scientific community. One prime example of such an initiative is FAIR EVA [1], a framework to automate evaluation, validation, and advising of FAIR data compliance, presented by Aguilar and Bernal at OR2022¹. Along those lines, One CGIAR decided to add a

¹OR2022 session reference with abstract: https://www.conftool.net/or2022/index.php?page=browseSessions&form_session=449

first generation of FAIR scoring algorithms in their new scientific reporting tool within its Performance and Results Management System (PRMS). Our presentation proposal builds on the unique opportunity to provide a first-hand account of the introduction of FAIR scores in a new mandatory research reporting tool and across a global research community consisting of scientists, librarians, and data managers, previously not exposed to such metrics. Our presentation aims to make the following key contributions that we outline more closely in this proposal:

- Presentation of the new One CGIAR reporting tool, the FAIR scores, and their user-centered design.
- Analysis of the research documentation behavior of the One CGIAR global agricultural science community during the first-ever reporting period, with a particular focus on FAIR scoring and subsequent (meta)data documentation updates.
- Discussion of design decisions regarding the FAIR scores and their expected evolution in the reporting tool. This will be closely aligned with a survey study and follow-up interviews that we will conduct and analyze before the OR2023 conference and that aim to closely map the communities' perceptions towards the introduction of organization-wide visible FAIR scores, particularly emphasizing the differing complexity across the FAIR criteria and the interplay between data managers, librarians, and scientists in collaboratively creating the best documentation and sharing outcome.

Design of PRMS and the FAIR scores

PRMS results from One CGIAR reforms aiming to create a common system across 15 research centers that houses theory of change management, data quality assessment, and other processes such as planning and reporting. The system aims to meet learning, accountability, and resource mobilization objectives. The PRMS reporting tool captures impacts, outcomes, and outputs, including knowledge products such as journal articles, datasets, reports, presentations, and briefs. Users accessing the PRMS for reporting are CGIAR staff, including researchers and data managers. The reporting process opens in December, to capture the progress achieved in the preceding calendar year. Scientific and technical knowledge products that get recorded in CGSpace², the centralized institutional repository, are then linked in the PRMS reporting tool through their uniform resource identifier – the handle. When the researcher/data manager reports a specific knowledge product, they add the handle in the PRMS. Next, through a tool called M-QAP (Monitoring, Evaluation, and Learning Quality Assurance Processor), metadata such as title, abstract, partners, and links to other data are automatically retrieved from the repository into the PRMS. This process guarantees interoperability between systems and saves time by reducing user input.

To check compliance with the CGIAR Open and FAIR Data Assets Policy³, the M-QAP tool has been set up to check for specific metadata fields based on a set of seven FAIR criteria. The selected criteria were the ones considered as easy to understand, follow, and review by researchers, librarians, and repository managers. Although we recognize the plurality of methods for assessing FAIR compliance, we aimed for a simplified version in this first iteration, in which each criterion has the same weight and in which the scores are easy to interpret (see Figure 1). The selection and description of these seven criteria resulted from a highly iterative user-centered process that involved several developers, User Experience (UX) experts, and potential users. Based on this process, that also profited from our previous extensive user-centered requirements research at CGIAR [2], we learned that the initial introduction of FAIR scores in a community should carefully strike a balance between challenges and rewards. We placed particular emphasis on adding some criteria that are easy to achieve (e.g., F1 and F3), to ideally increase FAIR scoring acceptance, while also adding more challenging ones that require a closer study of the FAIR principles (e.g., I2). Again, we note that this is the first design iteration and we expect that the FAIR scoring will iteratively and

² <https://cgspace.cgiar.org/>

³ [CGIAR Open and FAIR Data Assets Policy](#)

annually evolve between reporting periods based on the experiences and perceptions that we collect. We hope and expect that OR2023 participants will profit from these findings and discussions.

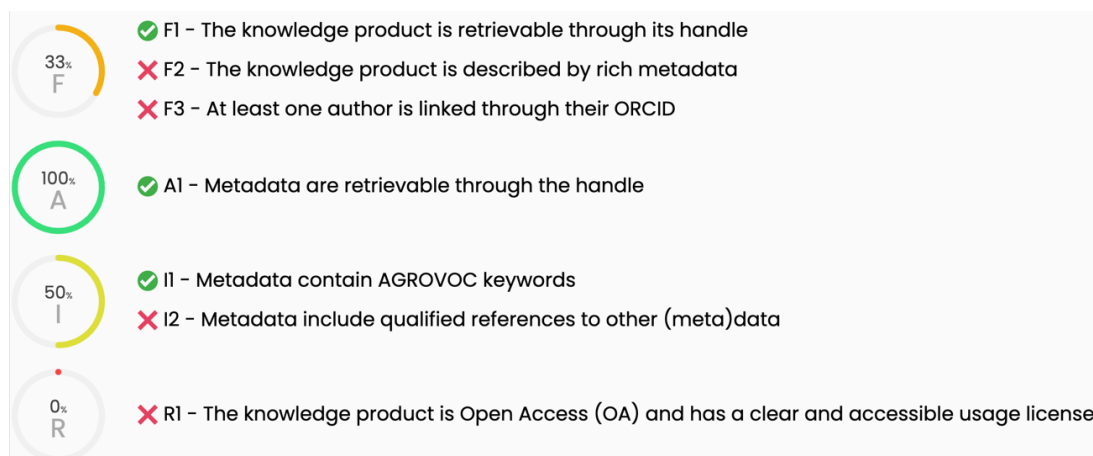


Figure 1. An example of FAIR criteria scores related to a knowledge product.

Data Analysis of the First Reporting Period

We registered a total of 418 complete datasets that were recorded by the One CGIAR research community. As shown in Figure 2 a, the mean FAIR scores assigned to the records after initial upload differed across the FAIR criteria. Mean scores across all 418 records were: Findable 85%, Accessible 100%, Interoperable 63%, and Reusable 77%. The researchers and data managers had the chance to update the (meta)data documentation and to improve the FAIR scores until the end of the reporting period. In total, we found that 163 records (i.e., 39% of all records) received at least one update. Out of those, 12 records received an improved FAIR score based on the updates. That amounts to 7.3% of the updated records and 2.9% of all 2022 records. Figure 2 b depicts how the FAIR scores improved for those 12 records. We note that the mean scores of the FAIR principles after record update exceeded the mean scores of the complete reporting data (compare to Figure 2 a). This is particularly interesting given the fact that they were previously consistently lower than the baseline reference in Figure 2 a. The only exception is the Accessible principle that remained at 100%.

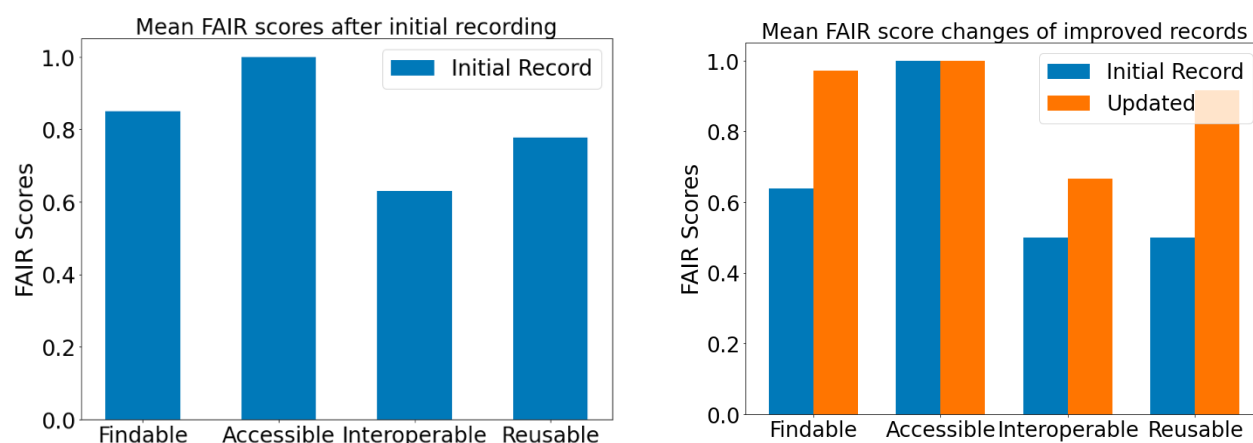


Figure 2. left (a): Mean FAIR scores of all records after initial recording; right (b): Mean FAIR score changes of the updated records that resulted in improved scores.

As shown in Figure 3 a, most 2022 reporting records relate to journal articles (61.5%), briefs (11.7%), and reports (6.5%). However, no FAIR score improvements were registered for journal articles. Instead, as depicted in Figure 3 b, most of the improved records relate to briefs, reports, presentations, and posters.

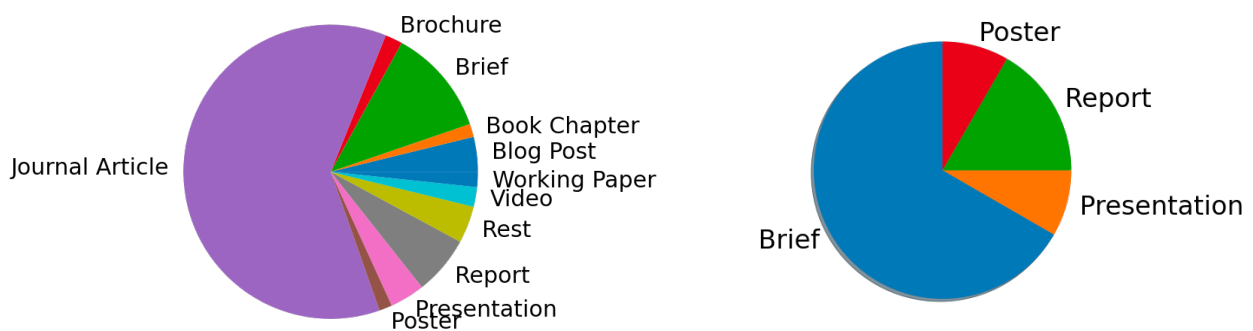


Figure 3. left (a): Overall knowledge product ratio; right (b): Distribution of improved records.

Discussion, Outlook, and Presentation at OR2023

The 2022 One CGIAR reporting period gave us a unique opportunity to provide a first-hand account of the adoption of a new global research reporting tool and the equally new automated FAIR scoring. While we note that the scores need to be iteratively refined, improved, and ultimately evolve across annual reporting periods, we underline the sensible nature of introducing such automated scoring algorithms for the first time in a research community. Overall, we are content to have received positive initial signals from the research community. Yet, we also directly note needs for improvement. For example, while the overall FAIR scores in Figure 2 indicate that our goal of striking a balance between FAIR score challenges and rewards worked out for most of the FAIR pillars, clearly the criteria chosen for the Accessible principle needs to be revised, as all records immediately scored 100%. We would be happy to present and discuss findings like these in more detail at OR2023.

Further, we commit to follow-up studies with users who participated in this first FAIR scores reporting period. In particular, we will run a survey study in February 2023 and conduct follow-up interviews Early 2023 to systematically map users' perceptions towards introducing community-wide visible automated FAIR scores. We will also present and discuss these findings at OR2023 in case this proposal gets accepted.

Overall, our work closely relates to the tensions between differing FAIR criteria complexity in automated FAIR scoring, as well as the scores' impact on documentation behavior and users' comfort level. We argue that these findings and discussion points are highly relevant to the OR community, as evidenced by current initiatives¹ [1] and the strong emphasis of OR2022 to study and consider user perceptions closely. Linking directly to the OR2023 theme, we hope our presentation will help further unlock FAIR compliance through sophisticated automated FAIR scoring and actionable user feedback.

References

- [1] Aguilar, F., & Bernal, I. (2022). FAIR EVA: Checking data FAIRness in data repositories.
- [2] Feger, S. S., Pertiwi, C., & Bonaiuti, E. (2022). Research Data Management Commitment Drivers: An Analysis of Practices, Training, Policies, Infrastructure, and Motivation in Global Agricultural Science. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-36.
<https://hdl.handle.net/20.500.11766/67317>