Case study on the Monitoring-Quality Assurance Processor-API

A tool to support CGIAR Quality Assurance process for peer-reviewed publications

November 11, 2021

Monitoring, Evaluation and Learning (MEL)

Monitoring, Evaluation and Learning (MEL) teams improve the decision-making and impact of research organizations through four areas of expertise: Monitoring and evaluation to plan, implement and evaluate the impact of projects and programs throughout the project lifecycle; knowledge management to capitalize on learning, dissemination and knowledge sharing; data management to collect data and ensure its quality and accuracy; and software development to build the MEL system on a platform. Each area has a coordinator relying on specialists and research fellows who help to facilitate planning, budgeting, reporting, and risk assessment on the MEL Platform, a tool facilitating the implementation of MEL. MEL Platform centralizes the collection, visualization and use of data for more informed decision-making and research impact of organizations.

AUTHORS

Valentina De Col[1], Sara Jani[1], Max Rünzel[2], Héctor Tobón[3], Manuel Ricardo Almanzar[3], Diu Seng See[4], Enrico Bonaiuti[1,5]

SUGGESTED CITATION

Valentina De Col, Sara Jani, Max Rünzel, Héctor Tobón, Manuel Ricardo Almanzar, Diu Seng See, Enrico Bonaiuti (2021). Case study on the Monitoring-Quality Assurance Processor-API - A tool to support CGIAR Quality Assurance process for peer-reviewed publications. International Center for Agriculture Research in the Dry Areas (ICARDA).

HANDLE

https://hdl.handle.net/20.500.11766/66480

[1] The International Center for Agricultural Research in the Dry Areas (ICARDA)
[2] CGIAR Advisory Services Shared Secretariat (CAS)
[3] Alliance of Bioversity International and CIAT
[4] Clarivate
[5] The International Potato Center (CIP)

DISCLAIMER

Version	Date	Originator(s)	Reviewer(s)	Description
1	16/07/2021	Valentina De Col	Sara Jani, Enrico Bonaiuti	Structure and content
2	28/07/2021	Valentina De Col	Max Rünzel	Content revision
3	17/09/2021	Valentina De Col	Sara Jani, Enrico Bonaiuti	Content revision
4	27/09/2021	Valentina De Col	Héctor Tobón, Manuel Ricardo Almanzar	Content revision
5	11/10/2021	Valentina De Col	Diu Seng See	Content revision

Index

Glossary

A4NH	CGIAR Research Program on Agriculture for Nutrition and Health
API	Application programming interface
AR	Annual report
Big Data	CGIAR Research Program for Big Data in Agriculture
CCAFS	CGIAR Research Program on Climate Changes, Agriculture and Food Security
CGIAR	Consultative Group for International Agricultural Research
CIAT	International Center for Tropical Agriculture
CIP	International Potato Center
CLARISA	CGIAR Level Agricultural Results Interoperable System Architecture
CRP	CGIAR Research Program
DOI	Digital object identifier
EiB	Excellence in Breeding
F.A.I.R.	Findability, accessibility, interoperability, and reusability
Fish	CGIAR Research Program on Fish
FTA	CGIAR Research Program on Forests, Trees and Agroforestry
Gender	CGIAR Gender Platform
Genebank	CGIAR Genebank Platform
GLDC	CGIAR Research Program on Grain Legumes and Dryland Cereals
ICARDA	International Center for Agricultural Research in the Dry Areas
ISI	Formerly known as the Institute for Scientific Information™
ISSN	International standard serial number
ISTIC	Institute of Scientific and Technical Information of China
Livestock	CGIAR Research Program on Livestock
Maize	CGIAR Research Program on Maize
MARLO	Managing Agricultural Research for Learning and Outcomes
mEDRA	Multilingual European DOI Registration Agency
MEL	Monitoring, evaluation and learning
MELIA	Monitoring, evaluation, learning, and impact assessment
MIS	Management information system
M-QAP	Monitoring, evaluation and learning quality assurance processor
OA	Open access
OICR	Outcome impact case reports
PIM	CGIAR Research Program on Policy, Institutions and Markets
QA	Quality assurance
Rice	CGIAR Research Program on Rice
RTB	CGIAR Research Program on Roots, Tubers and Bananas
SLO	System-level outcome
SMO	CGIAR System Management Organization
Wheat	CGIAR Research Program on Wheat
WLE	CGIAR Research Program on Water, Land and Ecosystem
WoS	Web of Science™

Executive summary

AIM | This report aims to present a first case study on the use and performance of the Monitoring, Evaluation, and Learning Quality Assurance Processor (M-QAP-API), a tool recently developed and introduced for the CGIAR quality assurance (QA) process for journal articles. Whereas, previously, the QA process relied on the manual check of all publications reported by CGIAR Research Programs (CRPs) and Platforms, this tool now allows an automatic, reproducible, reliable, and rapid way to assess the Web of Science (WoS) Core Collection (formerly known as ISI, the Institute for Scientific Information), along with the open access (OA) status of thousands of publications at once, thanks to the integration with different Application Programming Interfaces (APIs). Introduced for the CGIAR CRPs and Platforms 2020 Annual Report (AR), the M-QAP-API tool is exposed within the CGIAR CLARISA centralised service and it queries Clarivate (WoS), Scopus, Unpaywall, Crossref, Altmetric, and F.A.I.R metrics from GARDIAN by using the publication's digital object identifier (DOI). Data collected from the CLARISA platform were evaluated after and before manual checks by quality assessors, to appraise the performance, limitations and future applications of the tool.

MAIN FINDINGS AND IMPACT | The tool has allowed the automatic validation of more than 2,500 DOIs (97.4% of the total publications sent by CRPs and Platforms) and automatically confirmed—as covered in Web of Science Core Collection and OA—almost 90% and 80% of the DOIs, respectively. This, in turn, supports the validation of metadata according to CGIAR System Management Organisation (SMO) guidelines. Compared to a manual check, it has saved approximately 40 days of work and provided a standardized, rapid, precise, reproducible, time- and resource-saving solution for the CGIAR QA process of scientific publications.

RECOMMENDATIONS | The tool proved to be a valid support for the CGIAR QA process, CRPs and Platforms, ARs, CGIAR Annual Performance Report, and CGIAR Results Dashboard. Regular queries to the databases are suggested as changes might occur overtime to the indexing in WoS Core Collection and OA status of recent publications, but also for Altmetric and F.A.I.R. scores. The tool could also serve to support knowledge management teams when curating and validating metadata for publications through integration in official repositories. In future, more metadata could be extracted by the tool to expand data retrieval and collection, and to provide the basis for further analysis—for example, related to bibliometric and Quality of Science.

LIMITATIONS | For data completeness, the tool should be paired with the manual check provided by quality assessors for all publications: (i) without a DOI; (ii) returning no metadata; (iii) with incorrect DOIs; and (iv) waiting for indexing in the repositories after publication.

KEYWORDS | CGIAR Quality Assurance, peer-reviewed publications, API, Web of Science Core Collection, ISI, OA

What is the Monitoring-Quality Assurance Processor-API tool? A brief introduction of the tool

The Monitoring, Evaluation, and Learning Quality Assurance Processor (M-QAP) is a publications' metadata extractor that employs Application Programming Interfaces (APIs) from Clarivate (WoS), Scopus, Unpaywall, Crossref, Altmetric, and F.A.I.R metrics from GARDIAN (Fig. 1). The tool, M-QAP-API for short, is designed to ensure that publications with a DOI are validated against the above-mentioned databases. During the annual CGIAR QA process, the tool has been used to support the reporting of CGIAR results, including the CGIAR Results Dashboard. The M-QAP-API has been designed by the Monitoring, Evaluation and Learning (MEL) team at the International Center for Agricultural Research in the Dry Areas (ICARDA) with financial support from the CGIAR System Management Organisation (SMO). It has been integrated by the Innovations and Business Development team of the Technology Integration unit at the Alliance of Bioversity International and CIAT to support existing processes in CLARISA and other management information system (MIS) platforms, such as MEL and MARLO. Its code is open source[6].

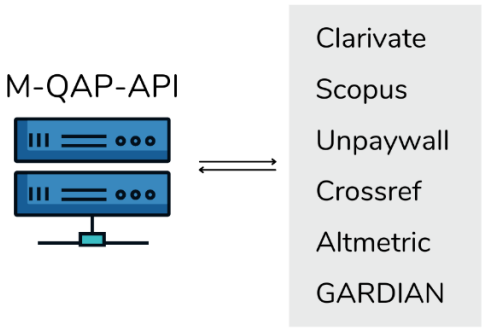

Figure 1: Simplified representation of M-QAP-API. The tool queries six different databases, namely WoS, Scopus, Unpaywall, Crossref, Altmetric, and GARDIAN, and extracts specific information.

Within the scope of CGIAR QA of peer-reviewed publications, the M-QAP-API has retrieved data and metadata from:

1. **WoS API Expanded**, powered by Clarivate, returns if the publication is indexed or not in the WoS Core Collection.
2. **Scopus API**, powered by Elsevier Developers, returns if the publication is in the Scopus citation database. In this context, it supports the validation of peer-reviewed publications[7].
3. **Unpaywall API** returns the response about the publication's OA status.

[6] https://github.com/icarda-git/M-QAP-API
[7] It is assumed that if the DOI is found in the Scopus database, the article can be considered peer-reviewed.

4. **Crossref API** retrieves the information about the DOI Registration Agency. This API has been added to the tool as Unpaywall only covers the Crossref Registration Agency[8] and excludes, for example, DataCite, of which 99% of articles are OA[9]. Querying the Crossref database allows the finding of the DOI registered by DataCite and sets them automatically as OA to improve coverage.

Besides the information retrieved by M-QAP-API tool for the QA, the APIs can return supplementary metadata. Appendix A includes an overview of the metadata used from each API for the QA, and a preview of additional metadata that could serve future automation and applications.

Why use the M-QAP-API tool?

The M-QAP-API was developed to validate, in an accurate, reliable and automatized way, the assessment of peer-reviewed publications submitted by CGIAR entities as part of their annual reporting. The use of this tool saves time (approximately 40 days of work) and resources—also upon data entry, as it is requested that all CGIAR CRPs and Platforms accurately submit the publications' DOIs. Previously, the assessment was performed via a manual check by different individuals using non-standard ways of verifying the WoS Core Collection and OA status. Moreover, the use of a benchmark academic database of reference, such as WoS, allows for a consistent, reliable and replicable retrieval of data, which, in turn, brings value to the overall process and reduces errors and misinterpretations.

Advantages and disadvantages of the M-QAP-API tool

1. It automates the validation process and analyzes of thousands of publications at once.
2. It only requires the DOI as an input.
3. It supports the consistent and reproducible collection of data.
4. It allows integration in MIS platforms to support users in filling publications' metadata and provide responses in real-time (e.g., on WoS Core Collection, OA). The tool has been already integrated and tested in both MEL and MARLO MIS platforms.
5. It relies on a paid account for the WoS API Expanded, whereas Scopus, Unpaywall, Crossref, Altmetric, and GARDIAN APIs are free.
6. It is not completely free from errors: One sub-set of the results ("DOIs not found", which, in the case of the present case study, is estimated to be around 0.2% of the total) requires validation via a manual check and/or contact to the WoS/Unpaywall technical support for further investigation.

[8] https://support.unpaywall.org/support/solutions/articles/44001900286-which-dois-does-unpaywall-cover-
[9] Unpaywall, personal communication.

7. Publications without a DOI (in this case study, 2.6%) cannot be processed by the tool and require validation via a manual check. Guidance for manual validation and interpretation of the result of the tool can be found in a previously published report[10].
8. The technical delay in the journal article indexing (from the publication to the record in the database), especially for more recently published articles, causes the API to not return any response. As such, it then requires either a manual check and validation or the performance of a second query after some time (weeks/months).

[10] https://repo.mel.cgiar.org/handle/20.500.11766/13115

The case study

CGIAR quality assurance (QA) process

The CGIAR 2020 QA process took place after CRPs and Platforms submitted their AR at the end of April 2021. It considered eight reported indicators[11], among which were peer-reviewed papers. The QA process is hosted in the centralized web service CLARISA, a service equipped with functionalities that enable MIS platforms (MEL and MARLO) to communicate with each other, collect, standardize, and aggregate information in the language needed for the System-Level reports[12] tool (Fig. 2). CLARISA is managed by the Innovations and Business Development team at Alliance of Bioversity and CIAT with the support of SMO, ICARDA, and the International Potato Center (CIP). This service was previously used for the 2019 QA and its functionalities have been further improved in 2021, including the addition of the M-QAP-API.

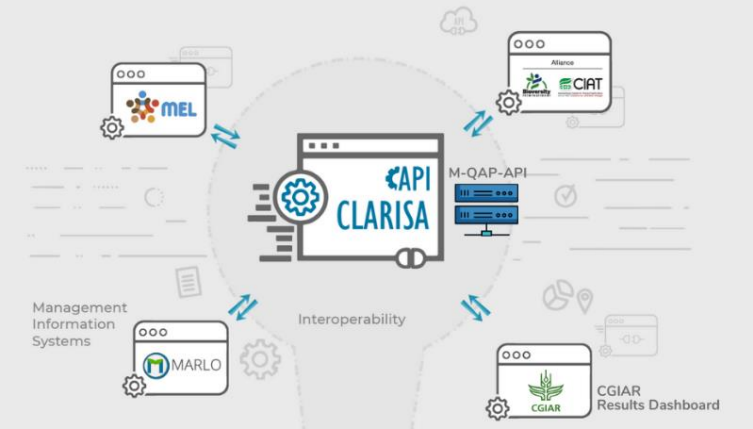

Figure 2: The M-QAP-API tool is integrated within the CLARISA environment, which guarantees interoperability among MIS platforms and other interfaces (figure adapted from the CLARISA website).

In May and June 2021, a group of evaluation experts for each indicator, called quality assessors, reviewed information sent during two rounds. The first round saw the validation or invalidate of the information submitted by the CRPs and Platforms, while the second acknowledged comments and changes received in the first round and provided requests for clarification by the CRPs and Platforms.

[11] Contributions to System-Level Outcomes (SLOs) targets, Policies, Outcome Impact Case Reports (OICRs), Innovations, Milestones, Capacity Development, Monitoring, Evaluation, Learning and Impact Assessment (MELIA) and Peer-Reviewed Papers.

[12] https://clarisa.cgiar.org/swagger/home.html#about

The quality-assured data on CGIAR key indicators ultimately feed the Annual Performance Report and serve the Results Dashboard, offering an overview of the results achieved by the CRPs and Platforms in research for development.

The main aim of this case study

As previously noted, the M-QAP-API was used for the first time in 2021 to assess peer-reviewed papers submitted by MIS platforms to CLARISA. The report aims to present preliminary results, in terms of the tool's performance and the possibility of use beyond the QA.

The M-QAP-API in action

Among the journal articles' metadata sent to CLARISA, and in line with SMO guidance (i.e., title, author(s) name(s), DOI, handle, article URL, WoS Core Collection status, OA status, journal name, volume, issue, pages, and year of online publication), is the DOI—a unique identifier used by the tool to query Clarivate, Scopus, Unpaywall, and Crossref databases through the respective APIs (Fig. 3).

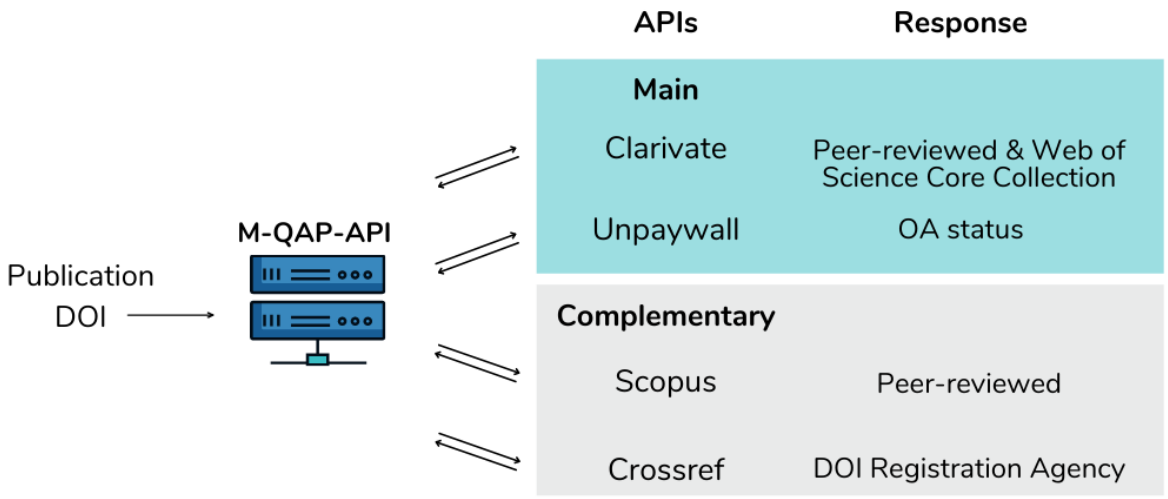

Figure 3: Schema of the main and complementary APIs and their respective response used within the QA.

For the 2020 QA, CLARISA received a total of 2,577 peer-reviewed publications[13] from 16 different CRPs and Platforms (Fig. 4 and Appendix B). Of these, 2,509 (97.4%) were provided with a DOI and therefore processed through the M-QAP-API.

[13] Among the total DOIs received in CLARISA, around 14% are duplicated as different CRPs contributed in the same publication with their indicators submitted separately.

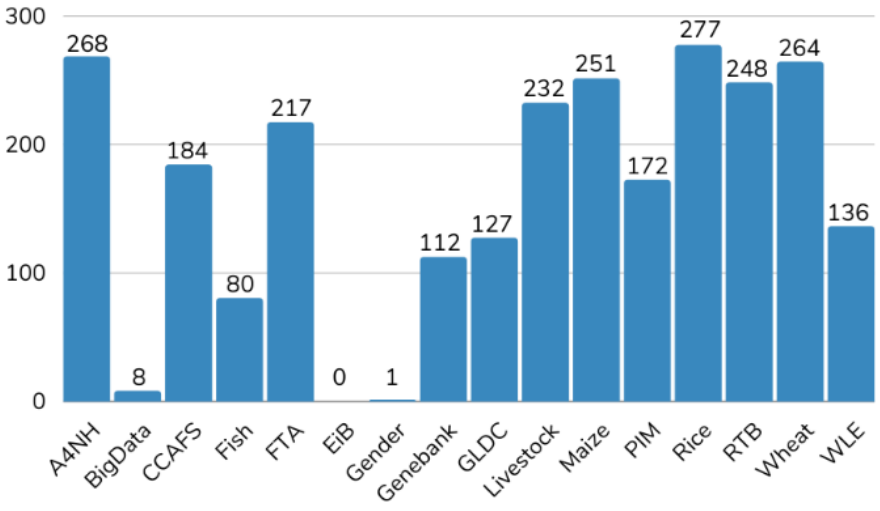

Figure 4: Numbers of peer-reviewed papers submitted to CLARISA by each CRP or Platform for the 2020 QA.

Results from the tool before and after manual QA

A. Web of Science Core Collection status

Journals covered in the WoS Core Collection need to meet a minimum of 24 Quality Criteria if they are to be covered in the Emerging Sources Citation Index. If journals meet an additional four Impact Criteria, they will be covered one or more flagship indexes, namely Science Citation Index Expanded, Social Sciences Citation, or Arts & Humanities Citation Index. These stringent criteria mean the WoS Core Collection covers highly reliable and impactful publications[14]. The indexing in at least one index of the WoS Core Collection[15], was confirmed through the WoS API.

Within the first assessment run by the CLARISA team on 2,509 publications with a DOI, the result shows that 2,244 publications (89.4%; Fig. 5) were automatically validated as WoS Core Collection by the tool. The remaining 258 publications (10.4%; Fig. 5) were identified as not being indexed in the WoS Core Collection and, depending on the data completeness, aggregated to the pool of records requiring manual check by the quality assessors. A non-significant percentage of papers (0.2%; Fig. 5) had a DOI that could not be processed by the tool because it was found to be invalid[16]; and was therefore added to the group requiring manual validation.

[14] https://clarivate.com/webofsciencegroup/journal-evaluation-process-and-selection-criteria/

[15] Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (AHCI), and Emerging Sources Citation Index (ESCI) are the four journal indexes of the WoS Core Collection™.

[16] An invalid DOI is when there is no data returned by the tool for any request through the APIs.

Following the manual validation performed by the quality assessors in the first round, the number of WoS Core Collection publications potentially[17] improved (+3.4%; Fig. 5). This increase mainly referred to the number of the DOI/journal articles not yet indexed in at least one of the WoS Core Collection Indexes (the time for indexing depends on the journal itself and the WoS indexing service, and can take up to six weeks[18]), and therefore these were not found by the WoS API.

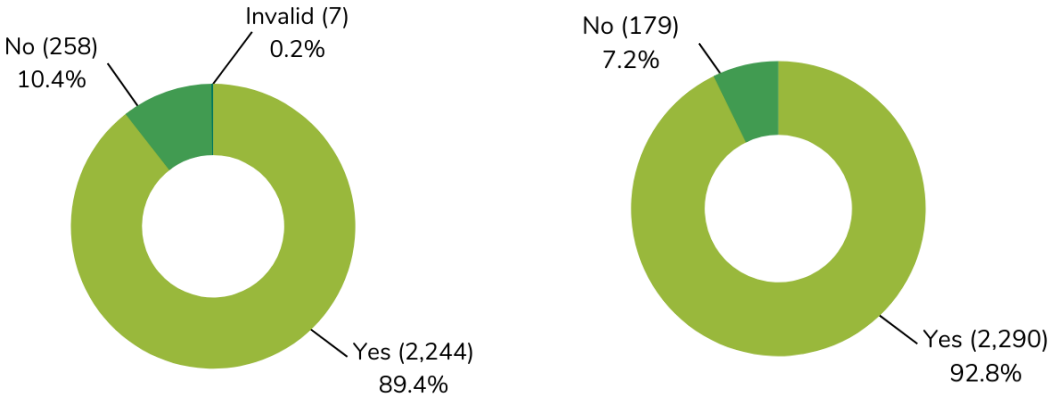

Figure 5: Results of the response of the M-QAP-API regarding the WoS Core Collection status before (left) and after (right) manual QA.

The status of invalid DOIs was also resolved by searching the publications in the MIS platform and/or via search engines, such as Google and Google Scholar, using other available metadata (i.e., title, handle) (Table 1).

To further estimate possible changes over time of the WoS Core Collection status (which can be followed by the tool), approximately two months after the first validation performed in CLARISA, all the DOIs/journal articles that initially returned a result for the WoS Core Collection status as 'No' were re-evaluated through the M-QAP tool—which is similar to the QA tool in CLARISA, but hosted within the MEL platform. The M-QAP tool was used within the MEL environment during the initial testing of the APIs, and allows authorized MEL users to query WoS, Scopus, Unpaywall, and DataCite by DOI through a Microsoft Excel input template. Its code is also open source[19].

[17] In the second round, CRPs and Platforms can either accept or refuse the suggestions of the quality assessors.
[18] Personal communication with the WoS technical support.
[19] https://github.com/icarda-git/M-QAP

The total number of articles found as indexed in WoS Core Collection improved by 1.9% (from 2,244 to 2,286). When comparing further with the group validated as WoS Core Collection by the quality assessor, only 12 articles were present in both list (Fig. 6).

Reasons for WoS Core Collection status result of 'No'	Cases before manual QA	Cases after manual QA
Invalid DOI	7	0
DOI/journal article is peer-reviewed but not indexed in the WoS Core Collection™	130[20]	179
DOI/Journal article not yet indexed in the WoS Core Collection™	NA	46
Journal article is indexed in the WoS Core Collection™ but not linked to its DOI	NA	0
DOI/journal article is not peer-reviewed or excluded from the QA for other reason(s)	NA	40
No clear response from the tool if peer-reviewed or not and included in WoS Core Collection or not	128	0

Table 1: Numbers of cases before and after manual QA when the WoS Core Collection status was confirmed to be 'No'. The disaggregation for different reasons follows a previous explanation[21].

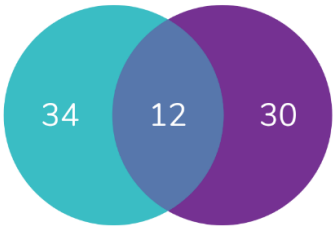

Figure 6: Venn diagram showing the number of WoS Core Collection publications validated by the assessor during the first round (left) and those found by the tool after two months through an independent check (right). In the middle is the 12 publications validated by both.

[20] Only considered in this category are publications found by the Scopus API or found to be peer-reviewed, but non-WoS Core Collection by manual QA. All the other publications not confirmed as peer-reviewed went directly to the QA.

[21] https://repo.mel.cgiar.org/handle/20.500.11766/13115

B. OA status

Within the first round of QA, 1,995 publications (79.5%; Fig. 7) were automatically validated as OA by the tool. After manual validation, the number of OA publications slightly decreased[22] (-1.2%, from 1,995 to 1,971). However, the overall ratio of OA/non-OA remained the same.

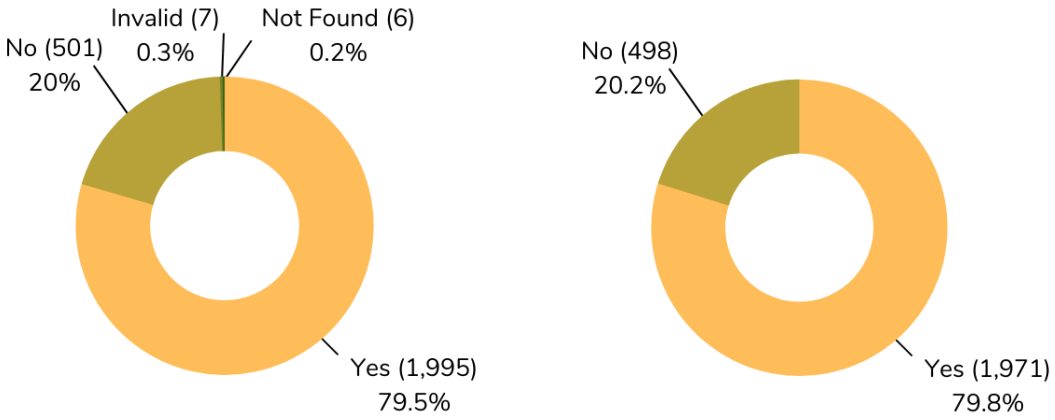

Figure 7: Results of the response of the M-QAP-API regarding the OA status before (left) and after manual QA (right).

As per the WoS Core Collection status, the access status of invalid DOIs and those not found by the tool were resolved by searching the publications using other available metadata (i.e., title, handle) (Table 2).

Reasons for OA status result of 'No'	Cases before manual QA	Cases after manual QA
Invalid DOI	7	0
DOI/journal article recently uploaded in a repository	NA	NA
DOI registered by an Agency that is not CrossRef or DataCite	NA	6
DOI/journal article is not peer-reviewed or excluded from the QA for other reason(s)	NA	40

Table 2: Numbers of cases before and after manual QA when the OA status was confirmed to be 'No'. The disaggregation for different reasons follows a previous explanation[23].

[22] The slight decrease is mainly due to the fact that many non-peer-reviewed journal articles that entered the QA were OA.

[23] https://repo.mel.cgiar.org/handle/20.500.11766/13115

Among the total, five DOIs (0.25%) were registered by the DataCite Registration Agency, followed by the Multilingual European DOI Registration Agency (mEDRA, four cases), the Institute of Scientific and Technical Information of China (ISTIC, one case), and Aititi Inc. (AIRIT, one case)[24]. The latter three agencies are not covered by Unpaywall, meaning the confirmation of the OA status had to be manually performed.

Among the list of OA links retrieved from Unpaywall, 46 (2.4%) were attributable to a CGIAR Center/Repository: 21 articles mentioned the OAR ICRISAT; 17 CGSpace; six the CIFOR Library; and two the IFPRI Library (Appendix C).

Similarly to the WoS Core Collection, the DOIs/journal articles non-OA (Fig. 7) were re-evaluated approximately two months after the QA through the same tool (M-QAP).

The total figure of OA articles slightly improved (+1,1%; from 1,971 to 1,992) —possibly concurrent to the end of the embargo period of some journal articles and/or upload of allowed copies into the repositories. Among the OA links retrieved, MELSpace appears for three articles[25], bringing the total number of articles stored in OA in a CGIAR Center/Repository to 49 (Appendix C).

[24] https://www.doi.org/RA_Coverage.html

[25] At the time of the QA, MELSpace had just submitted the request for being indexed among Unpaywall list of sources, and therefore, it appears among the repository during the final independent check.

Discussion, recommendations, and possible enhancements

The M-QAP-API tool processed more than 2,500 DOIs sent by CGIAR CRPs and Platforms at once, and automatically validated as WoS Core Collection and OA, respectively—almost 90% and 80% of the DOIs (Fig. 8). This has provided a rapid, precise, reproducible, and time- and resource-saving solution for the CGIAR QA process of publications.

Figure 8: Summary of the performance of the M-QAP-API tool.

Thanks to this tool, the manual check was reduced to less than 300 articles, and proved indispensable in resolving and validating DOIs not processed by the tool, invalid DOIs, and all journal articles lacking a DOI. Had the entire quality assurance process for 2,700 publications been performed manually, it would have required roughly 60 days of work[26]. The work of the quality assessors had allowed an increase in the number of peer-reviewed DOI/journal articles not indexed in the WoS Core Collection™, by manually searching the journal's international standard serial number (ISSN) in the WoS Master Journal List (MJL)[27] (Table 2). Moreover, the quality assessors were able to exclude from the submitted list of publications non-peer-reviewed journal articles and publications erroneously added as journal articles (40 in total, e.g., magazine articles). The manual check had an impact on the final percentage of OA publications, which slightly decreased after the QA. Therefore, automatic and manual validation are complementary and should work hand-in-hand throughout the QA.

As observed two months following the first QA, a subset of articles changed their WoS Core Collection and OA statuses over time. For the WoS Core Collection status, this can be mainly attributed to the technical time (up to six weeks[28]) needed by the WoS service to index the articles. Meanwhile, for the OA, articles become OA when released from their embargo (e.g., six, 12 or 24 months after their publication), meaning a copy (e.g., the accepted version) might have become available. Interestingly, the manual check could attribute 34 journal articles as WoS Core Collection that, after two months, were still not found by the WoS API—

[26] Estimation based on the experience of a quality assessor.
[27] https://mjl.clarivate.com/home
[28] Personal communication with the WoS technical support.

confirming the lag time between the publishing of the article and their indexing in the database. For these reasons, it is suggested to employ the tool for periodical checks to monitor changes in the WoS Core Collection and OA statuses, as this allows, for instance, the CGIAR Results Dashboard to be up-to-date on the WoS Core Collection and OA statuses of reported articles. This also applies to Altmetric and F.A.I.R. scores, as their values can change over time. MARLO platform has been successfully applying this approach by automating the update with Altmetric scores to update CGIAR Results Dashboard with the help of the M-QAP-API service in CLARISA. The periodical check through the tool could be programmed quarterly and possibly integrated within repositories in a way that means it stores some data (such as WoS Core Collection) but does not repeat the query and therefore lead to an increase in costs.

To add an additional layer of information, the CGIAR Results Dashboard could use more metadata from the current M-QAP-API and show in the future:
- The data source: i.e., WoS, Scopus, WoS and Scopus, none.
- The percentage of different OA statuses: i.e., green, gold, hybrid, bronze[29].

More metadata could be also added in the future to the M-QAP-API (Appendix A), which would expand the potential of the tool for future applications—including bibliometric analysis and assessment of the Quality of Science—and bring more visibility and impact overall to CGIAR scientific production.

Last but not least, the tool was previously and successfully implemented in MARLO and MEL. In particular, MARLO displays to users the metadata retrieved by the APIs and stores metadata information. The further potential of the tool is that it could be embedded in the repositories to support users, librarians and knowledge management staff in the immediate validation of journal articles, as well as provide a real-time assessment of the scientific impact of publications, and potentially guide decisions on investments when publishing scientific results.

[29] https://support.unpaywall.org/support/solutions/articles/44001777288-what-do-the-types-of-oa-status-green-gold-hybrid-and-bronze-mean-

Disclaimer

The count of journal articles is based on the publications with a DOI reported by each CRPs and Platforms, so it does not consider all publications without a DOI (68; 2.6%), whose WoS Core Collection and/or OA status was validated via manual check during the QA. Moreover, the dataset used for the calculation reflects a specific moment in time and changes that occurred throughout the QA process might have been only in part reflected in the final figures. This refers, for example, to articles removed at a further stage of the QA.

Moreover, during the trial of the tool in early 2021, a total of approximately 30 DOIs were sent to the Clarivate customer support and, therefore, issues of articles not found in WoS were resolved before the start of the QA process.

Acknowledgements

The development and improvements on the M-QAP-API tool were possible thanks to the financial support of the CGIAR SMO, as well as the WoS service subscription.
Enrico Bonaiuti from the CRP on Roots, Tubers and Bananas (RTB) provided the overall design of the process and support to the team.
Moayad Al-Najdawi from Code Land for System Development (CodeObia) developed the script, while the Innovations and Business Development Team at the Alliance of Bioversity International and CIAT implemented the script in MISs and provided feedback. Among the team, Héctor Tobón and Manuel Ricardo Almánzar provided support and the dataset to allow the analysis.
Valentina De Col from the CGIAR Research Program (CRP) on Grain Legumes and Dryland Cereals (GLDC) provided a review and analysis of the data, improvement of APIs' responses, and interaction with external services such as WoS and Unpaywall.
Sara Jani from the International Center for Agricultural Research in Dry Areas (ICARDA) provided an advisory role to the team in her capacity of Publications Repository Manager and metadata curator.
Max Rünzel from CGIAR Advisory Services Shared Secretariat (CAS) provided a critical review of the text.
Diu Seng See and Caroline Yeoh from Clarivate, and Richard Orr from Unpaywall, for their technical support during the testing of the M-QAP-API tool.

Appendix A

Metadata retrieved and retrievable from the different databases through the APIs

Database	Metadata retrieved for the QA	Additional metadata that could be retrieved in the future[30]
Web of Science (Clarivate)	High quality publications covered in the Web of Science (WoS) Core Collection	- Abstract - Author(s) - Author(s) ORCID - Author(s)' keywords - Funding agency and acknowledgement - ISSN/ESSN - Issue - Journal name - Keywords plus - Language - OA status - Organization(s) of the author(s) and address(es) - Pages - Publication date - Publication title - Publication type - Publisher city - Publisher name - Research areas - Volume - WoS category More at the WoS API webpage.
Scopus	Peer-reviewed	- Author(s) - ISSN/ESSN - Journal name - OA status - Pages - Publication date - Publication title - Publication type - Volume More at the Scopus API documentation.

[30] The list is not meant to be exhaustive but only showcase the potential of the APIs.

Unpaywall	OA	- Best OA evidence - Best OA host - Best OA licence - Best OA URL - Best OA version - Genre - Journal in DOJA - Journal is OA - Journal ISSN - Journal name - OA status - Published date - Publisher More at Unpaywall API documentation.
Crossref	DOI Agency	- Agency ID - Author(s) - Author(s) ORCID - Funder - Issue - Licence - Pages - Publication date - Publication title - Publisher - Volume More at Crossref API documentation.

Appendix B

Number of peer-reviewed papers involved in the QA process by CRPs and Platforms (2020 AR)

CRP or Platform	Number of peer-reviewed papers submitted to CLARISA	Peer-reviewed papers after QA	Difference
A4NH	268	263	5
BigData	8	8	0
CCAFS	184	183	1
EiB	0	0	0
Fish	80	80	0
FTA	217	216	1
Gender	1	0	1
Genebank	112	111	1
GLDC	127	119	8
Livestock	232	227	5
Maize	251	251	0
PIM	172	172	0
Rice	277	273	4
RTB	248	246	2
Wheat	264	264	0
WLE	136	130	6
Total	**2,577**	**2,543**	**34**

Appendix C

Links to CGIAR Repositories (2020 AR) for OA journal articles

OAR ICRISAT: http://oar.icrisat.org/

1. http://oar.icrisat.org/11358/1/10.1007_s00122-019-03512-z.pdf
2. http://oar.icrisat.org/11616/1/s00122-020-03563-7-2.pdf
3. http://oar.icrisat.org/11622/1/s10526-020-10015-0.pdf
4. http://oar.icrisat.org/11534/1/2020-Rathnakumar%2C%20GEInteractionsInQTLIntrogressi.pdf
5. http://oar.icrisat.org/11627/1/Mohammed2020_Article_GeneticVariationAndDiversityOf.pdf
6. http://oar.icrisat.org/11624/1/GRACE%20iron%20zinc%20protein%20and%20agronomic%20traits%20Final%201.pdf
7. http://oar.icrisat.org/11766/1/s12038-020-00087-6.pdf
8. http://oar.icrisat.org/11541/1/05_Participatory%20mapping%20_ML%20Techniques.pdf
9. http://oar.icrisat.org/10618/1/Weed%20research%20issues%2C%20challenges%2C%20and%20opportunities%20in%20India.pdf
10. http://oar.icrisat.org/11634/1/2020_Parasai-Sindh%20watershed-Jhansi.pdf
11. http://oar.icrisat.org/11512/1/Climatic%20variability%20Resilience%20framework.pdf
12. http://oar.icrisat.org/11054/1/Residue%20Level%20Paper%202019.pdf
13. http://oar.icrisat.org/11386/1/landscape_positions_dictating_crop_fertilizer_responses_in_wheatbased_farming_systems_of_east_african_highlands.pdf
14. http://oar.icrisat.org/11625/1/SR20007.pdf
15. http://oar.icrisat.org/11355/1/Understanding%20the%20response%20of%20sorghum%20cultivars%20to%20nitrogen%20applications%20in%20the%20semi%20arid%20Nigeria%20using%20the%20agricultural%20production%20systems%20simulator.pdf
16. http://oar.icrisat.org/11360/1/Productivity%20and%20profitability%20of%20maize-legume%20systems%20under%20CA.pdf
17. http://oar.icrisat.org/11705/1/jbaa009.pdf
18. http://oar.icrisat.org/11565/1/Falconnier_et_al_2020_GCB.PDF
19. http://oar.icrisat.org/11785/1/jac.12433.pdf
20. http://oar.icrisat.org/11620/1/plb.13147.pdf
21. http://oar.icrisat.org/11672/1/ESI%20IJE-%2082%283%29%5B23821%5D-pages-1-4%2C147-150.pdf

CGSpace: https://cgspace.cgiar.org/

1. https://hdl.handle.net/10568/102501
2. https://hdl.handle.net/10568/106928
3. https://hdl.handle.net/10568/108107
4. https://hdl.handle.net/10568/109187
5. https://hdl.handle.net/10568/110358
6. https://hdl.handle.net/10568/110611
7. https://hdl.handle.net/10568/105981
8. http://hdl.handle.net/10138/329706
9. https://hdl.handle.net/10568/111686
10. https://hdl.handle.net/10568/111275
11. https://hdl.handle.net/10568/110643
12. https://hdl.handle.net/10568/108643
13. https://hdl.handle.net/10568/105521
14. https://hdl.handle.net/10568/108318
15. https://hdl.handle.net/10568/110357
16. https://hdl.handle.net/10568/110886
17. https://hdl.handle.net/10568/110118

CIFOR Library: https://library.cifor.org/

1. https://www.cifor.org/library/7498
2. https://www.cifor.org/library/7505
3. https://www.cifor.org/library/7649
4. https://www.cifor.org/library/7651
5. https://www.cifor.org/library/7820
6. https://www.cifor.org/publications/pdf_files/articles/AKomarudin2001.pdf

IFPRI Library: https://www.ifpri.org/publication/ifpri-library-and-knowledge-management-website

1. https://www.ifpri.org/cdmref/p15738coll2/id/133559/filename/133770.pdf
2. http://www.ifpri.org/publication/what-intrinsic-value-fertilizer-experimental-value-elicitation-and-decomposition-hill

MELSPACE: https://repo.mel.cgiar.org/

1. https://mel.cgiar.org/reporting/download/hash/158ad45b813fa18727bc283742821965
2. https://mel.cgiar.org/reporting/download/hash/2d26fccf3292956a46018a4d1bff5a91
3. https://mel.cgiar.org/reporting/download/hash/6f73fd6b0ced8983f29b50196a3d0354