

Strengthening Scholarly Publishing at the University of Zambia: Custom OJS Plugins for Metadata Accuracy, DOI Management, and Repository Integration

By

Mubanga Chibesa (2025023035)

Supervisors:

Dr. Lighton Phiri

A report submitted to the University of Zambia in partial fulfilment of the requirements of the Postgraduate Diploma in Computer Science

THE UNIVERSITY OF ZAMBIA

LUSAKA

2025

Declaration

I, Mubanga Chibesa, do hereby declare that the work in this dissertation has not been previously submitted in candidature for any degree. This report is the result of my own work and investigations, except where otherwise stated. Other sources are acknowledged by given explicit references. A complete list of references is given.

Signature.....

Date

Copyright

All rights reserved. No part of this report may be reproduced or stored in any retrieval system or transmitted in any form or by any means, electronic, mechanical photocopying, recording, scanning or otherwise without the prior written permission of the author or the University of Zambia.

©Mubanga Chibesa, 2025.

Certificate of Approval

This report of Mubanga Chibesa has been approved as fulfilling the requirements or partial fulfilment of the requirements for the award of Postgraduate Diploma in Computer Science by the University of Zambia;

Examiner 1	Signature	Date
.....		
Examiner 2	Signature	Date
.....		
Examiner 3	Signature	Date
.....		
Chairperson/Board of Examiners	Signature	Date
.....		
Supervisor	Signature	Date
.....		

Abstract

The growing demand for accurate, interoperable, and easily discoverable scholarly outputs necessitates enhanced publication workflows in institutional repositories and journal management systems. At the University of Zambia, where scholarly publishing is increasingly transitioning to digital platforms, Open Journal Systems (OJS) has emerged as a central tool for journal hosting. However, limitations in automated metadata compliance, DOI integration, and repository deposition pose persistent challenges.

This research presents the conceptualization, development, and deployment of three interlinked OJS plugins designed to enhance the publication pipeline: (1) an Automated Metadata Verification Plugin that validates journal article metadata against institutional and international standards (2) a DOI Deposition Plugin that streamlines Crossref DOI registration and metadata syncing in real time; and (3) a DSpace Export Plugin that packages OJS-published content into DSpace-ready Simple archive bundles for seamless repository ingestion.

The proposed system has been implemented and tested within selected UNZA-hosted journals especially under the JABS editorial workflow. Preliminary results show substantial improvements in metadata completeness, cross-platform visibility, and compliance with repository requirements. Additionally, the plugins promote FAIR (Findable, Accessible, Interoperable, Reusable) principles by ensuring structured and verified data flows across platforms.

By automating routine tasks, reducing human error, and bridging siloed scholarly infrastructure, this plugin suite represents a scalable model for other institutions like the University of Zambia in the Global South seeking to align with international open access standards.

Acknowledgements

I would like to express my heartfelt gratitude to all those who played instrumental roles in the successful completion of this research report. Foremost, I extend my sincere appreciation to my supervisor, Dr. Lighton Phiri, whose unwavering guidance, encouragement, and support consistently challenged me to go beyond my limits and strive for continuous improvement. His understanding, constructive feedback, and confidence in my abilities provided constant motivation throughout this academic journey.

I am also deeply grateful to my family for their unwavering support. I extend my heartfelt thanks to my parents, Christine and Peter Chibesa, for their encouragement, guidance, and steadfast belief in me, as well as to my sisters, Memory and Alice, whose emotional support and uplifting moments contributed significantly to maintaining a positive and focused research experience.

To everyone mentioned, your support and contributions have had a profound impact on this work. I am sincerely thankful for the encouragement, inspiration, and strength you have provided throughout this academic journey.

Table of Contents

Abstract	1
Acknowledgements	2
Table of Contents	3
List of Tables	5
List of Figures	6
List of Abbreviations	7
1. Introduction	8
2. Related Work	8
2.1. The Development of Meta data Extractor Plugin for Open Journal System	8
2.1.1. Metadata Quality in Digital Repositories: A Survey of the Current State of the Art	
2.1.2 Custom OJS Metadata Extractor	8
2.2. System Integration and Interoperability in Scholarly Publishing	8
2.2.1. OJS and DSpace Integration	8
3. Methodology	8
3.1.Theoretical Grounding	8
3.2. Metadata Quality Gaps And Workflow Inefficiencies For Journal Publishing At The University Of Zambia	8
3.3.Design And Implementation Of Modular Plugin Suite To Automate Metadata Verification, Doi Registration, And Repository Export	9
4. Conclusion	10
References	10

List of Tables

Table 1: Extracted Metadata fields

Table 2: Plugin directory/file structure

List of Figures

Figure 1: CRISP-DM Model

Figure 2: Visibility of JABS across academic databases

Figure 3: Visibility of JONAS across academic databases

Figure 4: Visibility of JLSS across academic databases

Figure 5: Dspace export plugin user interface

Figure 6: Directory Structure of the DSpace SAF Batch Export

Figure 7: Directory Structure of a single article in DSpace SAF

List of Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
CRISP-DM	Cross Industry Standard Process for Data Mining
UNZA	The University of Zambia
DOI	Digital Object Identifier
OJS	Open Journal Systems
SAF	Simple Archive Format

CHAPTER ONE

1.1 INTRODUCTION

The digital transformation of scholarly publishing has reshaped how academic institutions produce, disseminate, and preserve research outputs. As universities increasingly adopt open access platforms and institutional repositories, the demand for accurate, interoperable, and discoverable metadata has become central to the visibility and impact of scholarly work. At the University of Zambia (UNZA), Open Journal Systems (OJS) has emerged as the primary platform for hosting academic journals, offering a flexible and scalable solution for editorial management. However, despite its widespread adoption, OJS implementations at UNZA face persistent challenges related to metadata quality, DOI registration, and repository integration.

These limitations are not unique to UNZA. Across many African institutions, scholarly publishing workflows remain fragmented, with manual metadata entry, inconsistent standards compliance, and siloed infrastructure undermining the discoverability and long-term preservation of research. The absence of automated tools to validate metadata, register DOIs in real time, and export content to institutional repositories contributes to low indexing rates in global databases such as DOAJ, Crossref, and OpenAlex. This, in turn, affects citation visibility, academic recognition, and alignment with international open science principles.

In response to these challenges, this research presents the conceptualization, development, and deployment of a suite of three custom plugins for OJS, each designed to address a specific gap in the scholarly publishing pipeline at UNZA:

- **Automated Metadata Verification Plugin:** Validates article metadata against both institutional and international standards, ensuring consistency, completeness, and machine-readability.
- **DOI Deposition Plugin:** Streamlines the process of DOI registration and metadata synchronization with Crossref, enabling real-time updates and reducing manual errors.
- **DSpace Export Plugin:** Packages published content into DSpace-ready Simple Archive Format (SAF) bundles for seamless ingestion into UNZA's institutional repository.

These plugins are designed to operate independently yet complement one another, forming a modular system that enhances the integrity and interoperability of scholarly outputs. The development process was informed by an empirical analysis of metadata completeness and online visibility across three UNZA-hosted journals, i.e JABS, JONAS, and JLSS reveal significant gaps in indexing and metadata dissemination. None of the journals were listed in DOAJ or AJOL, and inconsistencies were observed in Crossref, OpenAlex and Google Scholar coverage. These findings underscore the urgent need for automated solutions that align with the FAIR (Findable, Accessible, Interoperable, Reusable) principles of scholarly communication .

By automating routine editorial tasks, reducing human error, and bridging disconnected systems, this research aims to strengthen the scholarly publishing infrastructure at UNZA. The plugin suite not only improves metadata accuracy and repository compliance but also offers a scalable model for other institutions seeking to modernize their publishing workflows. Ultimately, this work contributes to the broader goal of enhancing the visibility, accessibility, and impact of African scholarship in the global research ecosystem.

1.1.1 The Role of Metadata in Scholarly Communication

Metadata has emerged as a cornerstone of scholarly communication, serving as the connective tissue that links research outputs to global discovery systems. Defined as “data about data,” metadata encapsulates essential descriptors such as titles, author affiliations, abstracts, keywords, and licensing [6]. This data enables search engines, repositories, and indexing services to retrieve, categorize, and rank scholarly content. Its strategic importance is underscored by the FAIR principles, which advocate for metadata that is Findable, Accessible, Interoperable, and Reusable [10].

High-quality metadata not only enhances visibility and citation impact but also ensures compliance with international standards such as Dublin Core [8], MARC, MODS, and OAI-PMH2. These standards facilitate interoperability across platforms like Crossref, Google Scholar, and institutional repositories, allowing for seamless data exchange and long-term preservation. Studies have shown that enriched metadata significantly boosts discoverability [1], Crossref, for instance, relies on over 15 metadata tags to accurately index journal articles [3].

Yet, metadata quality remains uneven, particularly in resource-constrained environments. Inaccurate or incomplete metadata can result in broken links, failed indexing, and diminished user experience. At institutions like the University of Zambia, where editorial workflows are often manual and fragmented, the lack of automated metadata validation tools has led to inconsistencies that undermine the visibility of local scholarship. Addressing this gap is not merely a technical fix, it is a strategic imperative for elevating the institution’s research profile and aligning with global best practices.

1.1.2 Open Access and Equity in African Scholarship

Open access (OA) was founded on the philanthropic ideal of democratizing knowledge, making scholarly outputs freely available to all, regardless of geography or economic status [4]. For African institutions, this vision holds transformative potential: it promises to dismantle entrenched barriers to participation in global research and foster inclusive knowledge ecosystems. However, the reality has often diverged from this promise. Commercialization of OA through article processing charges (APCs) and systemic biases in editorial and peer review processes have disproportionately marginalized scholars from the Global South.

African scholarship continues to face challenges in visibility, representation, and infrastructure [7]. Journals hosted on platforms like OJS may struggle to meet the metadata and indexing requirements of global directories such as DOAJ and AJOL, leading to exclusion from key discovery channels. This exclusion perpetuates a skewed research

landscape where African voices are underrepresented and undervalued.

In response, initiatives like the African Platform for Open Scholarship (APOS) and the African Principles for Open Access advocate for a socially just OA model one that emphasizes equity, multilingualism, and indigenous knowledge systems. These frameworks call for interoperable open standards, inclusive editorial practices, and institutional investment in open infrastructure. By aligning with these principles, the University of Zambia's efforts to enhance its scholarly publishing workflows through custom OJS plugins represent a meaningful step toward reclaiming agency in the global research ecosystem.

1.2 Background

The University of Zambia (UNZA) has long served as a leading academic institution in Southern Africa, producing research across diverse disciplines. In 2019, UNZA undertook a strategic transition from print-based scholarly publishing to digital platforms, adopting Open Journal Systems (OJS) as its primary journal management system. This move aligned with global trends toward open access and digital dissemination, offering the potential to streamline editorial workflows, improve accessibility, and enhance the visibility of institutional research.

Currently, UNZA hosts fifteen journals on its OJS platform, spanning fields such as agriculture, law, biomedical sciences, and natural sciences. While the adoption of OJS marked a significant step forward, the transition has revealed persistent structural and operational challenges. Editorial workflows remain fragmented, with delays in peer review, publication, and indexing. Metadata quality is inconsistent, often lacking the completeness and standardization required for effective indexing in global databases. Moreover, the absence of formal editorial guidelines and standardized processes has led to variability in journal quality and discoverability.

These challenges are compounded by limited automation and integration across systems. For instance, DOI registration is not consistently implemented, and there is no direct linkage between OJS and UNZA's institutional repository (DSpace), resulting in siloed content and missed opportunities for long-term preservation. As a result, UNZA's journals struggle to meet international publishing standards and remain underrepresented in key indexing services such as DOAJ, AJOL, and Crossref.

Addressing these issues is critical not only for improving operational efficiency but also for elevating the global profile of UNZA's scholarly output.

1.3 Problem Statement

Despite the adoption of Open Journal Systems (OJS) in 2019, scholarly publishing at the University of Zambia continues to face systemic challenges that hinder the visibility, accessibility, and impact of its research outputs. The fifteen journals currently hosted on the platform are constrained by workflow inefficiencies, poor metadata quality, and the absence of standardized editorial practices. These limitations result in delayed publications, inconsistent indexing, and reduced discoverability in global academic databases.

Furthermore, the lack of automated tools for metadata verification, DOI registration, and repository integration has created operational silos that fragment the publication pipeline. Without formal editorial guidelines or institutional policies to govern digital publishing, journal management remains ad hoc and vulnerable to human error. This undermines the institution's ability to comply with international standards and diminishes its presence in open access directories and citation indexes.

To address these challenges, there is a pressing need for a modular, scalable solution that enhances the technical and editorial infrastructure of UNZA's publishing ecosystem. The development of custom OJS plugins focused on metadata validation, DOI deposition, and repository export offers a strategic pathway to modernize workflows, improve metadata interoperability, and strengthen the institution's contribution to global scholarship.

1.4 Objectives

The primary objective of this research is to enhance the efficiency, interoperability, and discoverability of scholarly publishing at the University of Zambia by developing and evaluating a modular plugin suite for Open Journal Systems (OJS) that addresses metadata quality gaps, workflow inefficiencies, and repository integration challenges.

1.4.1 Specific objectives

- What are the existing metadata quality gaps and workflow inefficiencies within UNZA's current OJS-based journal publishing system?
- How can a modular plugin suite be designed and implemented to automate metadata verification, DOI registration, and repository export in alignment with international publishing standards?
- What is the impact of the developed plugins on the performance, usability, and overall effectiveness of editorial workflows, metadata quality, and integration with the institutional repository?

1.4.2 Research questions

- What are the existing metadata quality gaps and workflow inefficiencies within UNZA's current OJS-based journal publishing system?
- How can a modular plugin suite be designed and implemented to automate metadata verification, DOI registration, and repository export in alignment with international publishing standards?
- What is the impact of the developed plugins on the performance, usability, and overall effectiveness of editorial workflows, metadata quality, and integration with the institutional repository?

CHAPTER TWO

2. RELATED WORK

The automated publishing pipeline in the Open Journal System (OJS) hinges on a sequence of interrelated processes that guarantee metadata accuracy, persistent identifiers, seamless repository integration, and intuitive editorial workflows. This literature review synthesizes current research into four thematic areas, each directly supporting the study's objectives.

2.1 Metadata Quality and Automated Extraction

2.1.1 Metadata Quality in Digital Repositories

Metadata quality remains a foundational concern in the effective management of digital repositories. A comprehensive survey by Park highlights four core dimensions of metadata quality: accuracy, consistency, completeness, and interoperability [11]. Accuracy relates to the correctness of metadata values, including spelling, formatting, and semantic clarity. Completeness concerns the presence of all required metadata elements necessary for indexing, retrieval, and preservation. Consistency refers to the uniform application of metadata standards across records, while interoperability ensures that metadata can be exchanged and interpreted across heterogeneous systems and platforms. Together, these dimensions underpin the discoverability, accessibility, and long-term preservation of digital scholarly resources.

Despite their importance, these dimensions are frequently compromised in practice. Park identifies human error in manual metadata entry, inconsistent application of standards such as Dublin Core, and the absence of automated validation mechanisms as persistent challenges [11]. These issues are particularly pronounced in large-scale digitization initiatives where metadata is produced by multiple contributors with varying levels of technical expertise. The lack of standardized validation tools forces repositories to rely heavily on manual quality control processes, which are inherently error-prone and difficult to scale.

Subsequent studies reinforce these findings. Zeng and Qin emphasize the role of metadata education and professional training in mitigating human-induced inconsistencies [12], while Greenberg advocates for structured metadata auditing frameworks that enable systematic assessment and continuous improvement of metadata quality [13]. Collectively, this body of literature underscores the need for both technical and organizational interventions to address persistent metadata quality problems.

A notable limitation of Park's analysis is the absence of discussion on persistent identifiers (PIDs), particularly Digital Object Identifiers (DOIs), and their role in strengthening metadata reliability. This omission is significant for institutions such as the University of Zambia (UNZA), where DOI integration has the potential to substantially improve metadata persistence, citation tracking, and interoperability. DOIs provide stable, globally unique identifiers that remain resolvable even when the physical location of digital objects changes. Their adoption facilitates reliable citation, cross-platform linking, and integration with discovery services and aggregators. Prior studies demonstrate that DOI implementation enhances repository visibility and supports more robust scholarly communication

infrastructures [14, 15].

Beyond persistent identifiers, recent technological developments offer additional solutions to longstanding metadata challenges. Automated metadata generation and validation tools, including those based on natural language processing and machine learning techniques, can reduce reliance on manual entry while improving accuracy and efficiency. Semantic web and linked data technologies further enhance interoperability by enabling richer contextual relationships between metadata elements. Frameworks such as the Metadata Quality Framework (MQF) introduce quantifiable metrics for monitoring metadata performance, allowing repositories to systematically evaluate and refine their metadata practices over time.

2.1.2 Custom OJS Metadata Extractor

The inconsistencies in metadata quality identified in the literature are largely attributable to human error in manual metadata entry workflows [11]. This challenge is particularly acute in resource-constrained environments, including institutions with limited technical staffing, multilingual contributors, or diverse editorial boards. In such contexts, manual metadata population within Open Journal Systems (OJS) introduces a significant cognitive burden for authors and editors, increasing the likelihood of inaccuracies and inconsistent adherence to schema standards such as Dublin Core, NLM XML, or JATS. These deficiencies directly affect article discoverability, indexing in platforms such as Google Scholar and Web of Science, and downstream system integration [4].

The Rizqi et al. metadata extractor plugin directly addresses these operational challenges by automating metadata extraction within the OJS submission workflow. The plugin employs a rule-based extraction approach that processes uploaded article files, with initial support limited to DOCX formats to ensure optimal structural consistency. Using PHP's ZipArchive functionality, the plugin parses the underlying XML structure of submitted documents and applies heuristic rules derived from common scholarly publishing conventions. For example, the largest bolded text is identified as the article title, centrally aligned bold text is interpreted as author names, affiliations are detected using institutional keywords, and email addresses are extracted using syntactic pattern matching. Semantic anchors such as section headers ("Abstract," "Keywords," or discipline-specific equivalents) further guide the extraction process.

The plugin functions as an additive step within the OJS submission pipeline. Upon initial article upload, metadata extraction is automatically triggered, and the extracted values are presented to editors for verification and completion. This design preserves human oversight while substantially reducing manual data entry and the associated risk of error. Empirical evaluation reports an extraction accuracy rate of approximately 95%, alongside a 3.72-fold reduction in metadata population time compared to fully manual workflows. Remaining extraction failures are primarily linked to deviations from prescribed author templates, non-standardized affiliation naming, and irregular email formatting.

2.2 System Integration and Interoperability in Scholarly Publishing

2.2.1 OJS and DSpace Integration

System integration and interoperability have become increasingly critical as academic institutions transition toward fully digital publishing ecosystems. At UNZA, the adoption of Open Journal Systems in 2019 marked a strategic move toward open access dissemination. However, the absence of seamless integration between OJS and the institutional repository (DSpace), combined with inconsistent DOI registration practices, has resulted in fragmented workflows that limit discoverability, preservation, and compliance with international publishing standards.

Interoperability in scholarly publishing refers to the capacity of heterogeneous systems to exchange, interpret, and preserve metadata using shared standards and protocols. Xia and Sun argue that institutional repositories must support open interoperability frameworks such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to enable effective metadata exchange [16]. In the case of UNZA, the lack of automated metadata transfer between OJS and DSpace restricts the visibility of published research and undermines long-term preservation strategies. Empirical evidence from other institutions demonstrates the feasibility and benefits of such integration. Villarreal et al. document a successful implementation of custom metadata bridges between OJS and DSpace at Universidad Nacional de La Plata, resulting in improved indexing and repository coherence [17].

Persistent identifiers, particularly DOIs, play a central role in reinforcing interoperability and citation reliability. DOIs provide stable, actionable references that support consistent linking across platforms and enhance the integrity of citation metrics. Brase demonstrates that DOI adoption significantly improves discoverability and metadata robustness within scholarly infrastructures [14]. However, manual DOI registration workflows remain vulnerable to errors, delays, and schema non-compliance. Macgregor highlights the advantages of automated DOI deposition using Crossref APIs, noting reductions in error rates and improved metadata consistency [18].

The architectural design of publishing platforms further influences interoperability outcomes. Modular, plugin-based architectures offer scalable and flexible solutions that allow institutions to enhance functionality without disrupting core systems. Industry analyses suggest that modular publishing infrastructures enable targeted workflow improvements, reduce system-wide risk, and accelerate development cycles [19, 20]. Within OJS, such plugins can support specialized tasks including metadata validation, DOI registration, and repository export, enabling institutions to adapt their publishing workflows to evolving standards and operational needs.

Equally important is adherence to metadata standards and schema alignment. Consistent application of metadata schemas such as Dublin Core or JATS XML is essential for ensuring that exported records remain interoperable and machine-readable across repositories and indexing services. Prior studies emphasize that automated schema validation and conversion tools can significantly mitigate human error while enhancing metadata consistency and usability [11, 13].

In summary, system integration and interoperability are foundational to effective scholarly publishing. For UNZA, adopting a modular plugin-based approach that automates metadata verification, DOI registration, and repository export represents a strategic pathway toward improved operational efficiency, metadata quality, and global visibility.

2.2.2 DOI Registration and Crossref Integration

Digital Object Identifiers are central to the discoverability, citation, and long-term accessibility of scholarly content. Their integration into publishing platforms such as OJS is not merely a technical enhancement but a strategic requirement for institutions seeking alignment with international publishing standards and the FAIR principles—Findable, Accessible, Interoperable, and Reusable. For institutions in the Global South, DOI adoption also contributes to greater equity in scholarly visibility and participation within the global research ecosystem.

Crossref, as the leading DOI registration agency for scholarly publishing, provides standardized metadata schemas and services that support DOI registration, citation tracking, versioning, and metadata dissemination. In modern versions of OJS (3.x and above), the Crossref XML Export plugin enables automated DOI assignment and metadata deposition. However, such functionality is absent in legacy OJS installations still in use at UNZA, necessitating manual and fragmented DOI workflows that are prone to error and delay.

This limitation underscores the need for a custom, automated DOI depositor plugin tailored for legacy OJS environments. By automating metadata validation, DOI registration, and synchronization with institutional repositories, such a plugin would reduce human error, improve metadata quality, and ensure timely DOI assignment. More broadly, investment in this infrastructure aligns with UNZA's strategic goals of modernizing scholarly communication, strengthening open access publishing, and enhancing global research visibility.

3. METHODOLOGY

3.1 Theoretical Grounding

The theoretical framework guiding this research is the CRISP DM methodology. This framework highlights approaches in data mining by segmenting data mining projects into six distinct phases. These phases are business understanding, data understanding, data preparation, modelling, evaluation and deployment[5]. Each of these phases help project organisations and machine learning practitioners to have appropriate guidance throughout the whole life cycle of the project in order for them to meet the business requirements.

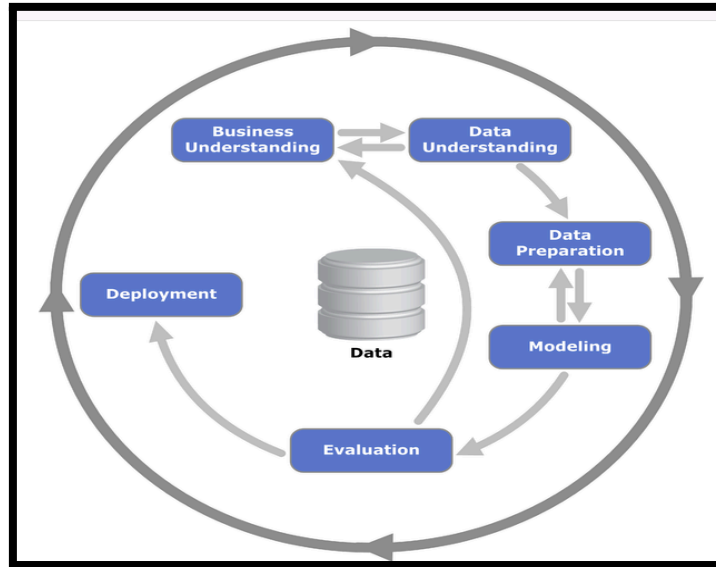


Figure 1. CRISP DM Model

Business understanding: This involves understanding the project objectives and requirements. It is in this phase that planning is conducted and this is important to enable us to carve out what is necessary and what is not necessary to our project.

Data understanding: Collecting and exploring initial data to familiarise with its content, quality and structure.

Data preparation: This involves cleaning, transforming, and integrating data to prepare it for modelling.

Modelling: Involves Selecting and applying appropriate model techniques to build and evaluate predictive or descriptive models.

Evaluation: At this phase assessing model performance to ensure it meets business goals and requirements.

Deployment: Deploying the model into production and monitoring its performance in its usable setting.

3.2. METADATA QUALITY GAPS AND WORKFLOW INEFFICIENCIES FOR JOURNAL PUBLISHING AT THE UNIVERSITY OF ZAMBIA

3.2.1 Research Design

This study adopted a quantitative research design, employing a computational Data Mining Approach to conduct an empirical analysis of the online visibility of Journal of Agricultural and Biomedical Sciences (JABS). The extract structured bibliographic information from Google Scholar, Crossref and Archive scholar. The aim is to systematically retrieve and analyze metadata related to journal articles published in or affiliated with the Journal of

Agricultural and Biomedical Sciences (JABS), Journal of Law and Social Sciences (JLSS) and Journal of Natural and Applied Sciences (JONAS) and analyze the percentage of these articles that are indexed on Google scholar, Crossref, Directorate of Open Access Journals (DOAJ), African Journals Online (AJOL) and Archive Scholar.

3.2.2 The Visibility problem

The poor visibility of the University of Zambia Journal articles can be attributed to metadata quality gaps as well as editorial workflow inefficiencies. The former is largely a product of the latter in that there is no dedicated editorial support at the institution. In order to ascertain the extent of the metadata quality gaps in the University of Zambia Journals, an empirical analysis was conducted with a key focus on the University's three most consistent journals namely; The Journal of Agricultural and Biomedical Sciences (JABS), Journal of Law and Social Sciences (JLSS), as well as the Journal of Natural and Applied Sciences (JONAS). Articles were extracted from each of the three journals and the extracted articles were used as the ground truth for comparison against other scholarly databases in order to understand the threshold of indexed articles as well as any metadata inconsistencies across scholarly databases.

Data Collection

I. Google Scholar

To assess the journals online visibility, we used the scholarly python library, which interfaces with Google Scholar to retrieve metadata about scholarly articles. A search string combining the journal name and its acronym was used, e.g:

search_term = "Journal of Agricultural and Biomedical Sciences" "JABS"

This search phrase is passed into the Scholarly API to query Google Scholar for potentially relevant publications. The collected data was stored in a python dictionary and then converted into a pandas dataframe with the final dataset being stored as a CSV file for further bibliometric analysis. The same process was followed for all the three other journals. Table 1 below shows the extracted metadata fields:

Table 1.Extracted Metadata fields

Variable	Description
Title	Article title
Authors	List of authors
Year	Year of publication
Journal	Source journal
Abstract	Summary of the article content (if available)
Cited by	Number of times the article has been cited
Link	Direct or indirect URL to the article
Subdomain	Host or domain providing access to the article

II. Crossref

After data was harvested from google scholar, we shifted our focus to crossref and we extracted metadata using the crossref rest API with the University of Zambia DOI prefix “10.53974”. The script harvested all articles on crossref assigned with the DOI prefix. The number of records specified in the script was “500” but only 241 articles were retrieved proving that only 241 records were indexed on crossref with the University of Zambia’s DOI prefix. Metadata fields such as title, DOI, publication date, author name were all retrieved and structured against each record. The records were saved into a csv file using pandas for further analysis.

III. OpenAlex

The goal was to collect metadata of scholarly articles associated with the three journals, i.e. JABS, JONAS and JLSS using the OpenAlex API, and save the results into a structured CSV file for further analysis. Each journal was queried individually using its ISSN (preferred) or Journal title (fallback). Using the ISSN is the most reliable approach, as OpenAlex indexes journals with strong metadata tied to ISSNs.

3.2.3 COMPARATIVE ANALYSIS

After this data was successfully collected, a comparative analysis was carried out to ascertain the number of published articles from JABS, JONAS and JLSS that were indexed in scholarly databases and to identify the metadata quality gaps. The analysis revealed that out of the 101 articles under the Journal of Agricultural and Biomedical Sciences that were extracted from the University of Zambia Journals platform, 92% of the articles were indexed on crossref, whereas 99% of the articles were indexed on OpenAlex and only 20% were indexed on Archive Scholar and 98% on Google scholar. None of the articles from the Journal of Agricultural and Biomedical Sciences were indexed on the Directorate of Open Access Journals (DOAJ) and African Journals Online (AJOL).

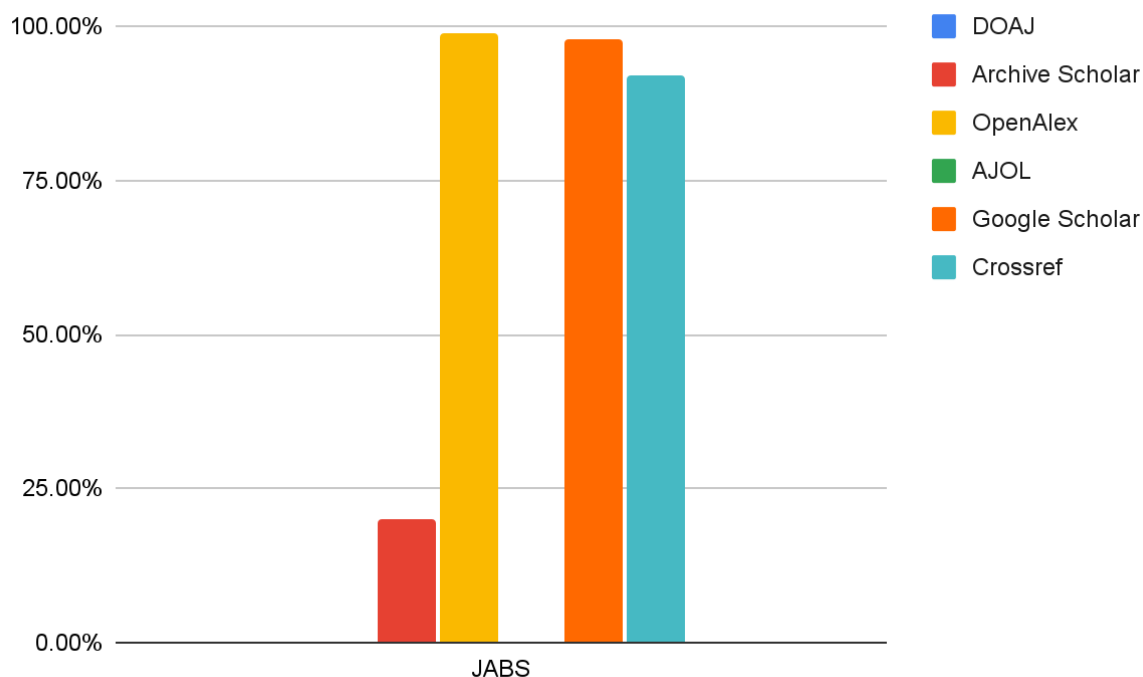


Figure 2. Visibility of JABS across academic databases

Similarly, the study revealed that out of the fifty five articles (55) articles from the Journal of Natural and Applied Sciences (JONAS), 58% were indexed on Open Alex, whereas 53% were indexed on Google Scholar and lastly 72% were indexed on Crossref. The journal had no visibility on platforms such as AJOL, DOAJ and Archive scholar.

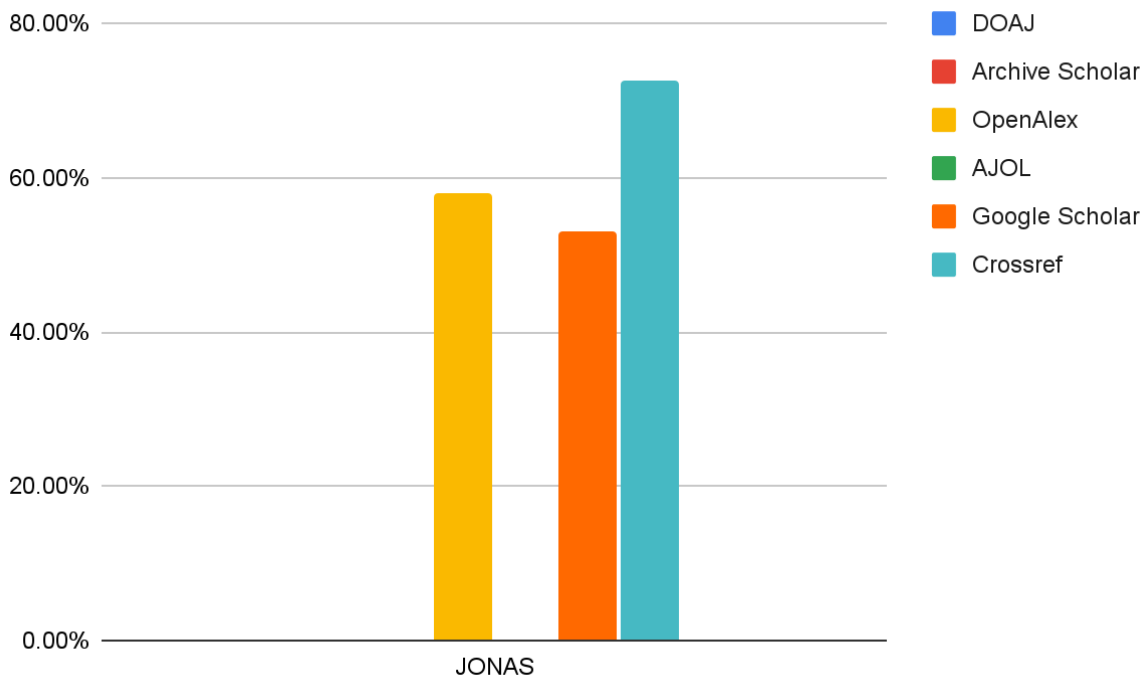


Figure 3. Visibility of JONAS across academic databases

Lastly, our Study revealed that out of the 88 articles from the Journal of Law and Social Sciences, only 48% were indexed on Google scholar, with 50% on Crossref and 91% on OpenAlex. The journal had no articles indexed on AJOL, DOAJ and Archive scholar as seen in the chart below.

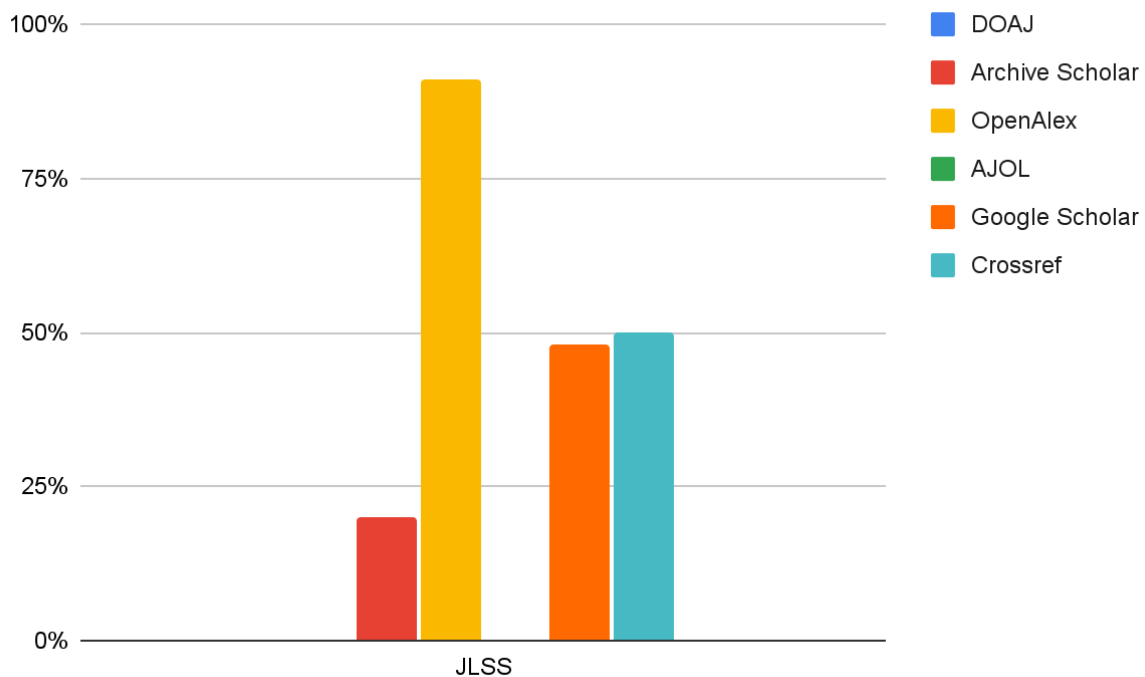


Figure 4. Visibility of JLSS across academic databases

Journals like the Journal of Agricultural and Biomedical Sciences may show decent publishing statistics, yet the visibility of their articles is still affected by other factors. For example, despite 99% of the Journal's articles being indexed on OpenAlex, only a small percentage of that are correctly attributed to the University of Zambia. This stark difference highlights a significant problem with the quality of metadata for these articles across databases.

Inconsistencies in metadata quality do not just affect the visibility of individual articles, they also harm the reputation and ranking of institutions. For instance, when ranking entities such as Edurank query the availability of research output from educational institutions on platforms like OpenAlex, an institution like the University of Zambia is bound to be at a disadvantage due to the misrepresentation of its true number of attributed articles.

What begins as a simple publishing mistake such as an uncaptured author name or an unrecorded affiliation can lead to a much larger problem affecting the visibility of an institution's entire research output. This analysis demonstrates the critical need for automated metadata verification tools within publishing systems like Open Journal Systems (OJS). Such tools would ensure metadata quality adheres to standards like Dublin Core, directly improving the visibility and accurate attribution of institutional research output.

3.3. DESIGN AND IMPLEMENTATION OF MODULAR PLUGIN SUITE TO AUTOMATE METADATA VERIFICATION, DOI REGISTRATION, AND REPOSITORY EXPORT

3.3.1 The DSpace Export Plugin: An Open Journal Systems Extension for Interoperability

The DSpace Export Plugin is an Open Journal Systems (OJS) extension developed to automate the packaging and exportation of published journal articles into the DSpace Simple Archive Format (SAF). This plugin was built to facilitate seamless interoperability between OJS and institutional repositories running DSpace, thereby significantly supporting open access dissemination and metadata standardization (Public Knowledge Project, 2024; Smith et al., 2023).

3.3.2 System Environment and Tools

The plugin was developed and tested within an OJS 3.3.0-15 environment installed on an Ubuntu 24.04 LTS server stack. The supporting software environment included Apache 2.4, PHP 8.2, and MySQL 8.0. Development was primarily conducted using VS Code. Core technologies leveraged in the plugin's development included:

- PHP (for plugin logic and backend integration)
- Smarty templating (for user interface rendering)
- JavaScript (jQuery) (for dynamic UI interactions)
- ZIPArchive (for efficient packaging of article data)
- OJS Plugin API (for foundational system integration)
- GitHub (for version control)

3.3.3 Plugin Architecture

The DSpace Export Plugin adheres to the OJS ImportExportPlugin class structure as defined in the OJS plugin API (Public Knowledge Project, 2024). This adherence ensures full compatibility and stability within the OJS ecosystem. The implementation comprises the following key files and directories:

Table 2. Plugin directory/file structure

File/Directory	Role and Function

DspaceExportPlugin.inc.php	The core file responsible for plugin registration with OJS and defining primary system routing.
DspaceExportHandler.inc.php	The Controller that manages the administrative user interface (UI) and executes the core export logic.
templates/index.tpl	The Smarty template used for rendering the layout and interactive elements of the administrative UI.
locale/en_US/locale.xml	An XML file containing all interface strings and localization data for the English (US) language.
README.md	Documentation providing essential information, including installation instructions, usage details, and developer notes.

3.3.4 Design and Implementation Process

The design phase was guided by several primary requirements necessary for a fully functional and compliant export utility:

1. To allow administrators to view all published articles within the journal.
2. To enable selection of specific articles for targeted export.
3. To accurately capture each article's metadata, including title, authors, affiliations, and associated galley files.
4. To generate a ZIP archive structured precisely according to the DSpace Simple Archive Format (SAF) standard.

3.3.5 Metadata Extraction

The plugin systematically leverages the OJS data access layer to retrieve comprehensive article metadata from several key DAO (Data Access Object) classes:

- PublishedArticleDAO: Retrieved article-level metadata (title, abstract, journal).
- AuthorDAO: Extracted author names and institutional affiliations.
- GalleyDAO: Obtained file paths for downloadable content (e.g., PDF, XML).

For clarity and compliance with metadata readability guidelines (OpenAIRE, 2023), each author's name is displayed alongside their institutional affiliation within the user interface.

3.3.6 User Interface Design

The administrative interface was implemented using OJS's native Bootstrap and Smarty UI components to ensure a consistent look and feel. Enhancements were introduced to optimize interactivity and user experience: a search bar and filter button for efficient article lookup, dynamic checkboxes for streamlined multi-selection, a "Select All" and "Export Selected" buttons for batch export functionality and lastly, author institution pairs displayed in a nested format to maximize clarity.

The design conforms strictly to PKP UI Guidelines (2022), maintaining compatibility with the OJS theme and adhering to role-based permission standards.

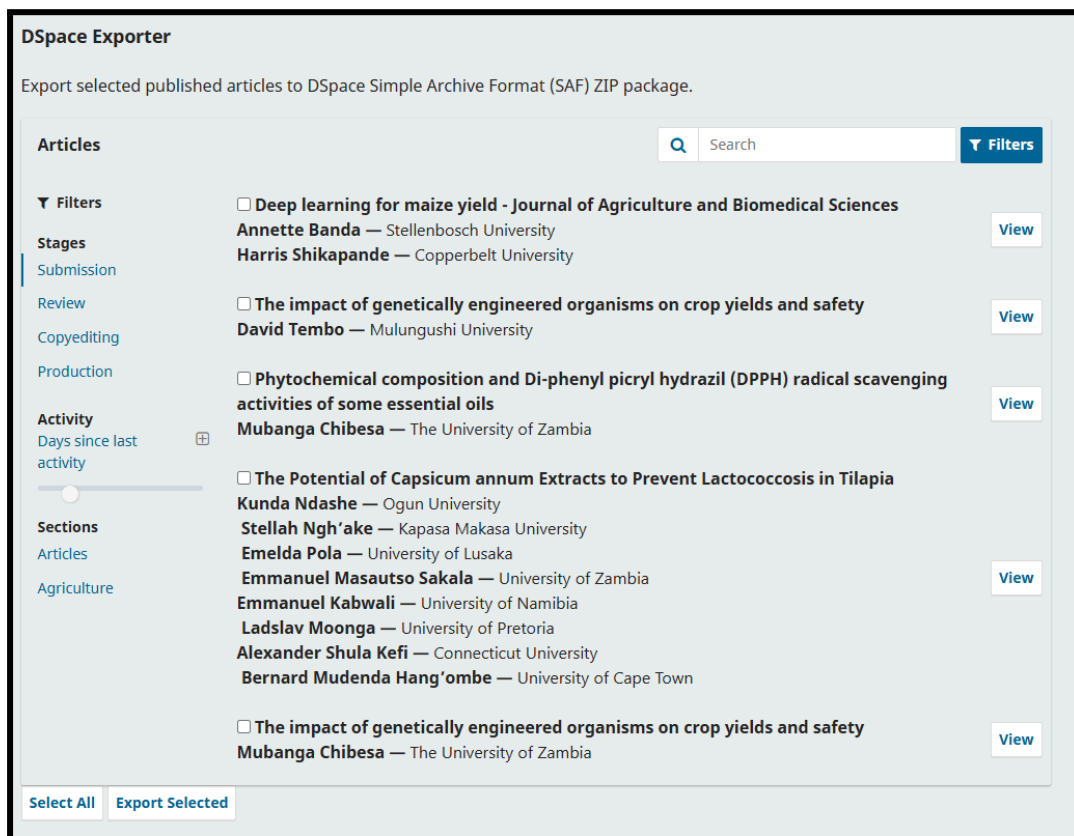


Figure 5. Example format displayed in the interface

Upon completion of the export process, the plugin generates a compressed archive containing a clearly defined hierarchical file structure for each exported article, compliant with the DSpace Simple Archive Format (SAF). This structure ensures that DSpace can ingest the metadata and content files correctly.

The resulting .zip file contains an /export/ root directory, within which each article is assigned its own numbered subdirectory (e.g., article_001/). Each individual article directory includes the following required components:

- **dublin_core.xml:** This is the mandatory metadata file, containing structured elements (such as title, creator, and identifier) compliant with the Dublin Core standard for bibliographic recordkeeping.
- **contents:** A manifest file that lists the file names contained within the directory, ensuring all associated content is tracked for import.
- **article.pdf (and other relevant galley files):** The primary content file (bitstream) of the journal article, such as the final PDF ready for repository dissemination.

This standardized architecture allows for batch processing and automated transfer of research output, directly facilitating institutional repository population.

Name	Size	^Packed	Type	Modified	CRC32
..			File folder		
item_000	6,935	4,611	File folder		
item_002	154,085	146,117	File folder		
item_003	357,210	336,360	File folder		
item_004	638,227	604,800	File folder		
item_001	937,974	706,572	File folder		

Figure 6: Directory Structure of the DSpace SAF Batch Export

Name	Size	^Packed	Type	Modified	CRC32
..			File folder		
contents	10	10	File	10/21/2025 10:...	D73D5E3A
dublin_core.xml	554	266	Microsoft Edge HT...	10/21/2025 10:...	DC4BE482
file1.pdf	6,371	4,335	PDF File	10/21/2025 10:...	F48FF747

Figure 7: Directory Structure of a Single Article in DSpace SAF

3.3.7 Integration and Maintenance

The plugin was seamlessly integrated into the OJS Tools > Import/Export module, making it accessible to journal managers and administrators. Access is strictly governed by OJS’s established role-based access control (RBAC) model. This integration facilitates maintenance

and ensures that the tool is managed and used by personnel with the appropriate administrative privileges.

3.4 Automated Metadata Verifier and Injector Plugin

The Automated Metadata Verifier and Injector Plugin constitutes a central technical component of the proposed enhancement framework for scholarly publishing workflows within Open Journal Systems (OJS). The plugin was developed to address recurring deficiencies in metadata accuracy, completeness, and standards compliance that arise from predominantly manual editorial processes. Such deficiencies directly undermine article discoverability, DOI registration success, interoperability with indexing platforms, and long-term repository preservation.

At a functional level, the plugin is implemented as a custom `ImportExportPlugin`, ensuring native integration with OJS version 3.3.0.15 and adherence to its plugin lifecycle and access control mechanisms. The plugin is structured within the OJS directory hierarchy under `plugins/importexport/metadataVerifier/` and follows established OJS conventions to facilitate maintainability and extensibility. Core components include the primary plugin class (`MetadataVerifierPlugin.inc.php`), which manages registration, authorization, and menu integration; a dedicated handler (`MetadataVerifierHandler.inc.php`) responsible for backend logic, AJAX requests, and validation workflows; Smarty-based templates for user interaction and reporting; locale files to support internationalisation; and auxiliary configuration files defining validation rules and logging behaviour.

The verification process is designed around a hybrid validation architecture. Where an OpenRouter API key is configured via the plugin's settings interface, the system invokes a large language model to perform document-level analysis of the submitted article PDF. In this mode, the plugin extracts and semantically interprets key bibliographic elements—including the article title, author names, institutional affiliations, abstract, keywords, language, and publication identifiers—and compares them against the corresponding metadata fields stored in OJS. This semantic comparison enables the detection of nuanced inconsistencies that extend beyond syntactic errors, such as partial author listings, reordered names, missing affiliations, or abstracts that diverge substantively from the final manuscript. To ensure resilience and operational continuity, the plugin incorporates a fallback mechanism based on regex-driven extraction and rule-based validation, which is automatically engaged when no API key is provided or when external model access is unavailable.

Editorial interaction with the plugin is mediated through a constrained, OJS-consistent user interface. Editors and journal managers can initiate validation for individual or multiple submissions from a centralized listing view that displays current metadata status indicators. Validation results are presented in a structured report format that enumerates detected issues, their severity, and the affected metadata fields. Where applicable, the system generates suggested corrections derived from the article content, allowing users to apply these changes selectively. This design preserves editorial control while significantly reducing the cognitive and administrative burden associated with manual cross-checking.

Upon successful validation, the plugin injects the corrected and enriched metadata back into the OJS submission object using the appropriate Data Access Objects (DAOs). This ensures

that metadata consistency is maintained across internal workflows and downstream services, including Crossref DOI deposits, OAI-PMH harvesting, and repository export mechanisms such as DSpace. Additionally, the plugin maintains detailed validation logs—recording execution timestamps, user actions, and issue resolution metrics—which can be exported for auditing and reporting purposes.

Collectively, the Automated Metadata Verifier and Injector Plugin operationalizes a data-centric, automation-driven approach to metadata governance within OJS. By combining configurable validation rules, AI-assisted document analysis, editorial oversight, and seamless metadata reintegration, the plugin enhances the reliability, interoperability, and global visibility of institutional research outputs while remaining sensitive to the infrastructural constraints commonly faced by journals in the Global South.

3.5 Automated DOI Depositor Plugin

The Automated DOI Depositor Plugin was developed to address persistent inefficiencies and error-prone practices in DOI registration workflows within Open Journal Systems (OJS). While OJS provides a native Crossref XML Export Plugin, that tool is fundamentally dependent on metadata manually entered into the system by editors or journal managers. In contexts where editorial capacity is limited and metadata entry is performed under time pressure, this dependency introduces a high risk of inconsistency, formatting errors, and incomplete records, often resulting in failed DOI deposits or inaccurate registrations. The Automated DOI Depositor Plugin was therefore designed as a corrective intervention that shifts the authoritative source of metadata from manually entered OJS fields to the article galleys themselves.

Technically, the plugin is implemented as a custom ImportExportPlugin, located under `plugins/importexport/doiDepositor/`, and conforms to OJS 3.3.0.15 architectural conventions. The core plugin class (`DoiDepositorPlugin.inc.php`) is responsible for plugin registration, role-based access control, and integration into the Import/Export interface. Deposit logic and API communication are encapsulated within a dedicated handler (`DoiDepositorHandler.inc.php`), while presentation and user feedback are managed through Smarty templates. Supporting components include locale files for internationalization, a configuration module for Crossref credentials and operational settings, and structured logging utilities to capture deposit outcomes and error states.

The defining feature of this plugin is its metadata acquisition strategy. Instead of relying on the OJS submission metadata as the primary input for DOI registration, the plugin performs automated extraction and normalization of metadata directly from the finalized article galleys (typically PDF files). Using a combination of structured parsing, pattern recognition, and optional AI-assisted analysis, the system identifies canonical metadata elements such as the article title, author list, affiliations, abstract, publication date, journal title, volume, issue, and page ranges. These extracted elements are then reconciled against Crossref schema requirements to construct a deposit-ready metadata package. This approach significantly reduces the propagation of human errors introduced during manual metadata entry and ensures that the DOI record reflects the authoritative, published version of the article.

In operational terms, the plugin executes a multi-stage workflow. First, it validates the availability and integrity of the article galley, ensuring that a definitive version of record exists. Second, it extracts and normalizes metadata from the galley, applying Crossref-specific constraints and formatting rules. Third, it performs pre-deposit validation to detect missing or ambiguous elements that could cause rejection by the Crossref API. Only after these checks does the plugin initiate an automated deposit request, transmitting the metadata directly to Crossref via its REST API. Deposit responses, including success confirmations or error messages, are captured and translated into human-readable feedback for editors.

From a user experience perspective, the plugin exposes a minimal and consistent interface aligned with OJS design constraints. Editors and journal managers can initiate DOI deposits individually or in batches, view real-time deposit status, and access detailed logs explaining the outcome of each transaction. Importantly, the plugin decouples DOI registration from the timing of manual metadata cleanup, allowing journals to register DOIs with a higher degree of confidence in the accuracy of the deposited records.

In contrast to the standard Crossref XML Export Plugin, which functions primarily as a metadata serialization tool, the Automated DOI Depositor Plugin operates as an end-to-end registration mechanism grounded in the published article itself. This conceptual shift redefines the DOI deposit process from a metadata-dependent administrative task into a publication-centric, automation-driven workflow. By aligning DOI registration with the actual content of article galleys, the plugin enhances the reliability, persistence, and scholarly credibility of UNZA-hosted journals, while simultaneously reducing editorial workload and increasing compliance with international indexing standards.

4.CONCLUSION

This study demonstrates how targeted extensions to Open Journal Systems can meaningfully strengthen institutional scholarly publishing workflows when designed with local realities and global standards in mind. The Automated Metadata Verifier and Injector, the Automated DOI Depositor, and the DSpace Package Exporter collectively address critical points of failure in the publication lifecycle—namely metadata inconsistency, unreliable identifier registration, and weak integration between journals and institutional repositories.

Rather than treating metadata management, DOI registration, and repository deposition as isolated tasks, the plugin suite approaches them as interconnected components of a single scholarly communication pipeline. The metadata verifier establishes a foundation of accuracy by aligning OJS records with the authoritative content of article galleys, reducing downstream errors in indexing and discovery. Building on this foundation, the DOI Depositor improves the reliability of Crossref registrations by sourcing metadata directly from the published article itself, thereby minimizing the effects of human error that commonly arise during manual data entry. The DSpace Package Exporter then ensures that the verified and published record is preserved and disseminated through the institutional repository in a structured, standards-compliant manner.

Beyond their technical functionality, these plugins contribute to a broader institutional objective: improving the visibility, credibility, and sustainability of university-hosted journals. For institutions such as the University of Zambia, where journals are often underrepresented in global indexing services, these tools offer a practical pathway toward improved compliance with international metadata standards and open access requirements. By reducing editorial workload and embedding quality control directly into publishing workflows, the plugins also support more consistent editorial practices across journals with varying levels of technical capacity.

Importantly, the modular design of the plugins allows for future extension, including support for additional indexing platforms, multilingual metadata validation, and deeper integration with researcher identifiers such as ORCID. As such, the work presented here should not be viewed as a closed technical solution, but rather as a flexible framework that can evolve alongside institutional needs and emerging scholarly infrastructure.

In conclusion, the plugin suite illustrates how modest but carefully designed interventions within OJS can produce significant gains in metadata quality, workflow efficiency, and research visibility. By aligning technical implementation with institutional priorities and global scholarly standards, the project contributes to more resilient and equitable scholarly publishing practices.

Reference

- [1] Aremu, B. 2024. *How to Increase Your Book Sales with Metadata: Reach Readers, Increase Discovery, Increase Sales*. AB Publisher LLC.
- [2] Ashour, A., Mirghani, M., Mohamed, M., Osman, W., Alqarni, M.H., Foudah, A.I., Zaki, A.A. and Samra, R.M. 2025. Natural product chemistry as an important topic in pharmacy education: Situational analysis and recommendations. *Currents in pharmacy teaching & learning*. 17, 9 (Sep. 2025), 102410.
- [3] Butt, B.H., Rafi, M. and Sabih, M. 2021. A systematic metadata harvesting workflow for analysing scientific networks. *PeerJ. Computer science*. 7, (Mar. 2021), e421.
- [4] Choudhury, T., Awasthi, A. and Jung-Sup, U.M. 2025. *Open Access as a Knowledge Platform for Implementing the United Nations Sustainable Development Goals: Open Access and Sustainable Development Goals*. Springer.
- [5] Cirillo, A. 2017. *R Data Mining: Implement data mining techniques through practical use cases and real world datasets*. Packt Publishing Ltd.
- [6] Getty Information Institute 1998. *Introduction to Metadata: Pathways to Digital Information*. Oxford University Press, USA.
- [7] Matheka, D.M., Nderitu, J., Mutonga, D., Otiti, M.I., Siegel, K. and Demaio, A.R. 2014. Open access: academic publishing and its implications for knowledge equity in Kenya. *Globalization and health*. 10, (Apr. 2014), 26.
- [8] Miller, S.J. 2022. *Metadata for Digital Collections*. American Library Association.
- [9] Stansfield, C., Dehdarirad, H., Thomas, J., Mathew, S. and O'Mara-Eves, A. 2025. Analyzing the Utility of Openalex to Identify Studies for Systematic Reviews: Methods and a Case Study. *Cochrane evidence synthesis and methods*. 3, 4 (Jul. 2025), e70038.
- [10] Wijnbergen, D., Kaliyaperumal, R., Burger, K., Bonino da Silva Santos, L.O., Mons, B., Roos, M. and Mina, E. 2025. The FAIR data point populator: collaborative FAIRification and population of FAIR data points. *BMC medical informatics and decision making*. 25, Suppl 1 (Jun. 2025), 211.
- [11] Jung-ran Park. 2009. Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & Classification Quarterly* 47, 3–4 (2009), 213–228. <https://doi.org/10.1080/01639370902737240>
- [12] Marcia Lei Zeng and Jian Qin. 2016. *Metadata*. 2nd ed. ALA Neal-Schuman, Chicago, IL, USA.
- [13] Jane Greenberg. 2010. Metadata quality: From evaluation to augmentation. *Cataloging & Classification Quarterly* 48, 5 (2010), 391–406. <https://doi.org/10.1080/01639374.2010.489416>
- [14] Jan Brase. 2009. DataCite – A global registration agency for research data. In *Proceedings of the 4th International Conference on Cooperation and Promotion of Information Resources in Science and Technology*. IEEE, 257–261. <https://doi.org/10.1109/COINFO.2009.66>

- [15] Martin Fenner, Mercè Crosas, Jeffrey S. Grethe, et al. 2018. A data citation roadmap for scholarly data repositories. *Scientific Data* 5, 1 (2018), Article 180259. <https://doi.org/10.1038/sdata.2018.259>
- [16] Jingfeng Xia and Li Sun. 2007. Assessment of self-archiving in institutional repositories: Depositorship and metadata quality. *Serials Review* 33, 1 (2007), 31–41. <https://doi.org/10.1016/j.serrev.2006.12.003>
- [17] Gonzalo Villarreal, Juan Pablo Fernández, and Ariel Ríos. 2015. Interoperability between Open Journal Systems and DSpace: A case study. *Information Services & Use* 35, 1–2 (2015), 25–34. <https://doi.org/10.3233/ISU-150742>
- [18] James Macgregor. 2023. Automating DOI workflows in Open Journal Systems using Crossref APIs. *Journal of Scholarly Publishing* 54, 3 (2023), 145–160.
- [19] Highwire Press. 2024. Modular publishing platforms and workflow extensibility. Highwire Press White Paper.
- [20] Syntax Stories. 2024. Why plugin-based architectures scale better in scholarly publishing systems. Syntax Stories Technical Blog.