When metadata is not enough: repository interoperability in a national library open-source discovery platform

Cuando los metadatos no son suficientes: interoperabilidad de repositories en una plataforma de descubrimiento de una biblioteca nacional

Dwi Fajar SAPUTRA (1,2), Taufik ASMIYANTO (2), Nina MAYESTI (2)

(1) Departament of Information Science, Universitas Pembangunan Nasional Veteran Jakarta,
dwifajar@upnvj.ac.id, https://orcid.org/0000-0001-6831-037X

(2) Doctoral Program in Information Studies, Universitas Indonesia
tasmiy@ui.ac.id, https://orcid.org/0000-0002-3477-2071
nina.mayesti@ui.ac.id, https://orcid.org/0000-0002-5508-5377

Resumen

Los servicios de descubrimiento a escala web (WSDS) dependen de la interoperabilidad de los repositorios institucionales para garantizar un acceso fiable al contenido académico. Sin embargo, esta interoperabilidad suele evaluarse únicamente a partir del registro de metadatos, sin considerar su sostenibilidad operativa. Este estudio analiza la preparación para la interoperabilidad de los repositorios en el ecosistema Indonesia OneSearch (IOS) mediante un enfoque multidimensional que integra la integridad de los metadatos, la sostenibilidad de la recolección y la capacidad de los metadatos. A partir de datos de registro y recolección de IOS, se adopta un diseño cuantitativo y exploratorio, complementado con técnicas de análisis visual como gráficos de densidad, diagramas de dispersión con regresión, diagramas de caja y mapas de calor. Los resultados muestran una relación débil y no lineal entre la integridad de los metadatos y la sostenibilidad de la recolección, lo que indica que unos metadatos completos no garantizan una interoperabilidad sostenida. La mayoría de los repositorios se concentran en niveles bajos de actividad de recolección, mientras que los repositorios de revistas científicas presentan mayor capacidad de metadatos y un mejor nivel de integración que los repositorios de datos y de tesis y disertaciones electrónicas (ETD). El estudio concluye que la interoperabilidad de los repositorios debe entenderse como un proceso dinámico y gradual, más que como un estado binario. La evaluación conjunta de la calidad de los metadatos y la sostenibilidad de la recolección resulta esencial para fortalecer la gobernanza y el desarrollo sostenible de los servicios nacionales de descubrimiento.

Palabras clave: Sostenibilidad de metadatos. Métricas de interoperabilidad. Servicios de descubrimiento web. Indonesia One Search. Indonesia.

Abstract

Web scale discovery services (WSDS) depend on the interoperability of institutional repositories to provide reliable access to scholarly content. In practice, repository interoperability is often inferred solely from metadata registration, while operational sustainability receives limited attention. This study examines repository interoperability readiness within the Indonesia One Search (IOS) ecosystem using a multidimensional framework that integrates metadata completeness, harvesting sustainability, and metadata capacity. Drawing on registry and harvesting data from IOS, the study adopts a quantitative and exploratory design supported by visual analytical techniques, including density plots, scatter plots with regression analysis, boxplots, and heatmaps. The findings reveal a weak and non-linear relationship between metadata completeness and harvesting sustainability, indicating that well-documented metadata does not necessarily result in sustained interoperability. Most repositories are concentrated in low harvesting sustainability categories, reflecting widespread inactive or discontinuous integration. Journal site demonstrate higher metadata capacity and are more likely to achieve well-integrated status, whereas repositories managing datasets and electronic theses and dissertations (ETDs) exhibit lower capacity and greater vulnerability to operational stagnation. The study concludes that repository interoperability should be conceptualised as a dynamic and graduated condition rather than a binary state. Assessing metadata completeness in isolation is insufficient, sustained harvesting and metadata capacity must be evaluated jointly. This integrative assessment provides empirical evidence to support improved governance and long-term sustainability of national web-scale discovery services.

Keywords: Metadata sustainability. Interoperability metrics. Web Scale Discovery Services. Indonesia One Search. Indonesia.

1. Introduction

The development of web-scale discovery services (WSDS) has been a key element in the integration of scientific information sources across repositories at national and global scales. Through interoperability protocols such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), WSDS enables the aggregation of metadata from various heterogeneous systems to support unified indexing and information retrieval. In this context, metadata is no longer understood solely as a descriptive tool, but as an operational infrastructure that determines the success of integration, the stability of the index, and the sustainability of discovery services (Baglioni et al., 2025; Knoth et al., 2023).

However, according to Asok et al., the evaluation of repository system quality generally still focuses on metadata completeness, particularly in relation to the principles of open science and increasing content visibility (2024). Some studies even show a positive relationship between metadata completeness and user engagement in specific contexts, such as institutional repositories (Rasuli et al., 2025).

However, this approach tends to be static and does not adequately consider operational dynamics in large-scale aggregation environments. Experience with international aggregation services shows that descriptively complete metadata cannot always be sustainably integrated due to platform differences, variations in institutional technical capacity, and inconsistencies in maintaining OAI-PMH endpoints (Knoth et al., 2023).

This condition indicates a gap between the registration-based assessment of metadata quality and the reality of dynamic, long-term interoperability. Interoperability does not stop at the repository registration stage; rather, it depends on the sustainability of harvesting activities and the continuous maintenance of technical readiness. This perspective is becoming increasingly relevant in the context of national aggregation, where the heterogeneity of institutions and repository management systems poses its own structural challenges (Baglioni et al., 2025; Knoth et al., 2025).

Indonesia One Search (IOS), as a national WSDS service, faces similar challenges in integrating thousands of systems with diverse characteristics and levels of readiness. Although the IOS registration system has provided an initial foundation for metadata integration, there are strong indications of stagnation in harvesting activities and disparities in technical readiness among repositories. To date, empirical research comprehensively evaluating the interoperability conditions of systems in WSDS by linking metadata completeness, operational sustainability, and technical readiness is still limited. If we relate this to IOS as the WSDS of the national library, data from reporting.onesearch.id (2025) shows a decrease in data entries from 2022-2023, with the number of entries dropping by 55.4%. Furthermore, no data is available for the years 2024-2025.

Based on this gap, this study aims to evaluate the interoperability of repositories within the Indonesian One Search ecosystem using national aggregated data. Specifically, this research analyzes the completeness level of registration metadata, maps the sustainability of harvesting activities as an indicator of long-term interoperability, and classifies repository readiness based on descriptive and operational dimensions. Conceptually, the research is built upon the framework that interoperability is the result of a multidimensional interaction between metadata quality, technical maintenance sustainability, and institutional capacity, so complete metadata does not automatically guarantee stable integration without consistent operational support.

To test the framework, this study employs a quantitative-exploratory approach by analyzing the Indonesian One Search registration and harvesting dataset. This approach allows for an empirical evaluation of the relationship between metadata quality and long-term interoperability, while also providing an evidence-based foundation for developing more adaptive metadata policies and national WSDS management.

2. Literature Review

2.1. Quality and Completeness of Metadata in Digital Repositories

The quality of metadata is a fundamental factor in the success of information retrieval systems and web-scale discovery services. High-quality metadata allows information sources to be effectively indexed, discovered, and utilized within a distributed digital repository ecosystem. One of the most widely studied dimensions of metadata quality is metadata completeness, which refers to the extent to which the required metadata elements are available and well-filled. Within the framework of classical metadata quality, completeness is seen as a minimum prerequisite for a record to function optimally within an information system (Bruce & Hillmann, 2004).

In recent decades, metadata completeness has been increasingly used as a quantitative indicator to assess system quality, primarily because it is relatively easy to measure and can be applied

across metadata schemas. Asok et al. (2024) emphasize that complete metadata is an essential foundation for open science practices, particularly in supporting the visibility, interoperability, and reuse of research data. Rasuli et al. also showed that metadata completeness contributes to increased visibility and user engagement, for example, in institutional repositories and library collection automation (2025).

Various approaches have been developed to systematically measure the completeness of metadata. Ochoa and Duval (2009) introduced the automated evaluation of metadata quality using element-presence-based metrics, while Király and Büchler (2019 developed a weighted approach that considers the relative importance of each metadata element.

Another approach considers the hierarchical structure of metadata, where an element is considered complete only if all its sub-elements are filled (Margaritopoulos et al., 2012). Although these approaches provide a strong quantitative basis for assessing the descriptive quality of metadata, most still view metadata as a static entity evaluated at a single point in time.

Beside completeness, other quality dimensions such as accuracy and consistency are also widely discussed in the literature. However, several studies indicate that incompleteness, format inconsistency, and filling errors remain common issues in institutional repositories, even in stable systems (Çolak & Eroğlu, 2025). This finding indicates that the quality of metadata depends not only on the schema or guidelines used, but also on the practices of ongoing metadata management and maintenance.

2.2. Repository Interoperability and the Sustainability of Metadata Harvesting

In a large-scale aggregation environment, the quality of metadata cannot be separated from the issues of interoperability and operational sustainability. Web-scale discovery services rely on the ability of repository systems to provide metadata that can be consistently harvested through interoperability protocols such as OAI-PMH. In this context, interoperability is not only determined by the completeness of metadata elements, but also by the stability of the endpoint, format compliance, and the continuity of system maintenance.

Global aggregators like CORE indicate that the heterogeneity of repository system platforms, variations in institutional technical capacity, and differences in metadata management practices are major challenges in cross-system metadata integration (Knoth et al., 2023). Manocci's study highlights that many repositories listed in national or international registries are not always sustainably harvestable due to URL changes, server failures, or a lack of technical maintenance (2022). This condition indicates that interoperability is a dynamic process susceptible to degradation over time.

Sustainable harvesting is an important indicator in assessing the interoperability readiness of repositories. Outdated metadata or repositories that are no longer responsive to harvesting requests can cause information to become obsolete, out of sync, or even disappear from the aggregator's index.

This phenomenon is often associated with the concept of metadata decay, which is the decline in the quality and relevance of metadata due to a lack of long-term maintenance. Although the dimension of timeliness has long been recognized as part of metadata quality, this aspect is relatively rarely operationalized systematically in repository evaluations (Kumar & Harinarayana, 2024).

Some recent studies are beginning to emphasize the importance of combining metadata quality evaluation with operational sustainability indicators. Baglioni et al. (2025) highlight the need for a registry-based interoperability approach that not only records the existence of repositories but also validates the technical readiness and sustainability of metadata integration.

A similar approach is also seen in studies on repository registries and national aggregation networks, which emphasize the importance of technical coordination and sustainable maintenance to ensure the stability of discovery services (Knoth et al., 2025).

3. Methodology

This study employs a quantitative-exploratory approach to evaluate the interoperability of repositories within the national web-scale discovery services (WSDS) ecosystem. The quantitative approach was chosen because it allows for objective measurement of metadata characteristics and interoperability performance on a large scale, while the exploratory nature is used to identify patterns, disparities, and trends that have not been extensively mapped in national-scale repository interoperability studies (Creswell & Creswell, 2018; Kumar & Harinarayana, 2024).

The research object includes repositories listed in Indonesia One Search as a national discovery service. Research data is sourced from registration metadata and harvesting records managed by the aggregator system. The use of aggregated

data is considered methodologically relevant because interoperability within the WSDS cannot be understood at the individual repository level, but rather as a systemic phenomenon emerging from the interaction between repositories, interoperability protocols, and aggregator infrastructure (Knoth et al., 2023; Baglioni et al., 2025).

Conceptually, the interoperability of repository systems in this study is treated as a multidimensional construct encompassing descriptive, technical, and operational aspects. This multidimensional approach aligns with the literature on metadata quality and interoperability evaluation, which emphasizes that metadata completeness alone is insufficient to represent long-term integration readiness, especially in large-scale aggregation environments (Ochoa & Duval, 2009; Margaritopoulos et al., 2012; Kumar & Harinarayana, 2024).

Data analysis was conducted by combining metadata completeness measurement, statistical distribution analysis, temporal harvesting classification, and comparative analysis across repository types. The use of presence-based scoring for metadata completeness metrics is a common technique in metadata evaluation studies to generate quantitative indicators that are replicable and comparable across systems (Ochoa & Duval, 2009; Király & Büchler, 2019).

Distribution analysis and comparison were used to identify structural variations in interoperability readiness across repository groups, as recommended in the institutional repository evaluation study (Çolak & Eroğlu, 2025).

The temporal dimension of interoperability is analyzed through the use of data harvesting to assess the sustainability of metadata integration over time. This temporal approach is used to capture the phenomena of interoperability degradation and metadata decay, which have been identified as important issues in global repository aggregation and service discovery studies (Mannocci et al., 2022; Knoth et al., 2023).

Thus, interoperability is not treated as a static condition, but rather as a dynamic process that depends on the sustainability of technical and operational maintenance.

Through this methodological approach, the research does not aim to establish a causal relationship, but rather to present an empirical mapping based on national aggregate data regarding the interoperability conditions of repositories.

This approach aligns with the practice of large-scale information systems evaluation, which places descriptive, classificatory, and trend analysis as the basis for policy decision-making and system development (Creswell & Creswell, 2018; Baglioni et al., 2025).

4. Discussion

4.1. Mapping of IOS Registration Field Metadata Categories

The Indonesian One Search (IOS) registration dataset contains metadata fields that not only function as a description of the repository's identity but also track the technical readiness and operational status of its integration into the national retrieval system. Due to the heterogeneous nature of the field characters, category mapping is necessary to ensure that each attribute can be read consistently within the interoperability evaluation framework.

In this study, all 29 fields are understood through four functional categories:

1. administrative registration metadata (Table I)
2. interoperability technical metadata (Table II)
3. operational harvesting metadata (Table III)
4. system administration metadata (Table IV)

The four components form a sequence that represents the repository integration cycle: from identity at registration, technical access for metadata exchange, activity trail of harvesting and indexing, to workflow control and change auditing (Figure).

Field	*Definition*	*Function*
id	Unique identifier of an entity in the IOS system	Record identification
title	Name of the repository, journal, or collection as registered in IOS	Entity identity
publisher	Name of the managing or responsible institution	Service accountability
description	Description of the repository	Additional contextual information
library_id	Institutional identifier within the IOS network	Institutional relationship and linkage

Table I. IOS Administrative Registration Metadata Fields

Field	*Definition*	*Function*
oai_url	OAI-PMH endpoint of the repository or journal	Harvesting access
metadataPrefix	Metadata format used (e.g., dc, oai_dc, mods, etc.)	Interoperability standard
setSpec	Set identifier within the OAI-PMH structure	Harvesting filter and segmentation
update_frequency	Update frequency selected by the repository manager	Synchronization scheduling
content_type	Type of system (journal, institutional repository, ETD, SLiMS UCS, others)	Platform identification
repository_software_id	Repository or publishing system used (OJS, EPrints, SLiMS, DSpace, etc.)	Platform interoperability
home_url	Main homepage of the repository or journal	Entity existence validation
cover_url	Cover image or visual representation of the repository (optional)	Visual identity and aesthetics

Table II. IOS Technical Interoperability Metadata Fields

Field	*Definition*	*Function*
harvest_date	Date of the most recent successful harvesting process	Sustainability indicator
harvest_status	Status of the harvesting result (e.g., success, fail, pending)	OAI communication quality
harvest_user	System or reviewer responsible for the harvesting process	Audit trail
harvest_url	Source URL used for harvesting	Endpoint validation
dedup_date	Date when metadata deduplication was performed	Post-harvest processing
dedup_status	Status of the deduplication process	Metadata cleaning indicator
dedup_user	User or process responsible for deduplication	Audit trail
solrupdate_date	Date when metadata were successfully indexed into Solr	Indexing success
solrupdate_status	Status of the indexing process	Discovery validation
solrupdate_user	System or process responsible for Solr indexing	Audit trail

Table III. IOS Operational Harvesting Metadata Fields

Field	*Definition*	*Function*
date_created	Date when the entity was created in the IOS system	Registration initiation
date_modified	Date when the metadata were last modified	Maintenance history
date_queued	Date when the entity entered the review queue	Workflow management
status	Registration status (e.g., approved, pending, rejected)	Reviewer validation
reason	Notes provided by the reviewer explaining the decision	Standards compliance assessment
notes	Additional remarks or technical notes	Supplementary technical comments
queued	Boolean flag indicating whether the entity is in the review queue	Queue position indicator
created_by	Administrator or user who created the entity	Audit
last_modified_by	User or system responsible for the latest modification	Audit
queued_by	User or system that placed the entity into the review queue	Audit

Table IV. IOS System Administration Metadata Fields

4.2. Data Preprocessing

Data preprocessing in this study is a fundamental step to ensure that the Indonesian One Search (IOS) dataset is in a suitable condition for analysis. The raw dataset obtained from the IOS registration and harvesting system exhibits variations in format, the presence of null values, and heterogeneity in metadata structure, which could potentially affect the accuracy of quantitative analysis. Preprocessing focuses on four main aspects: data cleaning, format consistency, semantic completeness, and mapping of analytical indicators. This series of steps ensures that each metadata field can be interpreted consistently within the framework of repository interoperability evaluation. With systematically processed datasets, further analysis can more accurately reflect the institution's metadata conditions within the Indonesian One Search ecosystem, while also supporting the validity and comparability of research results.

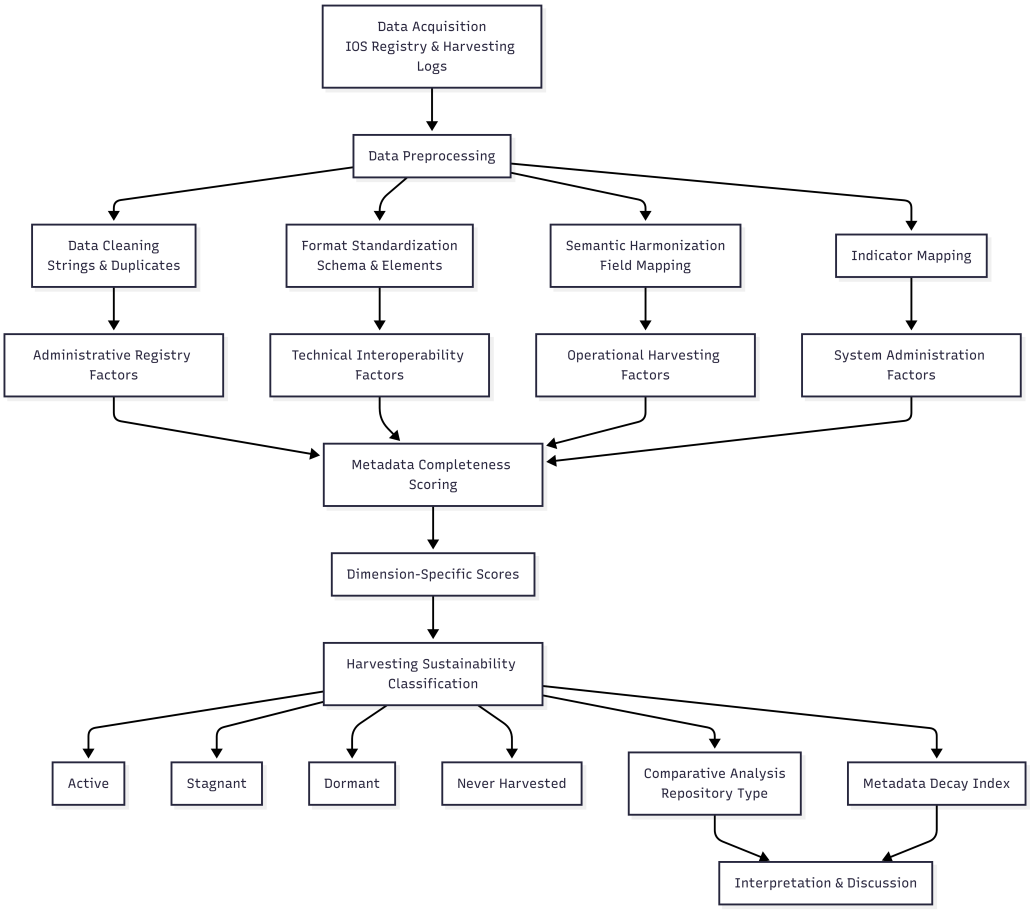


Figure 1. Research workflow for evaluating repository interoperability readiness in the Indonesia One Search (IOS) ecosystem

4.3. Analysis of Metadata Completeness: Distribution, Dimensions, and Systemic Patterns

Figure 2 presents the distribution of metadata health scores defined here as the aggregated completeness score across all metadata dimensions for each repository which indicates a concentration of repositories in the mid-to-low score range, with a small percentage of repositories at a very high level of completeness. This uneven distribution pattern reflects what is referred to in the literature as metadata quality skewness, which is the condition where most repositories only meet minimum metadata requirements, while more mature metadata practices are only found in a limited group of institutions (Park & Tosaka, 2015).

Some studies confirm that this type of distribution pattern is common in open and decentralized institutional repository ecosystems. Stvilia et al. (2017) explain that without continuous audit and feedback mechanisms, metadata tends to evolve unevenly, depending on the local capacity of the managing institution. In the national WSDS context, this means that the existence of repositories with high metadata scores does not automatically improve the overall quality of the system, as the majority of repositories remain at a minimum level of contribution.

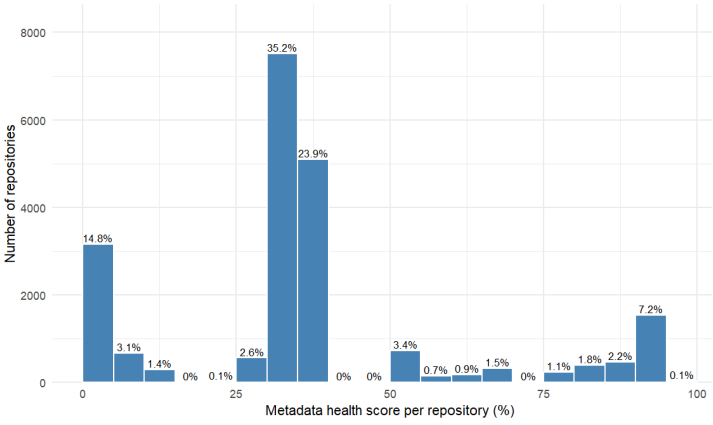


Figure 2. Distribution of metadata health scores across IOS repositories (histogram)

This distribution also strengthens the argument that interoperability assessment cannot be based solely on average values, but must consider the spread and concentration of scores. This kind of distributional approach has been recommended in the evaluation of large-scale information systems to avoid conclusions biased toward high-capacity institutions (Batini, Cappiello & Francalanci, 2016). Thus, the histogram of iOS metadata scores not only describes the technical condition of the metadata but also reveals structural disparities in the interoperability readiness of repositories.

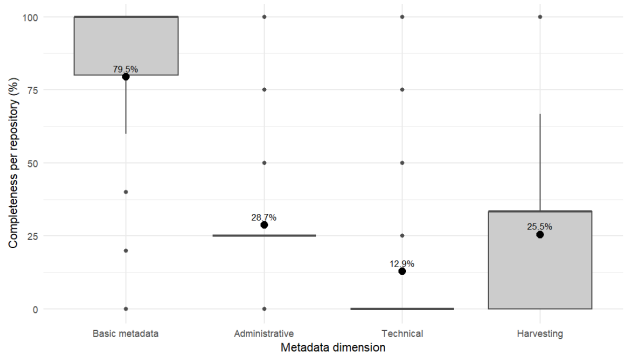


Figure 3. Metadata completeness by dimension across IOS repositories (boxplot)

This phenomenon aligns with findings in metadata management studies that distinguish between descriptive sufficiency and functional interoperability. Gilliland (2016) emphasizes that descriptive metadata is often prioritized because it is directly related to the visibility and legitimacy of collections, while technical and operational metadata is considered a less institutionally focused supporting layer. As a result, although the repository appears "complete" descriptively, its capacity to actively participate in the aggregation system remains limited.

Additionally, the low completeness in the technical and harvesting dimensions can also be interpreted as an indicator of weak integration sustainability. According to Higgins (2018), long-term interoperability in digital systems is highly dependent on process metadata, including the recording of harvesting and update activities. When this type of metadata is not managed consistently, aggregation systems risk functional degradation even if descriptive metadata remains available.

The differences between dimensions shown by the boxplot also support the multidimensional evaluation approach in this study. As stated by Zeng and Qin (2016), the quality of metadata cannot be reduced to a single indicator, but must be read as a combination of interconnected dimensions. This visualization strengthens the argument that the completeness of administrative metadata is not sufficient to represent the interoperability readiness of repositories within the national WSDS.

Figures 2 and 3 together show that the quality of repository metadata within the IOS ecosystem is uneven and highly dependent on the dimensions being evaluated. While basic metadata is relatively stable and complete, the technical and operational metadata that underpins long-term interoperability still shows structural weaknesses. This finding reinforces the view that WSDS interoperability cannot be evaluated solely through the completeness of descriptive metadata, but rather requires a multidimensional approach that considers score distribution and the operational dynamics of the repository.

4.4. Dynamics of Repository Harvesting Activity: Patterns, Challenges, and Sustainability

Figure 4 shows that most repositories in the Indonesian OneSearch ecosystem fall into the "never harvested" category, while the proportion of repositories actively harvested in the last year is very small. This pattern indicates that the integration of repositories into the national WSDS is often administrative, without sustained operational connectivity. In other words, the existence of a repository within a registration system does not automatically transform into an active metadata flow within the discovery index.

The dominance of the "never harvested" and "dormant" categories reflects a phenomenon known in repository aggregation literature as structural harvesting failure, which is a systemic and recurring integration failure, not just a temporary technical glitch. Mannocci, Baglioni, and Manghi (2022) show that at a large aggregation scale, many repositories become inaccessible due to endpoint changes, lack of system maintenance, or the absence of periodic monitoring mechanisms. In the context of iOS, this pattern indicates a gap between the registration process and the sustainable integration of metadata.

From an interoperability perspective, these findings confirm that harvesting sustainability is a more sensitive indicator than the completeness of registration metadata in assessing repository readiness. Repositories can have complete registration metadata, but still not contribute to WSDS if harvesting activity is not sustained. This condition strengthens the argument that interoperability needs to be understood as a dynamic

process that depends on technical and operational continuity, rather than as a static status at a single point in time (Higgins, 2018).

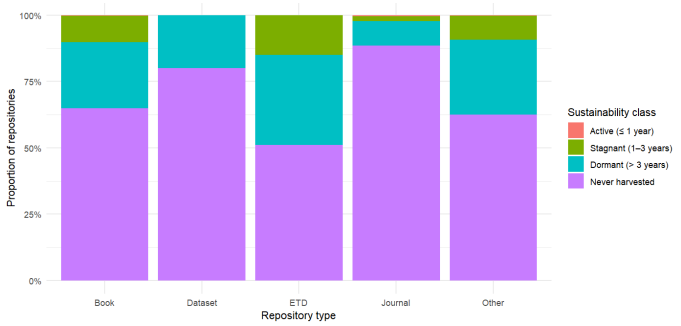


Figure 4. Distribution of IOS repositories across harvesting sustainability classes by repository type

Figure 4 (disaggregated by repository type) reveals a noticeable difference in the sustainability patterns of harvesting. The journal repository shows a relatively better proportion of harvesting activity compared to ETD repositories, datasets, and other types. This pattern reflects that repositories with regular publication cycles and a strong visibility orientation tend to have greater institutional incentives to maintain their technical interoperability. Conversely, ETD and dataset repositories are dominated by the never harvested and dormant categories. This indicates that although these repositories play a crucial role in the knowledge and open science ecosystem, managing their interoperability is often not a long-term priority. Repository management literature indicates that repositories with irregular content flows and fragmented management responsibilities are more prone to experiencing a decline in metadata integration quality (Park & Tosaka, 2015).

The low sustainability of harvesting on dataset repositories also underscores the complexity of research data interoperability. Batini, Cappiello, and Francalanci (2016) emphasize that dataset interoperability demands higher metadata investment and technical maintenance compared to text-based repositories. Without adequate policy and resource support, this type of repository tends to fail to maintain metadata integration within the national aggregation system.

The variation in sustainability across repository types also confirms that interoperability is contextual and influenced by domain characteristics, institutional practices, and technical governance. Therefore, the national WSDS evaluation needs to move beyond a uniform approach and consider monitoring and intervention mechanisms tailored to the type of repository. This finding reinforces the need for sustainable harvesting indicators as an integral part of repository interoperability evaluation, not merely a supplement to registration metadata analysis.

4.5. Integrating Metadata Completeness, Harvesting Sustainability, and Interoperability Readiness

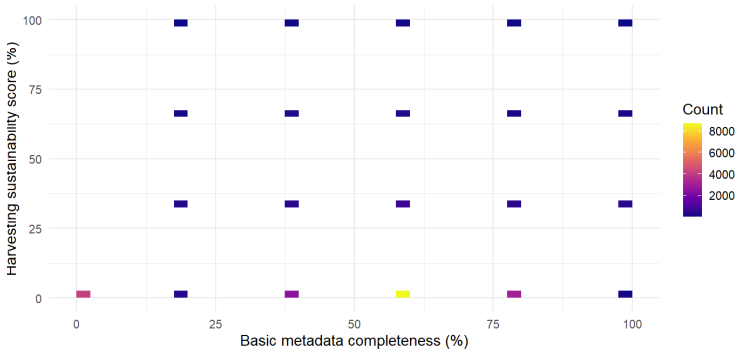


Figure 5. Density plot of metadata completeness versus harvesting sustainability across IOS repositories

The low repository density in areas with a combination of high metadata completeness and high harvesting sustainability indicates that ideal interoperability conditions are relatively rare. This finding supports the view that descriptive metadata is often prioritized for initial legitimacy and visibility purposes, while the technical and operational metadata that underpins sustainable harvesting receives less institutional attention (Gilliland, 2016; Zeng & Qin, 2016). Thus, structurally complete metadata does not necessarily reflect a repository's readiness to actively participate in the web-scale discovery services ecosystem.

The distribution of repositories with varying levels of metadata completeness but still at a low harvesting level also indicates a structural decoupling between metadata quality and the operational performance of interoperability. Global repository aggregation literature indicates that the sustainability of harvesting is more influenced by technical and governance factors, such as the stability of the OAI-PMH endpoint, repository software maintenance, and human resource sustainability, than by the mere completeness of registration metadata (Knoth et al., 2023; Higgins, 2018).

From a system evaluation perspective, this density pattern strengthens the argument that repository interoperability is a dynamic process susceptible to degradation if not accompanied by continuous monitoring and maintenance. The condition where metadata remains available but

is no longer actively harvested reflects the phenomenon of metadata decay, which directly impacts the timeliness and reliability of national discovery indexes (Higgins, 2018; Mannocci et al., 2022). Therefore, this visualization confirms the need to integrate metadata completeness indicators with sustainable harvesting indicators in assessing repository interoperability readiness.

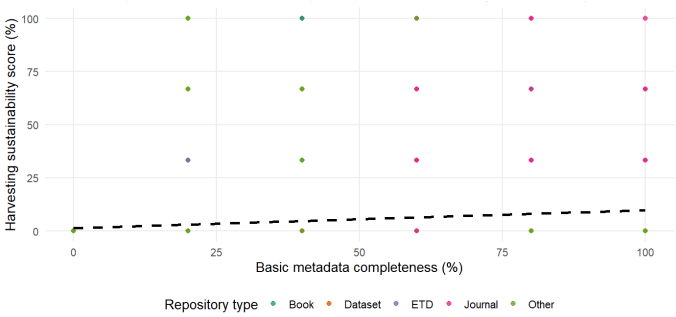


Figure 6. Scatterplot of metadata completeness versus harvesting sustainability by repository type

The wide distribution of points across various levels of metadata completeness, including repositories with high completeness but low harvesting values, suggests that metadata completeness is not the primary predictor of operational interoperability. This finding aligns with the literature emphasizing that the quality of descriptive metadata often serves as an initial prerequisite, but is not sufficient to guarantee long-term technical integration success in aggregation and discovery systems (Park & Tosaka, 2015; Zeng & Qin, 2016).

The variation in dot color based on repository type shows that the relationship between metadata completeness and harvesting sustainability is also contextual and domain-specific. Journal repositories tend to appear at a higher harvesting rate compared to other repository types, although they do not always have the highest metadata completeness. This pattern indicates that institutional factors such as publication rhythm, visibility demands, and management support have a greater influence on harvesting sustainability than metadata completeness alone. Studies on global repository aggregation show that repositories with regular content update cycles are better able to maintain technical interoperability, regardless of variations in the quality of their descriptive metadata (Knoth et al., 2023).

Conversely, the dataset and ETD repositories show a more inconsistent distribution, with many points at low harvesting levels despite varying metadata completeness. This condition reinforces the finding that the interoperability of this type of repository is highly dependent on the sustainability of technical maintenance and data management policies, not just on filling in registration metadata (Higgins, 2018). In the national WSDS context, this confirms that the interoperability improvement strategy cannot focus solely on enhancing metadata completeness, but needs to be directed toward strengthening operational harvesting mechanisms.

Conceptually, this visualization supports the argument that repository interoperability is the result of a multidimensional interaction between metadata, technology, and institutional practices. The weak relationship between metadata completeness and harvesting sustainability indicates a structural decoupling between descriptive readiness and operational performance. Therefore, a comprehensive evaluation of WSDS interoperability needs to integrate sustainable harvesting indicators and institutional context as key components of repository readiness assessment.

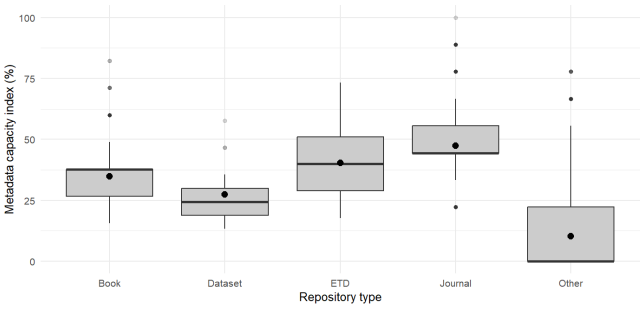


Figure 7. Metadata capacity by repository type (boxplot)

The high metadata capacity in journal repositories can be attributed to the demands of visibility, standardization, and the sustainability of scientific publications. Repository management literature indicates that journal repositories generally have stricter metadata policies and more stable integration with global aggregation and indexing systems, thus consistently supporting higher metadata capacity (Knoth et al., 2023; Park & Tosaka, 2015). This explains why journal repositories, despite not always having the highest metadata completeness, still demonstrate better interoperability readiness.

ETD repositories occupy a middle position with considerable variation, reflecting the heterogeneity of management practices at the institutional level. Some ETD repositories exhibit metadata capabilities approaching those of journal repositories, while others are at a lower level. This variation aligns with the finding that ETD management is highly dependent on institutional policies and the continuity of technical support, resulting

in uneven quality of the generated metadata (Higgins, 2018; Gilliland, 2016).

Conversely, the "other" dataset and category repositories show a lower median metadata capacity with a wide range of values. This condition indicates that although dataset repositories have significant potential in the context of open science, their metadata practices and interoperability are not yet fully established. Batini, Cappiello, and Francalanci (2016) emphasize that research data interoperability requires complex and sustained metadata investment; without adequate policy and resource support, dataset repositories are likely to experience metadata capacity limitations as reflected in this visualization.

The differences in metadata capacity between these repository types strengthen the argument that repository interoperability needs to be understood as a systemic capability, not merely the result of basic metadata completeness or simple harvesting activities. Metadata capacity serves as an integrative indicator, capturing the structural, technical, and operational readiness of the repository more comprehensively. Thus, this visualization bridges the previous completeness-sustainability relationship analysis toward the interoperability readiness classification of repositories in the next stage.

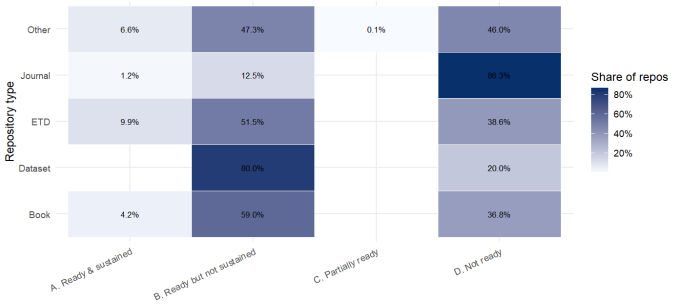


Figure 8. Interoperability readiness heatmap by repository type and MIR readiness class

Journal repositories hold the most prominent position in the Well-integrated class, indicating that the combination of metadata capacity, harvesting sustainability, and technical governance is more mature compared to other repository types. International literature confirms that journal repositories generally have more stable interoperability standards because they are tied to the global scientific publishing ecosystem, including indexers, aggregators, and open access initiatives (Björk, 2017; Knoth et al., 2023).

Thus, the interoperability readiness of journal repositories is not only influenced by metadata, but also by external pressures and institutional incentives.

Conversely, the dataset and ETD repositories are dominated by the Ready but not active and Partially ready classes. This reflects that although the repository already has metadata elements and basic infrastructure, the sustainability of integration and operational maintenance remains a major challenge. Studies on research data management show that dataset interoperability requires more complex coordination of metadata, policies, and technical resources compared to text-based repositories, which is why many dataset repositories fail to achieve full integration.

The "Not ready" category that still appears for some repository types, especially the "other" category, indicates structural gaps in interoperability readiness. The presence of repositories in this class indicates that the registration of repositories in the national WSDS has not been fully followed by continuous validation and mentoring mechanisms. The literature on discovery system evaluation emphasizes that without continuous interoperability governance, aggregation systems risk becoming mere lists of repositories rather than reliable retrieval infrastructures (Deodato, 2015; NISO, 2014).

Overall, this visualization confirms that repository interoperability readiness is tiered and highly influenced by domain context and institutional practices. This heatmap also reinforces the findings from the previous visualization that interoperability cannot be reduced to a single indicator, but is rather the result of integration between metadata capacity, harvesting sustainability, and system governance. Therefore, the interoperability readiness classification becomes an important instrument for the strategic and policy-oriented evaluation of the national WSDS.

5. Conclusion

This study demonstrates that repository interoperability within the Indonesia One Search ecosystem cannot be adequately assessed through metadata completeness alone, as harvesting sustainability and metadata capacity play decisive roles in determining long-term integration. The findings reveal structural disparities across repository types, where journal repositories tend to achieve higher interoperability readiness, while dataset and ETD repositories remain vulnerable to operational stagnation and metadata decay. By integrating metadata completeness and harvesting dynamics, this study contributes a multidimensional and evidence-based framework that advances the evaluation of web-scale discovery services beyond static metadata assessments.

The results offer practical insights for improving repository governance and monitoring mechanisms in national discovery infrastructures. Future research should incorporate longitudinal analysis and institutional policy variables to further refine interoperability readiness models and support sustainable repository ecosystems. Based on the findings, three governance recommendations are offered for IOS operators and analogous national WSDS.

First, automated OAI-PMH endpoint health monitoring should be implemented at quarterly intervals to detect harvesting failures before metadata decay becomes irreversible.

Second, differentiated support strategies are required by repository type: journal repositories need periodic standards audits, ETD repositories benefit from onboarding templates and institutional training, while dataset repositories require research data management policy assistance and deeper integration with national data infrastructure.

Third, registration workflows should include a minimum harvesting validation step—requiring at least one successful OAI-PMH harvest before approval—to ensure that the IOS index reflects operationally functional repositories rather than administrative registrations only.

CRediT author statement

Dwi Fajar Saputra: Conceptualization (lead); Methodology (lead); Investigation (lead); Formal analysis (lead); Data curation (lead); Writing – original draft (lead); Writing – review and editing (lead); Visualization (lead); Project administration (lead).

Taufik Asmiyanto: Supervision (equal); Methodology (supporting); Writing – review and editing (equal).

Nina Mayesti: Supervision (equal); Methodology (supporting); Writing – review and editing (equal).

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declares no conflict of interest.

Acknowledgments

The authors would like to thank the reviewers and editors for their constructive comments and suggestions, which helped improve the quality of this manuscript.

References

Asok, K.; Dandpat, S.S.; Gupta, D.K.; Shrivastava, P. (2024). Common metadata framework for research data repository: necessity to support open science. // Journal of Library Metadata. 24:1, 133-145.

Baca, M. (ed.) (2016). Introduction to Metadata. 3rd edn. Los Angeles: Getty Research Institute.

Baglioni, M.; Pavone, G.; Mannocci, A.; Manghi, P. (2025) Towards the interoperability of scholarly repository registries. // International Journal on Digital Libraries. 26, Article 2.

Batini, C.; Cappiello, C.; Francalanci, C. (2016). Data and Information Quality: Dimensions, Principles and Techniques. Cham: Springer.

Björk, B.-C. (2017). Open access to scientific articles: a review of benefits and challenges. // Information Services & Use. 37:4, 395-408.

Bruce, T.R.; Hillmann, D.I. (2004). The Continuum of Metadata Quality: Defining, Expressing, Exploiting. Chicago: ALA Editions.

Çolak, F.A.; Eroğlu, Ş. (2025). Evaluating metadata quality in institutional academic repositories of Turkish research universities. // Online Information Review. 49:7, 1335-1352.

Creswell, J.W.; Creswell, J.D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 5th ed. Thousand Oaks, CA: Sage.

Deodato, J. (2015). Evaluating web-scale discovery: A step-by-step guide. // Information Technology and Libraries. 34(2), 19-75. https://doi.org/10.6017/ital.v34i2.5745.

Gilliland, A.J. (2016). Setting the stage. // Baca, M. (ed.) Introduction to Metadata. 3rd edn. Los Angeles: Getty Research Institute, 1-19.

Higgins, S. (2018). Digital Curation: The Emergence of a New Discipline. London: Facet Publishing.

Király, P.; Büchler, M. (2019). Measuring completeness as a metadata quality metric in Europeana. // Proceedings of the IEEE International Conference on Big Data. IEEE, 2711-2720.

Knoth, P.; Herrmannova, D.; Cancellieri, M.; Anastasiou, L.; Pontika, N.; et al. (2023). CORE: a global aggregation service for open access papers. // Scientific Data. 10, 366.

Knoth, P.; Walk, P.; Cancellieri, M.; Upshall, M.; et al. (2025). USRN Discovery Pilot: increasing the discoverability of open access content through a national network. // Proceedings of the 20th International Conference on Open Repositories.

Kumar, V.C.; Harinarayana, N.S. (2024). Exploring dimensions of metadata quality assessment: a scoping review. // Journal of Librarianship and Information Science.

Mannocci, A.; Baglioni, M.; Manghi, P. (2022). "Knock Knock! Whos There?": a study on scholarly repositories availability. // Theory and Practice of Digital Libraries (TPDL 2022). Cham: Springer, 306-312.

Margaritopoulos, T.; Margaritopoulos, M.; Mavridis, I.; Manitsaris, A. (2012). Quantifying and measuring metadata completeness. // Journal of the American Society for Information Science and Technology. 63:4, 724-737.

NISO (2014) Open Discovery Initiative: promoting transparency in discovery. Baltimore: National Information Standards Organization.

Ochoa, X.; Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories. // International Journal on Digital Libraries. 10(2-3), 67-91.

Park, J.; Tosaka, Y. (2015). Metadata quality control in digital repositories and collections: criteria, semantics, and mechanisms, Cataloging & Classification Quarterly. 53(5-6), 595-613.

Rasuli, B.; Boock, M.; Schöpfel, J.; Van Wyk, B. (2025). The link between dissertation metadata completeness and user engagement in an institutional repository. // Scientometrics. 130:5, 2875-2899.

Reporting Indonesia OneSearch (2025). Global summary report: All institutions. Reporting Indonesia OneSearch. https://reporting.onesearch.id/index.php/global/all/summary

Stvilia, B.; Gasser, L.; Twidale, M.B.; Smith, L.C. (2017). A framework for information quality assessment. // Journal of the Association for Information Science and Technology. 68:1, 12-25.

Zeng, M.L.; Qin, J. (2016). Metadata. 2nd edn. London: Facet Publishing.

Enviado: 2026-01-09. Segunda versión: 2026-05-20.
Aceptado: 2023-06-18.