Unlocking potential: Harnessing the power of metadata for discoverability and accessibility

Julie Zhu*
Institute for Electrical and Electronics Engineers (IEEE), Piscataway, NJ, USA

Abstract. Metadata plays a crucial role in organizing, discovering, and accessing publications and products. This paper explores the complex metadata ecosystem and its impact on downstream functions. Focusing on author-provided metadata, it addresses the challenges faced by authors in terms of discoverability, and offers best practices for optimizing metadata. The paper also discusses publication-level metadata standards and related initiatives. It highlights the importance of KBART files for linking subscription products and explores the three types of linking in discovery services. Additionally, it emphasizes the role of metadata in search engine optimization and social media. The paper explores how metadata can enhance accessibility and emphasizes the need for collaboration across various teams and systems. Overall, it underscores the ongoing efforts to build robust metadata pipelines and improve the user experience in information discovery and interchange.

Keywords: Metadata ecosystem, author-provided metadata, discoverability, metadata challenges, best practices, publication metadata, KBART, linking, search engine optimization, social media metadata, accessibility, collaboration

1. Introduction: Metadata ecosystem

Metadata is an essential part of publications and products. We need metadata to display the key components of articles and publications, such as author-related information, publication information, article information, access information, funding information, and more. We need metadata for discoverability, linking, and access. We need metadata to track usage.

Metadata flows through multiple systems in various pipelines, from author systems, publishing systems, publishing platforms, indexing systems, link resolving systems, authentication systems, library systems, usage tracking systems, etc., and forms a huge metadata ecosystem. Metadata problems can occur at any time during any stage of the process, during metadata creation, enrichment, transfer, and configuration. Problems in the metadata upstream will impact the functions downstream. Fixing metadata downstream cannot possibly solve all of the possible problems.

I published a post in Scholarly Kitchen in 2019, titled "Building Pipes and Fixing Leaks: Demystifying and Decoding Scholarly Information Discovery & Interchange[1]." It includes a graph to show the complexity of the piping system. The chart reminds us that we content providers, discovery service providers,

---

and libraries - are all data plumbers. We need to continuously build data pipes and fix data leaks to ensure a better end-user experience.

This paper explores several key aspects of metadata and its critical role in various stages of the publishing process. While there are numerous types of metadata, this study focuses on three selected areas that significantly impact the overall metadata ecosystem. First, it delves into the importance of maximizing discoverability through best practices for authors' metadata. By optimizing article titles, author names, keywords, and abstracts, authors can enhance the visibility and accessibility of their scholarly work. Second, the paper delves into the advancements in publication metadata, specifically examining the impact of KBART[2] metadata for different types of linking in discovery services. It highlights the significance of accurate and comprehensive metadata for seamless navigation and access to content. Additionally, the study explores the role of metadata in search engine optimization and social media engagement, emphasizing the importance of adding metadata tags for improved visibility and engagement on these platforms. Third, the paper discusses the effective use of metadata to enhance accessibility, emphasizing practices such as optimizing persistent identifiers, PDF files, images, and other resources to ensure inclusivity and ease of access for all users. By addressing these key topics, the paper aims to shed light on the crucial role of metadata in various aspects of the publishing landscape and offers insights into best practices for optimizing metadata throughout the publishing lifecycle.

2. Maximizing discoverability: Best practices for authors' metadata

Author-provided metadata serves as the initial building block of the metadata pipeline, playing a crucial role in the discoverability of research articles. Authors frequently encounter challenges related to their articles' visibility in platforms such as Google Scholar, missing names in article results, incomplete display of articles in author profiles, and the need to enhance discoverability and usage.

A significant portion of the issues may come from author-provided metadata. For instance, misspellings or special characters in author names may cause Google Scholar to drop the author names from article search results. Even though in some countries, people have only given names without surnames, Google Scholar cannot handle that. If authors provide more accurate keywords, that will improve the chances of their articles being searched and found. Many authors have similar names, initials or even affiliations. Adding Author IDs such as ORCID ID can help with author disambiguation.

Inaccuracies in author-provided metadata contribute significantly to the challenges faced in the discoverability of research articles. Misspelled names and special characters can result in Google Scholar excluding author names from search results, and the platform currently has limitations in handling naming conventions without surnames. To improve search effectiveness, authors should use accurate keywords that align with their research content, enhancing article discoverability and search result inclusion. Author disambiguation is crucial due to the prevalence of similar names, initials, or affiliations. The inclusion of unique identifiers such as ORCID ID effectively distinguishes authors with similar attributes, ensuring correct attribution of their works and establishing a clear and recognizable digital presence within the scholarly community.

To address these challenges and support authors in optimizing their metadata, we have developed a comprehensive set of best practice tips. These guidelines are specifically curated to assist authors in

---

[2]NISO RP-9-2014, KBART: Knowledge Bases and Related Tools Recommended Practice. https://niso.org/publications/rp-9-2014-kbart, accessed September 14, 2023.

enhancing the discoverability, visibility, and accessibility of their scholarly work. By adhering to these recommendations, authors can maximize the impact and reach of their research on platforms such as IEEE Xplore, ultimately contributing to the broader dissemination and utilization of their findings[3].

- **Optimize article titles:** Authors should craft meaningful titles that accurately reflect the content of their articles. It is essential to include important terms upfront to improve searchability. Authors should consider the perspective of potential readers and utilize relevant keywords while keeping the titles concise. It is advisable to avoid special characters, abbreviations, and exaggerated claims, ensuring clarity and precision in the title.
- **Optimize author names:** Authors can optimize their names to ensure accuracy and consistency across publications. It is crucial to double-check spelling and special characters in author names. Capitalizing the first and last names helps maintain consistency and aids in identification. Additionally, authors are encouraged to include their ORCID ID, a unique identifier that enhances their visibility and attribution in scholarly literature. In cases where authors have only one name, duplicating it helps avoid any ambiguity.
- **Optimize keywords:** Authors should pay attention to selecting appropriate keywords that accurately represent the content of their articles. It is recommended to use a combination of thesaurus terms, encompassing both broad and narrow terms. Authors should strive to include indicative terms that capture the essence of their research. Thinking from the perspective of a potential reader and incorporating relevant search terms can significantly enhance discoverability.
- **Optimize abstracts:** The abstract serves as a concise summary of the article's content. Authors should focus on writing short and informative sentences that provide an overview of the research. Including important terms upfront in the abstract aids in searchability. Authors can also consider using repeated keywords and synonyms to further reinforce the relevance and visibility of their work.

By following these metadata best practices, authors can maximize the discoverability of their articles on IEEE Xplore. Optimized article titles, author names, keywords, and abstracts contribute to a comprehensive and effective metadata representation, enabling researchers and readers to easily locate and engage with their valuable contributions.

3. Advancing publication metadata: A multifaceted landscape

The world of publication metadata encompasses various stages, each guided by NISO standards to ensure effective management and utilization[4]. Let us explore some of these standards that contribute to different aspects of the publication lifecycle.

- **Information creation and curation** involve standards such as *JATS (Journal Article Tag Suite)*[5] and *Dublin Core Metadata Element Set*[6], among others. These standards provide structured frameworks

---

[3]IEEE Xplore Help. SEO Tips for Authors. https://ieeexplore.ieee.org/Xplorehelp/discovery-services/seo-tips-for-authors, accessed September 14, 2023.

[4]NISO Information Standards. https://niso.org/publications/standards, accessed September 14, 2023

[5]See: https://www.niso.org/publications/z3996-2021-jats, accessed September 14, 2023.

[6]See: https://www.dublincore.org/specifications/dublin-core/dces/, accessed September 14, 2023.

for capturing and organizing essential metadata elements, enabling comprehensive and consistent representation of scholarly content.

- When it comes to **information discovery**, the *Open Discovery Initiative (ODI)*[7] and *Access Licenses & Indicators*[8] *(ALI)* play pivotal roles. ODI focuses on enhancing the discoverability of scholarly resources, promoting interoperability and efficient access to information across various platforms. ALI, on the other hand, facilitates the identification and understanding of content usage rights, enabling better compliance and licensing practices.
- **Effective content linking** relies on standards such as *KBART (Knowledge Base And Related Tools)*[9] and *IOTA (Improving OpenURLs Through Analytics)*[10]. KBART enables accurate linking of resources by providing comprehensive title lists and holding information. IOTA leverages analytics to enhance the quality and reliability of OpenURLs, ensuring smooth and precise resource linking.
- **Authentication** is a crucial aspect, addressed by standards such as *ESPReSSO* (Establishing Suggested Practices Regarding Single Sign-On)[11] and *RA21*[12]. These standards promote secure and seamless access to scholarly content by establishing recommended practices for single sign-on authentication and identity management.
- To optimize **journal displays** on platforms, the *Presentation and Identification of E-Journals (PIE-J)*[13] standard provides guidelines for consistent and user-friendly journal presentation, enhancing the reading experience for users.
- Facilitating **publication transfers** between platforms, the *TRANSFER Code of Practice*[14] ensures smooth transitions while preserving metadata integrity and continuity of access.
- **Metrics** play a significant role in understanding the impact and usage of scholarly content. *SUSHI (Standardized Usage Statistics Harvesting Initiative)*[15] enables automated retrieval of usage statistics from platforms, while *Altmetric*s[16] provides alternative indicators of impact beyond traditional citation metrics.

By adhering to these standards, publishers, service providers, and libraries establish a foundation for effective metadata management, discovery, linking, authentication, display, transfer, and assessment. Embracing these standards fosters interoperability, consistency, and transparency, enhancing the accessibility, usability, and impact of scholarly publications.

4. Enhancing resource linking with KBART: A focus on link resolver knowledge bases

Among the various standards available, let us delve into one that holds significant importance in the library landscape: KBART (Knowledge Bases and Related Tools). KBART files serve as package-based

---

[7] See: https://www.niso.org/standards-committees/odi, accessed September 14, 2023.

[8] See: https://www.niso.org/standardds-committees/alo-revision, accessed September 14, 2023.

[9] See footnote #2.

[10] See: https://www.niso.org/standards-committees/iota, accessed September 14, 2023.

[11] See: https://www.niso.org/publications/rp-11-2011, accessed September 14, 2023.

[12] See: https://www.niso.org/standards-committees/ra21, accessed September 14, 2023.

[13] See: https://www.niso.org/standards-committees/pie-j, accessed September 14, 2023.

[14] See: https://www.niso.org./standards-committees/transfer, accessed September 14, 2023.

[15] See: https://www.niso.org/standards-committees/sushi, accessed September 14, 2023.

[16] See: https://en.wikipedia.org/wiki/Altmetrics, accessed September 214, 2023.

title lists, providing essential holding and title-level linking information for subscription products, aligning with the specifications set by the National Information Standards Organization (NISO)[17].

In the library ecosystem, prominent link resolver knowledge bases include EBSCO Full Text Finder, Clarivate's SFX, Alma, 360 Link, OCLC's Worldcat Linker, and several others. The range of related tools in this domain continues to expand, reflecting the evolving needs of libraries and users. Initially, link resolvers primarily served catalogs and abstracting and indexing (A&I) databases. However, over a decade ago, they gained significance as they started supporting discovery services, thereby enriching the user experience.

Today, KBART has wider implications beyond the realm of link resolvers. Elements such as ISSNs, included in KBART files, find relevance in search engine programs such as Google Scholar's Subscriber Link, CASA, and Universal CASA. This highlights the increasing integration of metadata standards such as KBART with search engines, promoting seamless access and discoverability of scholarly content.

By adopting the KBART specifications and incorporating them into their workflows, publishers, libraries, and service providers contribute to an ecosystem that facilitates efficient linking, discoverability, and access to valuable resources. The standardized approach provided by KBART benefits both information professionals and end-users, enhancing the visibility and usability of scholarly content across various platforms.

5. KBART fields: Enhancing metadata for improved resource linking

To facilitate effective resource linking and address common challenges associated with publication titles, the KBART standard provides a comprehensive set of metadata elements. These elements can be categorized into three groups: general, serials-specific, and monograph-specific.

In the general category, metadata elements such as publication_title_id and title_url play a crucial role in distinguishing publications with potentially unreliable titles due to misspellings, abbreviations, or other factors. These identifiers aid in accurately differentiating between similar or related publications, ensuring precise linking and access to the intended resources.

For link resolvers, the inclusion of publication IDs such as ISSNs (International Standard Serial Numbers) and ISBNs (International Standard Book Numbers) is essential. These identifiers enable link resolvers to store and establish connections between publications and articles. Accurate and complete ISSNs/ISBNs are vital to prevent gaps in holdings and avoid failures in OpenURL links. Furthermore, ISSNs are relied upon by Google Scholar for its Subscriber Link, CASA (Campus Activated Subscriber Access)[18], and Universal CASA programs, underscoring the significance of these identifiers in facilitating seamless access to scholarly content.

KBART Phase 2 introduced additional fields to address specific requirements. The parent_publication_title_id field establishes a link between a publication and its parent publication within a series, enabling users to navigate related content effectively. The preceding_publication_title_id field allows for tracking publication title changes by linking a publication to its previous title, ensuring continuity and accurate representation of historical metadata.

---

[17]See footnote #2.

[18]See: https://www.highwirepress.com/resources/data-sheets/casa-faq/, accessed September 14, 2023.

Moreover, the inclusion of access_type field serves to indicate whether a publication is paid or available as free/open access. This distinction is vital for users seeking specific types of content and helps inform subscription decisions for institutions and individuals.

Recognizing the evolving needs of link resolvers, the KBART Standing Committee is actively developing KBART III. This forthcoming iteration aims to address emerging complexities, such as indicating hybrid journals or flipped journals. By staying responsive to the evolving landscape, KBART III will continue to enhance the utility and effectiveness of metadata for link resolvers, supporting efficient access to scholarly resources.

The comprehensive metadata elements defined by the KBART standard significantly contribute to improved resource linking, accurate representation of publications, and seamless access to scholarly content. Publishers and providers can benefit from implementing these standardized fields to ensure enhanced discoverability, precise linking, and an optimized user experience.

6. Metadata for three types of linking in discovery services

Metadata is used for three different types of linking in Discovery Service Tools.

- **OpenURL linking:** The most traditional form of linking is OpenURL linking. It needs accurate metadata fields such as ISSN+volume+issue+page for a journal article, and ISBN+page for a book chapter. When ISSN, ISBN or some other metadata fields are incorrect in either the publisher data or in the link resolver knowledge base, the OpenURL links will break.
- **DOI linking:** Due to the frequent problems in OpenURL linking, a second form of linking was introduced: DOI linking. This is often more reliable than OpenURL linking, but not all articles have DOIs.
- **Direct linking:** To increase linking success, some publishers work with some discovery service providers to create Direct Links, using publisher-specific links including article ID or other metadata elements. Whatever the linking types, the quality of metadata is always highly important.

7. Metadata for enhanced search engine optimization and social media engagement

To optimize content visibility and reach a wider audience, the inclusion of metadata tags for search engines and social media platforms is essential. When it comes to web search engines like Google Search, it is crucial to incorporate unique metadata elements such as a distinctive title, descriptive summary, and canonical link URL. By doing so, publishers can enhance their content's discoverability and search engine rankings.

Similarly, for effective social media engagement, custom metadata tags specifically tailored for platforms such as Twitter can be utilized. These tags enable publishers to provide concise and captivating summaries, eye-catching images, and relevant hashtags, maximizing the impact of their content on social media channels.

Moreover, Google Scholar, as a specialized scholarly search engine, requires a comprehensive set of metadata tags. With over twenty special tags, notably beginning with citation_, Google Scholar caters to specific elements such as journal title, article title, volume, issue, page numbers, publisher information, and author details. Adhering to these metadata specifications ensures accurate indexing and representation of scholarly content on Google Scholar.

Additionally, for publishers participating in Google Scholar's Subscriber Link Program, the provision of metadata in XML format according to Google's guidelines is necessary. This includes essential information such as a journal's ISSNs (International Standard Serial Numbers) and the coverage range specific to each library customer. Complying with these specifications enables seamless access and streamlined linking for subscribers accessing scholarly materials through Google Scholar.

By integrating these metadata practices for search engine optimization and social media engagement, publishers can significantly enhance the discoverability, accessibility, and impact of their content across various digital platforms.

8. Enhancing accessibility through effective metadata practices

In today's publishing landscape, prioritizing accessibility has become a key focus for many publishers. Metadata plays a crucial role in improving accessibility, and careful attention is given to leveraging metadata to enhance this aspect. Publishers are actively exploring ways to optimize metadata for increased accessibility.

One important aspect is the optimization of Persistent Identifiers (PIDs) such as article DOIs, URLs, and author ORCID IDs. Publishers strive to include these identifiers whenever possible, as they facilitate reliable and persistent access to content. By incorporating PIDs into metadata, publishers contribute to a seamless accessibility experience for users.

Publishers can further enhance accessibility by optimizing PDF files. This involves adding relevant metadata to the PDF file properties, including file name, title, author names, and keywords. Placing important terms upfront in the metadata allows for better discoverability and accessibility of the content within the PDF document.

Additionally, publishers can optimize images by providing detailed captions and other relevant metadata. This practice aids users with visual impairments or those relying on assistive technologies to understand and engage with the content. Detailed image captions and metadata contribute to a more inclusive reading experience.

Beyond textual content, publishers also recognize the significance of metadata for software, codes, data, and reports. By incorporating metadata into these elements, publishers enhance the accessibility of these resources, enabling users to easily navigate, understand, and utilize them.

By adopting these metadata practices, publishers actively contribute to improving accessibility for a diverse range of users. The deliberate optimization of Persistent Identifiers, PDF files, images, and other resources ensures that content is more accessible, providing an inclusive experience for all users.

9. Collaborate to enhance metadata ecosystem

Ensuring the quality of metadata poses significant challenges for publishers. It is essential to recognize that metadata permeates various systems, both internally and externally. Internally, close collaboration among different teams and units is crucial. Additionally, external collaboration with library service providers, libraries, search engines, and end users is equally important.

For instance, distinct editorial and publication teams handle different content types such as journals, conference proceedings, standards, ebooks, and eLearning courses. However, the accuracy of metadata creation by one team does not guarantee the same level of accuracy across all teams. It becomes necessary

to monitor different teams and foster communication and knowledge exchange among them. Furthermore, even if metadata is created correctly, it does not guarantee proper storage across the various formats, databases, and systems such as XML files, FTP files, API files, and subscriber link files. Ongoing monitoring and troubleshooting are imperative.

Moreover, when handling metadata correctly internally, it is vital to ensure that external data partners accurately capture, store, and index the metadata. Valuable insights from inquiries and reports received from partners, customers, and end-users play a pivotal role in identifying and resolving metadata issues. This process is perpetual, requiring continuous efforts. Consequently, the focus remains on constructing robust metadata pipelines and rectifying any leaks that arise.

By fostering collaboration and maintaining vigilance, the metadata ecosystem can be strengthened, enhancing the overall quality and utility of metadata for publishers, service providers, libraries, search engines, and end users alike.

About the author

Julie Zhu is IEEE's Senior Manager, Discovery Partners. She cultivates and manages effective working relationships with Discovery Service, Link Resolver and Search Engine providers to maximize IEEE content visibility and findability in Discovery Service solutions implemented by IEEE institutional customers and search engines. She has been a member of NISO since 2009. She is a member of the KBART Standing Committee and a member of the ODI Standing Committee E-mail: Zhu.j@ieee.org.