

Leveraging Publisher Metadata for Universal Bibliographic Control

Satu Niininen

National Bibliography Services, The National Library of Finland, Helsinki, Finland.

E-mail address: satu.niininen@helsinki.fi



Copyright © 2025 by **Satu Niininen**. This work is made available under the terms of the Creative Commons Attribution 4.0

International License: <https://creativecommons.org/licenses/by/4.0/>

Abstract:

This paper investigates the integration of publisher metadata into bibliographic control to enhance the coverage of online publications in Finland's national bibliography. As the volume and diversity of online publications expands, traditional cataloging methods face significant challenges in achieving representative coverage of the national output. In response to these challenges, the National Library of Finland has implemented an automatic workflow for converting ONIX metadata into finalised bibliographic records.

The paper outlines the technical infrastructure and implementation process of the conversion workflow, while also discussing quality control measures to ensure the reliability of the national bibliography. Central to the discussion is the balance between the quality and quantity of metadata in relation to the Universal Bibliographic Control (UBC) core principle of metadata authoritativeness.

The establishment of a strategic partnership between libraries, publishers, and vendors is instrumental in fostering a metadata ecosystem that prioritizes data integrity, thereby supporting national bibliographies in their role of providing reliable access to information. Ultimately, the necessity for continuous adaptation in bibliographic practices is highlighted as a method to effectively meet the demands of the evolving digital publishing landscape.

Keywords: National Bibliographies, Universal Bibliographic Control, UBC, online publications, metadata

Introduction

The concept of Universal Bibliographic Control (UBC) serves as the essential backbone for creating and maintaining a national bibliography. However, adapting the UBC principles to the current, rapidly evolving digital publishing era presents certain challenges. Digital publishing has transformed the bibliographic landscape, increasing the volume and diversity of materials that require bibliographic control. Simultaneously the number of cataloging personnel is declining, creating a need for scalable solutions that traditional workflows cannot provide. As information proliferates across online platforms, the role of national bibliographic agencies (NBAs) is becoming increasingly complex, yet all the more necessary.

To enhance bibliographic control of online publications and to provide comprehensive access to Finland's national publishing output, the National Library of Finland has implemented an automatic conversion workflow for ONIX metadata from the publishing sector. While automatic metadata processing promises increased efficiency and coverage, it must navigate the delicate balance between metadata quantity and quality, a tension that strikes at the heart of UBC principles.

This paper examines the newly developed conversion workflow, outlining its impact on metadata quality and the strategic importance of fostering a collaborative metadata ecosystem with book trade stakeholders. The paper further discusses the challenges of maintaining bibliographic integrity when employing automated processes and explores the ways in which the NBA balances conflicting goals to maintain its role as a national repository of reliable metadata. Ultimately, the work presented here contributes to the ongoing discussion about the application of UBC principles in a digital age, offering insights into how national bibliographies can remain relevant and dynamic as publishing paradigms shift.

The Current State of Bibliographic Control in Finland

The Finnish national bibliography includes descriptions of the Finnish national output since 1488. It contains a wide range of material types, including print and online monographs, maps, music, sheet music, and periodicals. Currently, there are 2.3 million records in total, with an annual accrual of nearly 30,000 new records. Print monographs acquired as legal deposit submissions receive full level cataloging and authority control. This includes commercial, public sector, and university publications. Excluded materials are annual reports and financial statements of companies and institutions, ephemeral materials and dissertations below doctoral level.

For online publications, the state of bibliographic control has been significantly less comprehensive. With its current cataloging personnel, the NBA is able to describe only a fraction of all online materials published and collected

annually, making it necessary to allocate resources to prioritised materials only. Online materials assigned with full level descriptions are doctoral dissertations and legal deposit copies of publications that are vulnerable to loss as they are not collected or described by any other organisation, and are likely to be withdrawn from distribution within a short period of time. This set consists mostly grey literature, i.e. publications produced by organisations outside of the traditional commercial or academic publishing channels, leaving the bulk of commercial and academic online publications without any description other than the possible prepublication record.

This shortfall impacts several stakeholders as well as the National Library's own processes. Due to substantial gaps in coverage, the online publishing output cannot be consistently analysed or documented. As the bibliography also serves as the catalog for the national collection, comprehensive metadata is essential for the discoverability and use of legal deposit materials. The lack of descriptions has made it difficult to manage, access and preserve the online national collection, as these processes also rely on comprehensive metadata.

Collaboration with Publishers

The Finnish NBA has a longstanding and well-established partnership with representatives from the publishing sector. A significant milestone in this collaboration was achieved in 2009 with the establishment of the Finnish ONIX Centre. The centre is a joint effort between the National Library and Storia, a logistics service provider for the book trade. The ONIX Centre is tasked with managing the Finnish ONIX application, coordinating its use within the Finnish book market, and translating documentation and code lists as they are updated by EDItEUR. The centre includes participants from commercial and academic publishing houses as well as representatives of vendor platforms. Storia, as a member of EDItEUR, actively participates in the international development of the ONIX standard.

Storia also provides the National Library with continuous feed of ONIX metadata of all Finnish print, audio and electronic books in their catalog, which is then converted into prepublication records. This collaboration and the use of ONIX metadata as the foundation of prepublication records has been in operation for nearly twenty years, so the NBA has extensive experience in handling converted ONIX in a MARC21/RDA environment, and insight into its potential and limitations.

For the NBA, participation in the ONIX Centre is a strategically important position, offering direct contact to publishers and vendors, and the possibility to be influential in how ONIX is developed internationally and implemented nationally. Fostering a metadata ecosystem across the book industry landscape has been one of the long-term goals of the National Library. Beyond enhancing metadata quality, the initiative provides an opportunity to contribute to the implementation of new metadata elements both in book trade and in the

descriptive practices of Finnish libraries. As example of successful partnership is the recent introduction of accessibility metadata where the NBA has taken an active role. This involvement included collaborating through the ONIX centre in the creation of Accessibility Best Practices guidelines tailored for Finnish publishing houses. This has enabled the NBA to prepare for the European Union's accessibility legislation requirements for metadata well before the law came into effect, allowing the distribution of the resulting workload over a longer period of time. In the future, the partnership will also allow the automatic reuse of other relevant new metadata elements such as text and data mining opt-out information and other usage constraints for digital materials.

A robust network creates opportunities to enhance the integrity of ONIX metadata at its source by promptly notifying stakeholders of any inconsistencies. More reliable ONIX metadata leads to improved converted records, and benefits all institutions within the ecosystem, including publishers, vendor platforms, copyright management organisations, and libraries. The ultimate beneficiary of better metadata quality is the end user, whose search results are more reliable and accurate, whether using a vendor platform or a library catalogue. For publishers, the national bibliography acts as a showcase for their catalogue, making the accuracy of publication details a priority for both parties. In the Finnish book market, vendor platforms also enrich their proprietary keywords with subject headings from the national bibliography, further demonstrating that the benefits of cooperative effort are indeed mutual and aligning with UBC's goal of providing free and open access to trustworthy information.

Furthermore, through the partnership the NBA is ideally positioned to monitor trends and future directions in book publishing and retail. This includes observing changes in people's reading habits and their impact on published materials, formats, and distribution channels. Key shifts in recent years include the growing popularity of audiobooks, the transition of retail focus from online bookstores to subscription reading services, and the blurring of lines between podcasts and audiobooks. Anticipating similar changes helps the National Library strategically allocate resources in cataloging and legal deposit activities. According to annual publication statistics, the number of printed materials has not significantly decreased yet. However, in the future, online publishing may surpass print due to its ease and affordability, potentially becoming the dominant format in the market. If this occurs, the NBA may need to revise its selection practices to better reflect the prevailing situation, possibly shifting the focus of manual bibliographic control more towards digital content.

Outlining the Workflow

Given the largely positive experiences with using ONIX as a basis for prepublication records, the NBA has long aimed to extend the same process to handling electronic legal deposit items that cannot be described manually. This is a feasible endeavor because, in addition to prepublication metadata, the National Library also receives ONIX metadata with electronic legal deposit submissions of commercial publications.¹ However, unlike the conversion workflow for prepublication records, electronic legal deposit records are treated as automatically generated, finalised records without additional manual checking or editing. Eliminating manual intervention requires the process to meet higher standards and the produced records to be of higher quality and reliability than before. As a side benefit, improvements to the conversion have simultaneously improved the quality of prepublication records for print materials, which are still described manually.

The technical infrastructure of the workflow consists of a management system, a record import API and a replication component. The management system preprocesses the incoming data and tracks the state of items within the workflow. The record import API handles the conversion from ONIX to MARC21 and imports the record to the National Metadata Repository. The replication component allows the transfer of metadata from the repository to the national bibliography in a separate database. From there the metadata can then be used for different internal post-processing activities such as digital preservation.

Due to its complexity, the workflow took several years to develop, and it has been a joint effort between several sections within the National Library. The conversion was built on pre-existing ONIX to MARC21 crosswalk by OCLC² with an additional layer of refinements to fit better for the Finnish environment. The most time-consuming step in the development process was building a record merging functionality for the record import API. It is a component that detects if the publication has already been described in the shared repository. If a description is found, it updates the found record with metadata elements from the new converted record without losing anything valuable from the original record. This was a challenging task due to the size and complexity of the repository. Although it operates under shared description guidelines, with 300 institutions participating and over 17M records, there is notable variety which has complicated matching and merging of records.

The workflow was successfully implemented in late 2023. The converted records introduced a new record type with data elements and cataloging level not previously used in the shared repository, requiring updates to collective

¹ For more information on the collection of online legal deposits in Finland, see <https://urn.fi/URN:NBN:fi-fe2025051341628>

² The crosswalk is available at <https://www.editeur.org/96/ONIX-and-MARC21/>

descriptive guidelines by expert groups responsible for their maintenance. During the implementation phase, extensive communication was directed towards the cataloguing community to prepare for the transition. Through various informational webinars and workshops, other libraries and the wider cataloguing community were also able to provide feedback and suggestions for future development areas.

Since its introduction, the conversion workflow has processed a backlog of nearly 80,000 legal deposit submissions and created bibliographic descriptions that enable their discovery and use. For the national bibliography, this has been an outstanding improvement in the coverage and currency of online publications.

Quality Control

Although automatic workflows offer an attractive and cost-effective solution for metadata creation, they have obvious drawbacks. Bibliographies have been a result of intellectual work for a good reason, and even with the looming promise of AI soon fixing everything, currently there is no realistic alternative for manual description if the result is evaluated by data accuracy. In data conversions there is always something lost in translation, but measures can still be taken to minimise the risks.

Individual records may not undergo systematic manual intervention, but large batches are regularly reviewed for potential errors or other anomalies. When changes are made to the technical implementation of the workflow or new metadata sources are integrated into production, a team of experts reviews converted records in a test environment. Existing production pipelines are also reviewed regularly through spot checks to detect any corrective needs early on. Detected issues are organised into a backlog, analysed, and given a priority rating by a team of specialists from the NBA and other sections of the National Library. A diverse representation of services is considered important, as not all detected issues can or should be addressed. It is essential to weigh various perspectives when determining which ones require corrective action. Quality and performance enhancing improvements are implemented incrementally, demonstrating a commitment to continuous improvement. Furthermore, the ONIX data goes through sanity checks prior to conversion and if critical errors are encountered, for example the title is missing or content type is undefined, the workflow sends the publication to manual description workflow.

Known conversion errors can be roughly categorised into two types:

1. Input errors in source data
2. Interoperability errors between standards

The first category covers a range of cases where the source ONIX file contains errors. These can be misspellings, missing data elements, or data entered in the wrong ONIX field. Such errors can be impossible to detect beforehand in the source data but, if possible, they will be fixed retrospectively through mass

edits. Essentially, recurring errors are possible to be identified and fixed in the conversion, but random ones are not. So, for example, if the ONIX file contains a misspelling in the book title, the same error will unfortunately also recur in the final converted record, unless it is detected in routine spot checks. However, if a similar error recurs consistently, for example if a publication year is entered into the wrong ONIX field, conversion can be adjusted to handle the error. It is also possible to disregard specific data elements from specific sources or choose to prefer certain ONIX sources over others.

Errors caused by interoperability issues between standards represent a more complex problem type, stemming from the distinct purposes and target audiences of ONIX and library description standards. The focus of ONIX is on the commercial side of book data, and this fundamental difference can lead to field mismatches, where certain ONIX fields do not have direct equivalents in MARC21, making accurate data mapping difficult. Conversely, MARC21 may require more granular bibliographic information than ONIX provides.

A promising quality-enhancement feature is currently in development to refine converted records by utilising external data sources. The initial implementation of these lookup tools has been designed to verify ISSN identifiers provided in ONIX data against series authority records, ensuring accuracy in the series statement fields of converted records. In the future, this functionality can be extended to further enrich converted records with metadata elements extracted from their print equivalents, which will continue to undergo manual bibliographic control. This approach will allow contributor fields, authorised access points, and subject headings to be aligned with the manually described print versions. Once in production, this will significantly improve the reliability and cohesion of converted records while the NBA still operates in a MARC21 environment. Furthermore, when shifting to linked data, ONIX data elements such as publishers' proprietary work identifiers have shown potential to serve as an additional aid in instance clustering.

There should be continuous dialogue with the library community about quality issues and acceptable error types, particularly with institutions that participate in the shared metadata repository. It is important that members of the cataloguing community feel encouraged to reach out if they encounter data elements in converted records that complicate their cataloguing tasks. By doing so, the quality assurance process can be expanded and enhanced, engaging a broader community beyond the National Library.

However, the most efficient quality assurance method is to ensure source metadata is of consistent quality. The role NBAs hold within the national metadata landscape positions them as a connecting point between publishing and library sectors. They also have the necessary infrastructure and resources to take a leading position in fostering a collaborative metadata ecosystem for

all stakeholders who produce, supply, aggregate, or utilise publications and their metadata.

Alignment with UBC Goals and principles

Converted or otherwise automatically generated records cannot be considered equivalent to manually curated ones, and claiming otherwise would misrepresent the truth. This undeniably compromises the fundamental UBC goal of ensuring authoritative and reliable metadata. However, it is equally important for national bibliographies to remain relevant and dynamic by accurately and promptly reflecting the full scope of national publishing output. Excluding an increasing number of materials and material types does not offer a viable solution, as it would diminish National Bibliography's representativeness and go against the UBC goal of NBAs taking responsibility for the publishing output of their own region. In addition to the bibliography itself, other UBC stakeholders such as publishers also require comprehensive and up-to-date metadata. It is evident that there is tension between conflicting goals, so is it possible to find a satisfactory compromise between quality and quantity? While authoritativeness might remain unattainable through automatic processes, a defined set of standards can still be met, and to achieve this, the UBC statement offers a solid frame of reference for navigating a landscape of mutually exclusive goals.

The UBC statement provides a framework for the NBA's operations, but as a conceptual model, it does not offer practical guidelines for its implementation. It also does not address the mission and purposes of national bibliographies within their specific national contexts, which also define their operational parameters. National bibliographies are essential information resources, aiding selection and acquisition in libraries, serving as cataloguing tools, and verifying authorship and publication history. They function as datasets with applications across various societal sectors, supporting statistical analyses and research activities, particularly in digital humanities. Culturally, they preserve data of national works for future generations, providing insights into a country's identity and uniqueness. However, there are significant differences between countries in how these responsibilities and obligations are prioritised and due to this the same solution may not be applicable across all NBA contexts. (2022, p. 29)

To achieve an optimal balance, it is necessary to experiment, aligning with the UBC's call for ongoing improvement and adaptation to technological advancements. In Finland's national bibliography, nearly all bibliographic records have traditionally been full-level descriptions. A lower description level has been given only to a small number of exceptions, such as collection-level descriptions of ephemera materials. A significant shift took place in 2019 when, due to technical requirements of the new library system, all prepublication records had to be given a National Bibliography authentication code, effectively including them in the bibliography. Since then, the NBA has

explored incorporating different description levels for different publication subsets, aiming to improve coverage for publications that cannot be given full level descriptions.

To succeed, the process requires transparency on two different levels. Firstly, it is critical to explicitly define which materials can be described on a lower level through automated processes and which continue to require full-level descriptions. Secondly, automatically generated records should clearly state their provenience so they can be distinguished and separated from manually curated descriptions. An ideal implementation would be a user-friendly filtering option on the national bibliography search service, allowing users to choose between exploring a “full-level view” (full-level descriptions only, significant gaps in coverage) or a “mixed-level view” (different descriptive levels, better coverage and currency). This would allow the national bibliographies to better serve the diverse information needs of their various user groups.

Future Directions

Moving forward, further development of the workflow will proceed on two parallel fronts. Firstly, efforts are geared towards enhancing both the technical infrastructure and the quality of converted records. Secondly, the focus will also be on incorporating new data sources into the workflow.

Refinement of conversion rules is an ongoing task, but the emphasis is shifting towards enrichments processes, for example by automatically linking contributors in converted records to NBA’s name authority files through ISNI identifiers. As an ISNI (International Standard Name Identifier) Registration Agency, the National Library has collaborated in several cross-sectoral projects that aim to broaden the implementation of ISNIs across domains. Currently ongoing projects are carried out in collaboration with copyright management organisations, and previous work has been done with publishers and vendors encouraging them to adopt ISNIs more extensively in their systems and workflows.³ As a result, several publishers currently include ISNI identifiers in their ONIX data, opening the opportunity to more reliably identify contributors in the conversion workflow.

In 2024, the next phase was initiated, with the same conversion and record import workflow being developed for open access publications in institutional repositories with Dublin Core metadata. The data conversion from Dublin Core to MARC21 is currently in review phase, and the first pilot repository will move to implementation phase later in 2025. Most university and research organisation repositories in Finland are centrally maintained by the National Library. Therefore, networks are already in place, but extending the conversion process to include the handling of repository metadata introduces a new

³ For more information, see <https://www.kansalliskirjasto.fi/en/projects/adoption-isnis-copyright-management-organisations>

dimension to this cooperation. This shift will require more effective dialogue to ensure that the metadata across publication repositories remains consistent and cohesive.

Subsequent third phase will involve DDEX metadata of online music. The initial version of data conversion has already been developed, but the record import process is expected to be more complicated than in ONIX and Dublin Core workflows due to the need to handle both constituent unit and host item records.

Conclusion

This paper has examined how publisher metadata can be leveraged to enhance Universal Bibliographic Control. As the digital landscape evolves, traditional bibliographic methods need to adapt to manage the increasing volume of online content. When implemented carefully, automated workflows for converting ONIX metadata have shown promise in improving the currency and coverage of national bibliographies.

In addition to developing practices and policies within NBAs, there is a need for outreach initiatives directed toward book trade stakeholders in order to foster a mutually beneficial metadata ecosystem. Achieving close collaboration between libraries, publishers and vendors requires a shared vision that prioritises data integrity and an atmosphere of mutual trust.

One of the strengths of UBC is that as a high-level statement it leaves room for interpretation, allowing it to be flexibly applied across diverse contexts. However, to support this flexibility on a more pragmatic level, NBAs also need discussion, guidelines and use cases on how to make sustainable choices when aligning their selection and description policies with the changing publishing landscape. As the explosion of online content continues, UBC's mission to provide access to reliable information becomes increasingly important, and it is a crucial component in maintaining accurate metadata in this evolving environment.

References

IFLA Bibliography Section (2022). *Common Practices for National Bibliographies in the Digital Age*.

<https://repository.ifla.org/handle/123456789/2001>

IFLA Bibliography Section (2012). *IFLA Professional Statement on Universal Bibliographic Control*. <https://repository.ifla.org/handle/123456789/448>