

The International Standard Name Identifier: extending identity management across the global metadata supply chain

Andrew MacEwan^(a)

a) The British Library

Contact: Andrew MacEwan, andrew.macewan@bl.uk

Received: 12 April 2021; **Accepted:** 3 June 2021; **First Published:** 15 January 2022

ABSTRACT

This article describes how ISNI is being adopted as a common identifier across disparate sectors of publishing. Whether publishing and distributing recorded music, film or text ISNI is making good identity management a staple element in the global metadata supply chain. As the content creation industries become more engaged with the value of embedding good metadata from the point of publication libraries can look forward to benefitting from a truly global revolution in the metadata supply flow. A case study describes how a British Library project has taken ISNIs already in the British National Bibliography and cross-matched them with data from UK publishers' own databases to embed ISNIs into the book supply chain. It also describes plans for ongoing publisher engagement through implementation of ISNI assignment into its cataloguing-in-publication workflows for UK legal deposit.

KEYWORDS

Authority control; Identity management; Identifiers; Names.

Introduction

According to the standard ISO 27729 the International Standard Name Identifier was originally conceived as a “bridge identifier” with the ambition that it would be used for the identification of public identities of parties involved throughout the media content industries in the creation, production, management, and content distribution chains. This paper provides a brief update on how this ambition is beginning to be realised through the growth in adoption of ISNI in different publishing supply chains. Whilst this is important for the growing utility of ISNI in breaking down metadata silos in relation to efficient name identification it is also important to contextualise this as part of a broader trend that is seeing the business of producing well-controlled metadata become part of the business of publishing in the age of digital supply and demand. This paper, however, will focus on ISNI as an exemplar of this trend and will report in particular on a British Library case study describing our engagement with a group of UK book publishers and other agencies to embed ISNIs in the book supply chain.

Metadata silos and the supply chains

Different forms of creative content are distributed in supply chain metadata silos specific to each content type. The standards followed in each supply chain are well documented on websites promoting their use. Text publishing is supported by metadata supplied in the ONIX schema, with enhanced subject access through THEMA subject codes and additional product control provided in the form of trade identifiers: ISBN, ISSN, EAN barcodes, DOI, etc., as described at the EDItEUR website (EDItEUR, n.d). The music industry mirrors this with the DDEX schema standard, underpinned by the use of identifiers to express products at varying levels of granularity: ISWC, ISRC, RIN, RDR, etc. all described at the DDEX website. Metadata standards for the film industry are described most comprehensively at the website for the Entertainment Industry Identifier Registry (EIDR, n.d). Library standards have the advantage of attempting to accommodate and describe different content types in common standards, but even so libraries too have also worked in their disconnected silos reflecting historical divisions in curation of different content types. At the British Library our Sound Archive, our general catalogue, and our manuscripts and archives are catalogued in separate databases that reflect the major differences in the types of content and the standards that we use to describe them.

Library metadata itself exists in a silo in the context of the global supply chains. We rely on crosswalks and mappings, such as ONIX to MARC, to re-use data from the supply chain in our library based schemas. We also rely heavily on industry standard identifiers like the ISBN and the ISSN to build efficient automated workflows that allow machine matching based data enhancements from multiple sources. Co-operative cataloguing, the efficient re-use and sharing of metadata between libraries, where possible via automated workflows, is a staple activity fundamental to the efficient realisation of bibliographic control in the library world. In recent years the same theme of better metadata standards to support efficiency, automation and re-use have become a hot topic in every commercial supply chain in the publishing world. There is interest both in improving end-to-end metadata supply chains within each content industry and in building crosswalks between supply chains where appropriate commonality exists. In a Whitepaper on identifiers for artists

(Movielabs, 2019), the company Movielabs reviewed existing approaches to name identification such as VIAF, ORCID and ISNI as potential models for managing identities for the film industry. The paper notes that the widespread adoption of ISNI in the music industry is a factor recommending ISNI adoption in the film industry, given high levels of commonality linking the sectors, rather than pursuing invention of another name identification standard.

An early example of building better metadata solutions around commonality was the collaboration on the “RDA/ONIX Framework for Resource Categorisation” (JSC-AACR, 2006) that connected the work of the revision of the Anglo-American Cataloguing Rules with the development of the ONIX standard in the publishing industry. In 2014 the Linked Content Coalition published a paper, “Principles of Identification” (Paskin & Rust, 2014) that highlighted the content neutral potential of ISNI as a name identifier that could be used across multiple supply chains. Most recently the UK standards body, Book Industry Communications, has launched a Metadata Capability Directory (Matthews, 2020) to promote and improve the use of metadata standards in the end-to-end text publishing supply chain. The Directory is intended to be a platform where the use of standards across the supply chain can be compared, deficiencies and opportunities identified, and collaboration on solutions initiated. In the music industry the by-line on the DDEX website perhaps best summarises the conversations and initiatives that are taking place in every supply chain: “DDEX is a standards setting organisation focused on the creation of digital value chain standards to make the exchange of data and information across the music industry more efficient.” (DDEX, 2021)

This brief outline of the wider supply chain serves to highlight ISNI's place in the digital ecosystem of the global supply chains, but it also serves as a reminder that library metadata exists in the context of those supply chains and has the potential to benefit from the growing commercial interest in making metadata work better.

ISNI's place in the supply chain

The focus of the rest of this paper is on ISNI as a specific exemplar of a content neutral standard for name disambiguation that is starting to fulfil its purpose as a bridge identifier across sector specific silos for metadata. The foundation of ISNI in library metadata means that it already provides identification for authors, musicians, actors, editors, producers, artists and supports identification of both individuals and groups or organisations. In recent years, adoption has been strongest in the library sector and the music sector, with building blocks in place to encourage more widespread use in the book supply chain. ISNI's ability to work across so many specialist domains is based on a hub and spoke model in which Registration Agencies and Members provide sector expertise but work with a common database in the ISNI Assignment System, maintained by OCLC.

ISNI at work in the supply chain

In the music industry the ISNI membership list is growing. YouTube, Apple, Spotify and both major and minor record labels are set to be users of ISNIs and a growing network of music metadata

organisations specializing in rights, credit and attribution of content to artists and performers are providing the engine rooms for the supply of ISNIs to the music industry. Currently listed on the ISNI website from the music sector (alongside YouTube, Apple and Spotify) are SoundExchange, Quansic, Qanawat, Consolidated Independent, Jaxsta, @Musiekweb, Muso.AI, The ISRC Team and Soundways. (ISNI, n.d.) The last of these, Soundways, is a sound engineering company that has built an ISNI Registration Service within its Sound Credit system. Soundways describe it themselves on the ISNI website: “Sound Credit’s ISNI registration system is part of its larger system for music crediting, using Sound Credit’s new massive cloud profile feature. Once music creators and engineers set up a free profile, they can be instantly credited simply by entering an email address, swiping a card at a kiosk, or selecting their profile in an app. Any credited profile in Sound Credit will automatically attribute their ISNI code to every project involving that creator, along with other identifier codes such as the IPI/CAE or IPN that users can optionally enter” (Sound Credit, 2020). The interface with the ISNI central database emphasizes search and entering rich metadata to ensure that each ISNI is unique in the central database whilst local control of identities is maintained in the Sound Credit system itself.

An example of similar intention in the book publishing industry came in January 2020, when the Frankfurt-based technology and information provider MVB took on the role of an ISNI RAG operating in Germany, Austria and Switzerland. The first step will be to assign automatically an ISNI to all creators listed in the Verzeichnis Lieferbarer Bücher (VLB), the books-in-print catalogue used in the German-speaking world. In a second step, publishers whose books are listed in the VLB will be able to register new ISNIs for the creators of their works – directly from the catalogue, and free of charge. (MVB, 2020)

The British Library and ISNI

The British Library has a long standing involvement with ISNI from being a member of the ISO 27729 International Standard Name Identifier Committee to draft the standard to becoming one of the Founding Members of ISNI acting jointly with the Bibliothèque nationale de France to co-represent the Conference of European National Librarians (CENL) on the ISNI Board. Working with the Bibliothèque nationale and OCLC we supported the foundational work to build the initial ISNI database from VIAF and other data sources. The BL and the BnF have continued to provide quality assurance services to the ISNI International Agency for the ongoing maintenance of the ISNI database.

When the British Library became an ISNI Registration Agency in its own right this marked a strategic shift in our goals for authority control away from name disambiguation in the British National Bibliography (BNB) and in our catalogues towards bridging data silos and exploiting the potential of a numeric identifier to build and embed identity management into the supply chain. There are three guiding principles for our implementation of ISNI:

1. Embed ISNI in all our cataloguing workflows
2. Automate processes as far as possible
3. Engage with the supply chain

Pursuing these principles involves overcoming significant challenges. The British Library’s cata-

loguing workflows with regard to authority control use the LC/NACO file. We hold a complete mirror copy of the LC/NACO file in our Aleph cataloguing system and maintain currency with the other LC/NACO nodes through daily file exchanges. Integrating ISNI into our authority control workflows will require ISNIs to be uploaded into this LC/NACO shared resource. Conversations and planning for this to happen at scale are ongoing with the Library of Congress and the Program for Cooperative Cataloging, but it is evident that capturing and loading all the ISNIs already associated with NACO records within the ISNI database will take place over an extensive time period. In the meantime we have focused on getting ISNIs into our legacy bibliographic data and engaging with the UK publishing supply chain. Happily these two endeavours have worked in concert as will be described below.

A British Library case study in supply chain engagement.

Serious engagement with publishers and other actors in the UK supply chain was initiated in two facilitated meetings in early 2018. In January 2018 Publisher Licensing Services, an organization providing collective licensing and rights management services for the publishing sector, and an ISNI member organization, hosted a meeting for publishers to discuss the potential use of ISNI for improving identification of publishers and imprints in the supply chain. This discussion led to a follow up meeting in March hosted by Book Industry Communication to explore the wider topic of ISNI for authors, publishers and imprints. A colleague from the Bibliothèque nationale joined this meeting to give a presentation on their integration of ISNI into their cataloguing-in-publication workflows for French legal deposit. Thanks to further advocacy and promotion by EDItEUR the interest sparked by both these meetings led to the establishment of an informal UK Publishers Interest Group comprising the following organisations:

- Bibliographic Data Services (BL's CIP subcontractor)
- Book Industry Communication
- British Library
- Cambridge University Press
- EDItEUR
- Hachette UK
- International ISBN Agency
- Harper Collins
- ISNI International Agency
- Nielsen Book (UK ISBN Agency)
- Pan Macmillan
- Penguin/Random House
- Publisher Licensing Services
- Bloomsbury

Early on the group settled on a remit to explore practical solutions for disseminating ISNIs that were already established in the ISNI database into bibliographic product records that were already held in common by publishers and aggregators and the British National Bibliography. It was agreed that the quickest way to demonstrate value at scale and to introduce ISNIs into the supply

chain was to exploit what ISNI had already achieved in building its database of identifiers. Since the group as a whole had many different levels of capability for handling varieties of ONIX and MARC data it was also settled upon to make CSV files the medium of exchanging data between the British Library and the publishers themselves.

The starting point for the work was to get ISNIs into the British National Bibliography. Names in records in the BNB are the established name forms found in the LC/NACO file. We already had staff experienced in working with the Virtual International Authority File (VIAF) to associate VIAF and NACO IDs with the Linked Open Data version of the BNB. We also already had established links from ISBNs for product records, names in those records and LC/NACO IDs. By using the VIAF links we were able to pull across all ISNI-LC/NACO associations already established in VIAF clusters and bring the ISNIs back into the BNB. This provided us with a base file of 3,160,908 names in BNB records with assigned ISNIs for working with publishers' product data.

Each of the publishers in the working group provided us with sample files and later full back files as we developed the matching processes. Publishers provided us with a name string, their proprietary in house author ID, and its associated products. ISBNs were the key match point for identifying the target records and our staff developed algorithms to ensure we associated only confident matches between the LC/NACO name string and the publisher's name string to assign the corresponding ISNI. Differences between original publisher data and BNB catalogued data meant there were a variety of issues to work with: different name forms, punctuation and character set issues, reverse name forms, presence or absence of names for translators or illustrators, multiplicity of product ISBNs for the same work. The process was refined over time. Early results were quite variable between publishers and percentages of assignment relatively low in the first round of work. After several iterations and an expanded group of publishers' files to work with the latest results are as given in the table below.

Publisher	Number of names	Number of matches	Success rate
Atlantic	1,201	954	79%
Bloomsbury	43,558	28,420	65%
BurleighDodds	681	35	5%
ChannelView	1,392	1,146	82%
Canongate	521	363	70%
Cambridge University Press	21,298	16,292	76%
Dorling Kindersley	2,103	1,409	67%
Hachette	10,857	7,820	72%
Harper Collins	13,406	8,107	60%
Kogan Page	1,117	708	63%
Liverpool University Press	1,498	1,064	71%
PanMacmillan	1,642	1,332	81%
Penguin	14,297	9,638	67%



Publisher	Number of names	Number of matches	Success rate
Pluto	1,589	1,107	70%
Random House	24,060	16,127	67%
Taylor&Francis	107,871	68,878	64%
Total	247,091	163,400	66%

Fig. 1. Publishers' Author Name Matching Results

Generally, we have achieved a high level of consistency in the results and feedback from those publishers who have integrated the ISNIs into their own databases has confirmed the accuracy of the assignments from their side. An additional benefit that has come out of the work is cross deduplication of authors between publishers and in some instances deduplication within a publisher's own author file. The figures for deduplication are as given in the table below (Figure 2).

Publisher	Number of de-duplicated IDs (across all publishers)	Number of de-duplicated IDs (within publisher)
Atlantic	0	12
Bloomsbury	5409	619
BurleighDodds	0	0
ChannelView	0	0
Canongate	139	0
Cambridge University Press	3363	2225
Dorling Kindersley	398	62
Hachette	1940	746
Harper Collins	2350	119
Kogan Page	0	0
Liverpool University Press	0	4
PanMacmillan	524	33
Penguin	3449	846
Pluto	0	4
Random House	4386	1788
Taylor&Francis	8650	9916
Total	30608	16374

Fig. 2. Publishers' Authors Names Deduplication Results

Whilst the deduplication across publishers was an anticipated benefit of sharing a common supply chain author identifier, the cleanup of duplicates within a publisher's own data was an unexpected bonus, but one that demonstrated additional value in working across data silos. A further early bonus of this project with publisher data is the first example of a provided ISNI being re-used by Harper Collins in an ONIX record for a new publication by one of their authors. (Figure 3)

```

<TitleElement>
  <TitleElementLevel>01</TitleElementLevel>
  <NoPrefix/>
  <TitleWithoutPrefix textcase="02">Boy Giant</TitleWithoutPrefix>
  <Subtitle>Son of Gulliver</Subtitle>
</TitleElement>
<TitleStatement>Boy Giant: Son of Gulliver</TitleStatement>

<Contributor>
  <SequenceNumber>1</SequenceNumber>
  <ContributorRole>A01</ContributorRole>
  <NameIdentifier>
    <NameIDType>01</NameIDType>
    <IDTypeName>HCP UK Author ID</IDTypeName>
    <IDValue>4121</IDValue>
  </NameIdentifier>
  <NameIdentifier>
    <NameIDType>16</NameIDType>
    <IDValue>0000000121251907</IDValue>
  </NameIdentifier>
  <PersonName>Michael Morpurgo</PersonName>
  <PersonNameInverted>Morpurgo, Michael</PersonNameInverted>
  <NamesBeforeKey>Michael</NamesBeforeKey>
  <KeyNames>Morpurgo</KeyNames>
</Contributor>

<Contributor>
  <SequenceNumber>2</SequenceNumber>
  <ContributorRole>A12</ContributorRole>
  <NameIdentifier>
    <NameIDType>01</NameIDType>
    <IDTypeName>HCP UK Author ID</IDTypeName>
    <IDValue>1897</IDValue>
  </NameIdentifier>
  <NameIdentifier>
    <NameIDType>16</NameIDType>
    <IDValue>000000012147035X</IDValue>
  </NameIdentifier>
  <PersonName>Michael Foreman</PersonName>
  <PersonNameInverted>Foreman, Michael</PersonNameInverted>
  <NamesBeforeKey>Michael</NamesBeforeKey>
  <KeyNames>Foreman</KeyNames>
</Contributor>
  
```

Fig. 3. Example ONIX record containing ISNIs

Future work with publishers

The above results reflect the work we have achieved so far but the UK Publishers Interest Group continues to meet and we have more work to do. Although we do not think we can achieve much more improvement in the match rates through further improvements to our matching processes there may be improvements to be gained via more direct work with the ISNI database itself. Although the ISNI database began its life with a series of regular uploads of relevant records from the full VIAF database the last of these took place in 2016. Since then ISNI has worked with direct authority file loads from the increasing numbers of national libraries who have joined the ranks of the ISNI membership. The British Library has recently completed work on preparing an update file from its own copy of the LC/NACO database from 2016 to the present for submission to the ISNI database to bring LC/NACO up to date in the ISNI assignment system. Where possible this was enriched by associating title and ISBN data with the LC/NACO records extracted from the BL's own catalogues and the LC Books All file to facilitate the matching and the rich record assignment processes in the ISNI Assignment System. Following this load the total number of assigned ISNIs associated with a LC/NACO identity stands at 5,553,823 persons and 602,288

organisations. The results from this update will be used to re-run and fill some of the gaps in the publishers' results.

Following the above step the dialogue with the publisher group will move onto another stage. The high assignment rates already achieved have already opened up a conversation around and an appetite for 100% ISNI coverage in publishers' data. There are several avenues to explore for achieving this. The work to date has been an experimental project with the goal of seeding ISNIs at scale into the databases at the beginning of the supply chain. It has also been a mutually beneficial project to both publishers and the British National Bibliography. If we have run out of automated means to populate both the BNB and the publishers' databases then one option to provide a more intensive, manual level of intervention to fill the gaps could be a priced service for the remaining legacy data, acting in our role as an ISNI Registration Agency.

The other unresolved question though is the provision of new ISNIs for future authors. We are working on two solutions for this. One is the provision of a Registration Service portal for individual ISNI assignment requests. The second is the integration of ISNI assignment into our CIP workflows for our Legal Deposit intake and the development of a feedback loop to publishers along the lines already pioneered by the Bibliothèque nationale. Since the integration option accords with our lead principle of embedding ISNI into our workflows the steps to that are described first before concluding with an outline of the functionality of the Registration Service Portal.

Building a CIP workflow for ISNI assignment – next steps

As already noted earlier the British Library's cataloguing-in-publication programme has been contracted out and for many years has been provided by a company called Bibliographic Data Services (BDS). BDS have been a member of the ISNI UK Publishers Interest Group since its inception and they are closely engaged with the goal of embedding ISNI as a supply chain identifier. As part of the current CIP workflow BDS use and supply name headings from the LC/NACO file in their records. As a precursor to implementation of ISNI the British Library has already supplied a reconciliation file for corresponding LC/NACO – ISNI equivalents for BDS to use to facilitate automatic assignment on the back of their use of LC/NACO. The next step will be to update this file with an additional correspondence file based on the forthcoming update of the LC/NACO file in the ISNI database. Once this is in place BDS systems and workflows are primed and ready for implementation. At this point in time the details of a feedback loop to the publishers supplying BDS with pre-publication information to inform CIP work has yet to be determined, as does the potential role for BDS acting as a Registration Agency for original assignments, but the workflows as building blocks to inform those decisions will be in place.

Providing an ISNI Registration Self-Service Portal

The final piece of the British Library's engagement with the supply chain has been the development of an online service for individual requests. As part of our provision of quality assurance services to the ISNI-IA we respond to user queries and feedback, often leading to requests for updates and additions to existing records and requests for new assignments. We have firsthand experience of a wide

level of interest in ISNI amongst smaller publishers and directly from authors, artists, and performers across all repertoires of creative content. We are also acutely aware that at this level of interest and engagement the fact that only ISNI Registration Agencies and ISNI Members can register new ISNIs through privileged access to the ISNI assignment system interfaces is a barrier for those without the means to engage at the membership level. Since we are dealing with requests of all kinds at an individual level that is growing alongside the broadening engagement in ISNI we also know it is cumbersome and costly to deal with these individual requests sent to us through system-generated emails. As part of another project, involving cross-sector engagement with the music industry initiated by the British Library's National Sound Archive, the Mellon Foundation provided us with specific funding to build an End User Portal for ISNI assignment requests. We have now completed development of this system and it became operational in February 2020. The portal supports three main functions: Search, Request, and Add Data. The portal mediates these functions to interact directly with the ISNI system. The first two of these mirror the capability developed recently by Soundways in their Sound Credit system described above. As with the Sound Credit system the BL Service requires users to register on the system to access the functionality. Search is the critical first step to ensure that a pre-existing ISNI is not overlooked before submitting a request for a new ISNI. As a second check, when a request is submitted, it passes through the matching algorithms in the ISNI system in case a similar name identity does already exist. The Add Data function is an additional aspect of the service that will allow the many end users who want to enrich an ISNI record with additional titles or links to do so directly and easily. Editing existing data is not permitted because the ISNI database is built from the metadata of its Members and contributors and only members can edit their own data in an ISNI record. The British Library will regularly monitor and quality assure all activity and transactions that go through the portal.

Concluding reflections

This paper has sought to present a short update on the growing adoption of ISNI as a name identifier supporting different metadata supply chains. Although only a selection from all the activities going on across the ISNI network of 65+ Agencies and Members, it has provided examples that highlight drivers behind the interest from the supply chains. Drivers that position engagement with ISNI in the broader context of a more developed interest in the value of high quality metadata as an essential component in supply chain management for commerce, discovery and the attribution of rights. It has sought to contextualize the implications of this for libraries and our common interest in bibliographic control by showcasing just one approach, developed by the British Library, at engaging with the UK book publishing supply chain. We depend much on the supply of publishers' metadata but we have only had limited influence on bringing it into convergence with libraries' metadata requirements. Although authority control is only a single component of library metadata it has long been one of our most expensive metadata creation activities. Shifting the task of authority control into simultaneous management of ISNIs in the supply chain provides us with an opportunity to share that cost and to share its value. Metadata conceived and developed in the library sector, redefined as identity management, becomes a shared, common goal and the global supply chain becomes part of the solution.

References

- DDEX, n.d. Accessed June 2021. <https://ddex.net/>
- EDItEUR, n.d. Accessed June 2021. <https://www.editeur.org/>
- EIDR, n.d. Accessed June 2021. <https://www.eidr.org/standards-and-interoperability/>
- ISNI, n.d. Accessed May 2021. <https://isni.org/>
- Joint Steering Committee for Review of AACR, “RDA/ONIX Framework for Resource Categorisation”. Last modified august 3, 2006. <https://www.loc.gov/marc/marbi/2007/5chair10.pdf>
- Matthews, Peter. 2020, “Introducing the BICMetadata Capability Directory”. EDItEUR website last accessed June 2021. <https://www.editeur.org/3/Events/Event-Details/561>
- MVB. 2020. “MVB becomes an ISNI Registration Agency”. Press release posted on ISNI website, January 2, 2020. <https://isni.org/page/article-detail/mvb-becomes-an-isni-registration-agency/>
- Movielabs. 2019. “Creating a Talent Identifier for the Entertainment Industry”. Last modified August 28, 2019. https://movielabs.com/talentid/Talent_ID.pdf
- Paskin, N & Rust, G. April 2014. “Linked Content Coalition Principles of Identification”. Linked Content Coalition website. <http://doi.org/10.1000/287>
- Sound Credit. 2020. “Music industry ISNI registrations now free and automated”. Press release posted on ISNI website October 23, 2020. <https://isni.org/page/article-detail/music-industry-isni-registrations-now-free-and-automated/>