ISNI, a top tool for quality enhancement, smooth data flows and efficient internal processes

Ann Van Camp
Contemporary Collections, KBR (Royal Library of Belgium), Brussels, Belgium
ann.vancamp@kbr.be

Sven Lieber
Contemporary Collections, KBR (Royal Library of Belgium), Brussels, Belgium
sven.lieber@kbr.be

Abstract:

The International Standard Name Identifier (ISNI) is an ISO certified standardised number for uniquely identifying contributors to creative works. By resolving the problem of name ambiguity in global data flows, ISNI serves as an excellent tool to enhance the quality of authority data and acts as a bridge identifier across multiple domains and databases. In this presentation we will explain three major benefits of implementing ISNI at KBR, the Royal Library of Belgium: ISNI 1) enables us to enhance data quality, 2) facilitates smooth data exchanges, and 3) improves the efficiency of internal processes like cataloguing or monitoring legal deposit.

Keywords: quality control, authority data, ISNI, interoperability, efficiency

0 INTRODUCTION

The International Standard Name Identifier (ISNI) is an ISO certified standardised number for uniquely identifying contributors to creative works. By resolving the problem of name ambiguity in global data flows, ISNI serves as an excellent tool to enhance the quality of authority data and acts as a bridge identifier across multiple domains and databases. In this presentation we will explain three major benefits of implementing ISNI at KBR, the Royal Library of Belgium: ISNI 1) enables us to enhance data quality, 2) facilitates smooth data exchanges, and 3) improves internal processes like cataloguing or monitoring legal deposit. This presentation will only focus on quality control for authority data. Regarding quality control for bibliographic data at KBR, we refer to an article of Péter Király and Hannes...
Lowagie in the latest *IFLA Metadata Newsletter*¹. Before we will outline the affordances of implementing ISNI, we begin to comment on our choice to adopt ISNI at KBR.

### a – Why KBR needs ISNI

The trigger to adopt ISNI at KBR was the need for trustworthy and interoperable authority data in view of monitoring legal deposit. Monitoring legal deposit is particularly challenging in Belgium for two reasons. First, Belgian national heritage has always been and is still located on the crossroads of at least three language communities: 1) *Vlaanderen* or the Dutch-speaking community, 2) *Fédération Wallonie-Bruxelles* or the French-speaking Community, and 3) the *Deutschsprachigen Gemeinschaft* or the German-speaking Community. Secondly, Belgian legal deposit legislation includes not only all printed works published in Belgium – where the Belgian publisher is required to deposit –, but also all printed publications published abroad and written by a Belgian author who is domiciled in Belgium – in that case, the Belgian author is required to deposit.

Due to these challenges, KBR has two major needs in order to efficiently monitor legal deposit: we need to gather this dispersed and multilingual info from diverse and often siloed sources, and we also need to track and trace Belgian publications published anywhere in the world or translated in any language. With the aim of swiftly retrieving quality information on Belgian authors and their publications, KBR has started using ISNI, a persistent, language neutral, machine readable bridge identifier for contributors to creative works (mainly authors). Moreover, as national library of Belgium, one of the core functions of KBR is to publish a Bibliography of Belgium. In times of the Internet, accurately identifying the people who created our national heritage as primary entities becomes equally important as describing their works. Thus, a trustworthy name identifier is essential.

### b – Why KBR prefers ISNI to other name identifiers²

For KBR, the human data curation, high quality standards and scope of ISNI were decisive arguments to prefer ISNI to VIAF, Wikidata or ORCID.

**ISNI versus VIAF**

Of course, ISNI is not the first, nor the only name identifier which exists. Already in 2003 OCLC created a *Virtual International Authority File* (VIAF), that has since become a worldwide reference for libraries, archives and musea. When the central ISNI database was initiated in 2011, identity data from VIAF was used to seed the new database. So a great many of the existing (particularly the older) ISNI assignments included inputs from VIAF. Although OCLC manages both the VIAF-database and the ISNI-database – acting as ISNI Assignment Agency –, both initiatives differ fundamentally in several respects. As Anila Angjeli, Andrew MacEwan and Vincent Boulet point out, the main difference consists in the fact that ISNI provides “a unique, global, cross-domain, standard, persistent identifier for persons and organisations involved in creative contents”, and is “endowed with its own policymaking body and permanent control infrastructure”, which relies “both on computational methods and on human expertise”, whereas VIAF while applying an

---


² This part is translated and slightly adapted from a previous article in Dutch: Van Camp, Ann (2021, December) “ISNI als uniek verbindingselement”. *META: Tijdschrift voor bibliotheek & archief*, Jaargang 97, no. 9: 28-32.
automated clustering procedure “remains neutral towards differences in the cataloguing policy of its data contributors” and “does not create any data itself. It only combines and brings together data provided by its members”.

ISNI versus Wikidata
A more recent database which “seems to be evolving into a major authority hub”4, is Wikidata, the central knowledge graph to manage structured data of all Wikimedia Foundation projects like Wikipedia and Wikimedia Commons. In Wikidata you may find any entity or real world object, not solely persons and organisations from the creative sector. Anyone may contribute, as long as the data comply with the criterium of notability. Table 1 summarizes the comparison of ISNI with VIAF and Wikidata, based on the articles of Anila Angjeli, Andrew Mac Ewan & Vincent Boulet (2014) and of Carlo Bianchini, Stefano Bargioni & Camillo Carlo Pellizzari di San Girolamo (2021).

ISNI versus ORCID
A third name identifier complementary to ISNI, is ORCID, which is specifically meant to identify researchers, in connection with their publications, and their affiliations. ORCID and ISNI function independently from one another, but in 2013 they agreed to align both identifiers. Concretely, ISNI has reserved a set of numbers for ORCID, ranging from 0000 0001 5000 0007 to 0000 0003 5000 0001.

Table 1 – Comparison of ISNI with VIAF and Wikidata

<table>
<thead>
<tr>
<th></th>
<th>ISNI</th>
<th>VIAF</th>
<th>Wikidata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>Provide a unique, persistent, global and cross-domain identifier</td>
<td>Serve as worldwide reference for libraries, archives and musea</td>
<td>Centrally manage structured data of all Wikimedia Foundation projects</td>
</tr>
<tr>
<td>Type of entities</td>
<td>Creative sector: persons and organisations</td>
<td>Bibliographic entities: persons, organisations, works, expressions, geographic names, …</td>
<td>Any entity</td>
</tr>
<tr>
<td>Management</td>
<td>ISNI International Agency</td>
<td>OCLC</td>
<td>Wikimedia Foundation</td>
</tr>
<tr>
<td>Approach</td>
<td>Bottom-up: possible to contribute as registration</td>
<td>Top-down: harvests and clusters data of ca. 70</td>
<td>Bottom-up: freely editable by anyone</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>agency, member, data provider or end user</th>
<th>online sources selected by OCLC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quality</strong></td>
<td>Automated matching &amp; human curation (quality team + worldwide ISNI community)</td>
</tr>
<tr>
<td><strong>Number of records</strong></td>
<td>Ca. 13.07 mio persons (July 2022)</td>
</tr>
</tbody>
</table>

## 1 QUALITY ENHANCEMENT

When KBR became an ISNI Registration Agency in October 2020, a first step was retro identification matching our KBR person records with those described in the international ISNI database. In order to prepare a **bulk load submission of 800K records**, we worked to align our data with the ISNI elements and values\(^5\). That alignment implied adding certain data fields to our authority records or standardising the notation using Python scripts.

Actually, it was even while considering the option of becoming a registration agency that we had already started to examine the ISNI data elements and to **align our data**. In Spring 2020, we decided to add certain fields to describe authority data in our library management system. Besides MARC 21 field 024$a to include a standard identifier like ISNI or field 370$c to mention an associated country, the most important data fields were the fields to specify the birth date (046$f) and the death date (046$g) for a person, because before we only used the field 100$d which may contain various dates associated with a name (e.g. year of birth, death, flourishing period, …) and even qualifiers such as ‘ca.’ (circa) or ‘fl.’ (floruit). For a human eye, combined dates or text strings are perfectly understandable, but it goes without saying that notations like “ca. 1525-1569” are not easily machine readable, nor compliant with the ISO 8601 date format used by ISNI.

To **ensure a satisfying matching rate** with the ISNI database, we **enriched and cleaned our authority data** before submitting them to ISNI. From our catalog back-end we exported person authorities with additional publication samples in a tab delimited format specified by ISNI\(^6\). We developed Python scripts processing this file row by row to enrich and clean the data in a memory-efficient way. Based on available lists with additional birth and death dates, we could enrich a few hundred authorities. Furthermore, we standardised date values to comply with the ISO 8601 format YYYY-MM-DD, YYYY-MM or YYYY. This included extracting year values within brackets or replacing internally used None values, such as “?” or “[s.d]”; especially the last part was important as, according to ISNI, a potential match might be discarded due to date values not recognised by the ISNI data processing. Last but not least, the scripts also performed other corrections such as specifying “ISNI” as type for already present identifiers instead of VIAF when we algorithmically could determine that the identifier is an ISNI identifier\(^7\).

---


\(^7\) A 16 digit number starting with 0 and possibly ending with ‘X’
A next step was the **matching and assignment process** itself, that was performed by the ISNI Assignment Agency and closely monitored by the ISNI Quality Team, which is composed of members from the Bibliothèque nationale de France and the British Library. This batch load procedure manages quality control in four stages: 1) first quality checks regarding the format and richness, 2) a test data load, 3) an accept data load, and 4) a production data load. The ISNI Assignment Agency provided KBR with statistics on the results of the batch load and the ISNI Quality Team gave us valuable feedback on anomalies in our data that required further examination, for instance:

1. some names were not properly split in forenames and surnames because they were delimited with a semi-colon instead of a comma;
2. some person records actually appeared to be organisation records;
3. since the Quality Team found duplication and name errors in the provisional records and in part of the assigned records, they proposed a second data load in the ISNI Accept system after having adapted the configuration of the batch load process so that it would only allow merges with data that is already present in the ISNI database;
4. finally, the full data load in the ISNI Production system (= the live ISNI system) resulted in the assignment of 470,811 ISNIs to KBR records.

As Table 2 shows, the conclusion of this thorough review was the confirmation to upload more than 470K records in the ISNI Production system, or about 60% of what KBR initially submitted.

### Table 2 – Statistics on the results of the batch load

<table>
<thead>
<tr>
<th>Number of KBR records</th>
<th>Test Data Load</th>
<th>Accept Data Load 1</th>
<th>Accept Data Load 2*</th>
<th>Production Data Load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2021-11-25</td>
<td>2021-12-08</td>
<td>2022-02-21*</td>
<td>2022-03-21</td>
</tr>
<tr>
<td>read</td>
<td>822,590</td>
<td>822,590</td>
<td>822,590</td>
<td>822,590</td>
</tr>
<tr>
<td>errors</td>
<td>262</td>
<td>262</td>
<td>262</td>
<td>262</td>
</tr>
<tr>
<td>selected</td>
<td>822,328</td>
<td>822,328</td>
<td>822,328</td>
<td>822,328</td>
</tr>
<tr>
<td>assigned</td>
<td>684,636</td>
<td>684,755</td>
<td>470,717</td>
<td>470,811</td>
</tr>
<tr>
<td>provisional</td>
<td>134,275</td>
<td>134,144</td>
<td>27,187</td>
<td>27,180</td>
</tr>
<tr>
<td>of which:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>possible duplicates</td>
<td>74,964</td>
<td>79,034</td>
<td>6,660</td>
<td>6,653</td>
</tr>
<tr>
<td>single source KBR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>records</td>
<td>3,205</td>
<td>3,207</td>
<td>20,527</td>
<td>20,527</td>
</tr>
<tr>
<td>suspect</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

* New configuration: allows only merges with data that is already present in the ISNI database

The **listing of ISNI assignments** we then received allowed us to perform another quality check. For 35K records which already mentioned an ISNI in our library management system, we compared the results of manual look-ups with the automatically assigned ISNIs. In 99% of the records, both numbers corresponded, and for the remaining 1%, reviewing those contradictory ISNIs revealed mix-ups or duplicates, either in the KBR catalogue, or in the international ISNI database.

---

2 SMOOTH DATA EXCHANGES

In addition to the benefits for internal quality control, ISNI has also proven to be very useful as bridge identifier while exchanging data with third parties. The integration of ISNI in the KBR catalogue enabled us to enrich records with key information that trustworthy third parties had on the same ISNIs, and to swiftly compare KBR data with data from other national libraries for research purposes.

The integration of KBR authority data with data from other sources took mainly place in view of legal deposit and in the context of the BELTRANS research project on intra-Belgian translations NL-FR or FR-NL in the period 1970-2020. As we have presented during the DH Benelux 2022 conference, in order to create a FAIR data corpus of translations and contributors, we have gathered a plethora of other data sources in addition to the KBR-catalogue and developed a semi-automatic workflow of cleaning, enriching and integrating different types of heterogeneous data. The data sources vary from high-quality and structured information of other national libraries (BnF, KB, etc.) or the ISNI International Agency to crowd-sourced data sources such as Wikidata or smaller less structured databases specialised in translations (Index Translationum, Vertalingendatabase, DLBT, etc.) and even lists in Excel from public institutions that offered translation grants. The workflow is ongoing work but its components are publicly available in a GitHub repository under the MIT licence.

During this data integration process and especially while applying a BELTRANS-relevant filter such as the Belgian nationality of the author, illustrator, scenarist or publishing director, ISNI has proven to be an indispensable tool. It served as a bridge identifier to link all the Belgian persons from our KBR catalogue, an ISNI data dump, a BnF data dump, a KB SPARQL query and a Wikidata SPARQL query. We plan to further enrich the KBR catalogue with the integrated data from the BELTRANS linked database. For instance, since 2020, we have already doubled the amount of persons identified as Belgians in our catalogue, going from 14K to 28.6K Belgian authorities. Given that ISNI contains some 40K Belgian persons and Wikidata even 51K of ‘Belgian’ citizens, we still have plenty of authorities to label with the country ‘Belgium’, probably mostly persons linked to our heritage collections who lived in the Low Countries before the creation of Belgium.

---

11 https://www.bnf.fr/fr
12 https://www.kb.nl/
13 https://isni.org/
14 https://www.wikidata.org/
15 https://www.unesco.org/xtrans/bsform.aspx
16 https://letterenfonds.secure.force.com/vertalingendatabase/zocken
17 https://dlbt.univie.ac.at/
18 https://github.com/kbrbe/beltrans-data-integration
3 EFFICIENT INTERNAL PROCESSES

The next step is how to integrate ISNI in data flows for legal deposit and cataloguing. As we have explained in a previous article on how to monitor legal deposit\(^\text{19}\), at KBR we have started to import from selected bibliographic databases into our library management system pre-records of monographs that are subject to legal deposit legislation, but that have not been declared or deposited yet. Importing those pre-records presents multiple advantages:

- first of all, it enables KBR to permanently monitor and assess the level of compliance with legal deposit;
- it allows the Secretariat to easily register any incoming deposit, even when it arrives without any declaration (neither on paper nor online);
- it saves time for the cataloguing by the Agency of Bibliographic Information since the basic description has already been encoded;
- it prevents KBR from inadvertently purchasing works that are required to be deposited;
- centralising all information on the publications subject to legal deposit and on their publishers or authors facilitates and enhances collaboration between different teams.

For now, while importing those basic book descriptions, we have only automated the linking of the publication to its corresponding publisher thanks to a correlation list which makes the connection between a KBR authority and the corresponding publisher or imprint of Boekenbank\(^\text{20}\), the database of the Flemish ISBN-agency and book distribution platform managed by Meta4Books.

Another optimisation of the workflow for legal deposit and cataloguing that KBR pursues, is inspired by the examples of the Bibliothèque nationale de France (BnF) and the British Library (BL), who each in their own way are embedding ISNIs into the book supply chain. In France, the national library BnF has set up an ‘ISNI Request’ data flow based on the legal deposit declarations in ONIX\(^\text{21}\). This pioneering project was presented in an article by Jean-Charles Pajou (2016) and as follows during the IFLA WLIC 2017 in Poland:

The National Library of France (BnF, Bibliothèque nationale de France) is testing a new tool, called “ISNI Demande” (“ISNI Request”) which is designed for ingesting data on authors coming from publishers, creating provisional authority records, sending them to the ISNI international database, sending ISNIs back to publishers and giving librarians opportunities to use external data for establishing authority control on resources.\(^\text{22}\)

During the Frankfurter Buchmesse 2017 the French publisher Hachette Livre pointed out the mutual benefits of this cooperation between the national library and publishers: BnF gains more efficiency in processing Legal Deposit declarations and in cataloging with the help of


\(^{21}\) ONIX stands for ‘Online Information eXchange’. ONIX standards are designed by EDItEUR to support computer-to-computer communication and are all expressed in XML. ONIX for Books is “the international standard for representing and communicating book industry product information in electronic form” (EDItEUR, n.d. Accessed June 2022. https://www.editeur.org/).

ISNIs in ONIX feeds, while Hachette Livre benefits from the automated ISNI assignment for new contributors and from enhanced metadata quality\textsuperscript{23}. Since then, the ISNI request data flow has been fully implemented at BnF.

In the United Kingdom, on the other hand, the British Library has engaged with the UK book publishing supply chain with the aim “to exploit what ISNI had already achieved in building its database of identifiers”\textsuperscript{24}. As Andrew MacEwan describes in his update on the growing adoption of ISNI across the global metadata supply chain, they matched the LC/NACO name strings of the British National Bibliography and the publisher’s name string to assign the corresponding ISNI, taking into account the ISBNS of associated works.

Following those examples of the BnF and the BL, KBR also reached out to the book publishing professionals in Belgium. In Flanders, Meta4Books has shown a clear interest in integrating ISNI in the Dutch book supply chain: adding a proper unique identifier is part of their strategy to enrich data on ‘commercially relevant’ authors in a more consistent way, with the purpose to facilitate data flows and increase sales. Therefore, we have recently sent a compilation of 8678 authors with ISNI who are likely to be relevant for them. To facilitate the matching process with their data from Boekenbank, we have also added if known the ISBN-10 and ISBN-13 of an associated work, the name of an associated publisher, associated dates, etc. However, such a manually compiled list is not meant as a structural solution, but rather as an invitation to start using the ISNIs that have already been assigned to quite some Dutch speaking authors.

It is still very much work in progress, but for KBR the goal is to set up a (semi-)automated data flow to exchange bibliographic data and ISNIs of contributors with the Belgian book supply chain. If Meta4Books (Boekenbank) or publishers themselves provide KBR with ONIX files including ISNIs of contributors, then our cataloguing team can automatically link those book descriptions to the corresponding authority records. Or vice versa, if certain authors do not have an ISNI yet, KBR will systematically request ISNI assignments for contributors to deposited publications. This month, we have successfully tested automated ISNI requests via API in order to obtain ISNIs for new authorities (persons) who appear in the monthly lists of the Belgian Bibliography. What we have not developed yet, is an automated feedback loop from KBR to the book supply chain to share newly assigned ISNIs, but in the meanwhile the ISNI International Agency already offers several solutions to access existing ISNIs: via the public database or via the Linked Data resources\textsuperscript{25}. That way, ISNI will aid KBR in partially automating the cataloguing process, as well as in efficiently monitoring compliance with legal deposit. At the same time, the book supply chain will be able to reap the fruits of high quality metadata.

\textsuperscript{23} Audrain, Luc (2017, October 10) “ISNI assignment through Legal Deposit”. Presentation at EDItEUR Supply Chain Seminar during the Frankfurter Buchmesse, Frankfurt, Germany, Accessed July 2022.


\textsuperscript{25} https://isni.org/
To conclude, since KBR has become an ISNI Registration Agency, we are gaining better data quality, improved interoperability and more efficiency.

Acknowledgments
The authors wish to acknowledge the valuable cooperation with the Agency of Bibliographic Information, in particular with Hannes Lowagie and Julie Van Woensel.

References


