A Global strategy for authority metadata and identifiers
Lessons learnt from VIAF and ISNI and how to build a sustainable future

Discussion paper

July 2016

This paper is jointly proposed by the Bibliothèque nationale de France (BnF) and the British Library (BL). It summarizes our experience with VIAF and ISNI with the goal of **launching a discussion on future steps** that are needed to build successfully on these initiatives.

The use of both VIAF and ISNI has been spreading through the global information community in recent years to reach a critical point of interest, usage and maturity which we feel requires taking a fresh, high-level look at their respective and joint development, business and governance models.

This discussion paper proposes a first analysis of the current situation, in order to set the stage for a future joint effort, within the library community[1], to establish a consistent and sustainable strategy for their evolution.

1-Usage of VIAF and ISNI in the library community and beyond

The two major authority data repositories for names of persons and corporate bodies from the library sector that are currently maintained by OCLC, **VIAF** (*Virtual International Authority File*) and **ISNI** (*International Standard Name Identifier*), are **at the crossroads of major evolutions and key-challenges.**

Both repositories (combining an authority file – VIAF – and a registry of entities with unique identifiers – ISNI –), contribute to the overall exposure, discoverability and dissemination of library resources and services, in the perspective of the Semantic web and Linked Open Data (LOD). They empower interoperability and data exchanges with other business branches and communities (e.g. publishers, rights holders, cultural industries, researchers and digital humanities, etc.). **VIAF and ISNI have thus become important building blocks for many stakeholders** who started using them for innovation and new cooperation and business in the bibliographic and digital areas.

Both repositories address authority data (entities about people, works, things, places…), but they serve very different purposes:

- VIAF is built on existing national authority files and is used as **a source in the daily cataloguing processes** in libraries. It acts as **a data provider** for cultural heritage communities and research projects. It is also **a hub for the Web of data**: all major cultural Linked Data projects and services are linked to VIAF and it is even widely used by Wikipedia and Wikidata.

---

[1] This paper does not address the wider strategic context for ISNI as it is adopted outside the library sector.

- ISNI is **a registry** providing **stable and reliable identifiers** for public identities for persons and organisations. Those identifiers can be used to **make data exchanges processes more efficient**, for instance between libraries and publishers and rights holders, with the promise of sharing the costs in metadata production and dissemination across communities. Building on its status of **international ISO standard**, ISNI pushes for **a unique data model**.

Those differences can be summarized in the table below. The different origin, purpose and usage of the two registries explain that they both provide the library community with a service of a specific nature. While their similarities and differences may seem subtle to an outsider audience, they need to be clearly expressed in order to articulate the two registries for the sake of sustainability.

	VIAF	ISNI
Governance	An OCLC service, with an advisory board of library representatives (VIAF Council)	An ISO standard with a Maintenance agency, the ISNI International Agency (ISNI-IA). OCLC acts as a technical service provider
Business model	The service is provided free of charge by OCLC	Financial contribution by founding members and fee-based attribution of identifiers
Scope	Persons, organisations, works, expressions and geographical names	Persons and organisations
Source of the data	Libraries authority files, other contributors from the cultural heritage domain (museums, archives…)	Various data contributors (libraries authority files, rights management societies, researchers' databases, the music industry, the book supply chain) ; ongoing batchload of data and online assignments from ISNI Registration Agencies and Members for ISNI assignment
Data model	VIAF remains neutral between different changing standards and rules	ISNI is recognized as an ISO standard and pushes for a unique data model. ISNI has a developed quality control infrastructure and has developed data policies to reconcile the wide source base of data contributions.
Identifier persistence policy	Clusters created based on the alignments of data can change frequently	The aim is to stabilize the identified entities as persistently and reliably as possible. ISNIs are only assigned to clusters with a sufficient confidence rating assigned by the algorithms.
Updates	VIAF is updated monthly	Ongoing data loads according to the needs of Registration Agencies and Members; Online creation and editing of ISNI records by Members, Registration Agencies and also by the ISNI Quality Team (BL & BnF responding to End User requests)
Data dissemination and licensing	VIAF is distributed as Linked Data under an open license (ODC-By[2])	The ISNI registry is publicly available and can be searched via an API. ISNI data is under an open license with attribution. There are plans to develop ISNI as LOD
Main target audience	Libraries, cultural heritage, education and research	Global and multi-sector (libraries, archives, research, music and film industry, text publishing supply chain, rights management societies, etc.)
Example of usage	Source for catalogers; hub for aligning library entities with other repositories (such as Wikipedia and Wikidata)	Bridge between library and publisher data in order to automate the creation of links to authority records

In summary, the library community currently relies on VIAF for the dissemination and visibility of their data, while they develop a strong interest for ISNI in particular for the promotion of data exchanges between libraries and publishers / rights holders. Many institutions are currently investing on these services, creating pilot projects but also operational services. For this reason, managing

---

[2] http://opendatacommons.org/licenses/by/1.0/

VIAF and ISNI is no longer a matter of R&D. For many institutions, including OCLC, **the development and sustainability of these services will have a growing impact on their respective and mutual roadmaps.**

2-Current concerns regarding VIAF and ISNI

In this context, **many questions regarding the future and sustainability of both VIAF and ISNI have arisen.** These questions are of different nature.

a. **Technical concerns:** the technical landscape of VIAF and ISNI is impacted by various factors: the growing integration of VIAF in OCLC internal processes and services such as WorldCat and FAST, the complex interactions between VIAF and ISNI on the technical level[3], the constant evolution and improvement of alignments algorithms[4]… Frequent changes made to systems and data have sometimes been perceived by users as a "black box" effect. Such changes need to remain transparent for the community that is currently reusing these data and services in their own operational systems. There is a need for **a shared, public roadmap** between OCLC and other stakeholders, partners and users, so that the technical stacks of VIAF and ISNI evolve in a way that is consistent, transparent and agreed upon by the community as a whole.

b. **Governance model:** the ability to manage such a roadmap relies heavily on the efficiency of governing bodies for VIAF and ISNI, such as the VIAF Council and the ISNI-IA Board. It is important to ensure that these bodies have the ability to provide the expected input in terms of decision-making and strategic orientations[5]. **An analysis of their current functioning and the elaboration of requirements for a sustainable future** would be two important prerequisites for a revisit of the global governance model of VIAF and for developing a dialogue with the ISNI-IA Board concerning the library community's input to the strategic development of the ISNI standard.

c. **Market positioning:** While ISNI benefits from its status of ISO standard there is perceived competition with other identifiers (ORCID, in particular, which targets research and academic

---

[3] Providing data from VIAF to ISNI and retrieving data from ISNI isn't an automatic process. Moreover, both services are managed by different teams located on two continents. The opportunity of maintaining two databases facing similar challenges is regularly addressed.

[4] The OCLC technical team in Dublin has been highly responsive when it came to solving problems raised by data contributors in the past few years. The VIAF algorithm is now more transparent than in the past. Still, the 40 libraries or national networks which take part in VIAF are being kept informed of some major changes in the File sometimes after those changes have been completed, without participating in the decision-making process.

[5] The "VIAF Council" is an advisory body for strategic developments of VIAF. However, the key strategic evolutions for VIAF are insufficiently discussed in this body or, if they are, not always taken into account by OCLC. For instance, the question of interrelations between VIAF and ISNI has been addressed on several occasions since a technical meeting held in Paris in April 2013. Some recommendations were provided by the VIAF council, but there was no follow-up. There is a lack of strategic engagement among VIAF partners, possibly because representation in the group is more on the level of expertise than decision-making.

communities) which are growing their market position.. ISNI membership has also grown significantly in the last 12 months[6] and represents a significant commitment by many public institutions, state services and private partners seeking a common key to the reuse of massive data and also the development of new services and investments. Although there is an evolving conversation among libraries concerning the transition from authority control to identity management using persistent identifiers, ISNI is only perceived as one among many identifiers in the marketplace. The internal relationship between ISNI and VIAF both leveraging libraries' authority data serves to further dilute market positioning. This situation calls for **a more aggressive marketing and development agenda for ISNI** in the library community.

d. **Business model:** OCLC has invested significant resources for keeping VIAF a free service, which is truly appreciated. As far as ISNI is concerned, the situation is certainly more critical. The ISNI organization, still very much based on good will from its founding members, does not have the appropriate structure to scale up properly to meet an exponential growth in membership. The initial business model has some clearly apparent limitations, but a new one is still to be invented. OCLC has also invested far more resources in ISNI than it has received benefits so far. Considering the strategic interest that libraries have in both services, they could definitely consider participating more actively in their sustainability. In this perspective, **VIAF and ISNI business model should be reviewed not in parallel but together** as, from the end user perspective, they form a unique set or cluster of imbricated issues and resources.

3-Short term proposals to move forward collaboratively

BnF, BL and other major European and American academic libraries have become aware of this situation at a strategic level, and would like to deal with it in. We know that OCLC has its own doubts and questions. **We are sure that together we can find common solutions.** A strategic discussion among these partners could cover at least two major sets of issues:

1- **Clarification of the scope and goals of VIAF and ISNI**, both in terms of data and services. The distinction between VIAF and ISNI as exposed above should more precisely examined. The discussion could consider the respective user targets, functionalities and purposes of ISNI and VIAF services, the scope of their data, the data validation process in workflows, data quality, provenance, publicity and stability issues.

2- **Business and governance models for VIAF and ISNI:** based on shared concerns regarding the overall economic balance of these repositories and services, libraries along with OCLC could work on building alternative business models taking into account the needs of stakeholders (data remaining technically and legally open, while some services and products could come with a fee). We could open a discussion to put such issues on the table, taking into account investments made by all parties, services built on these investments, and the economic situation of stakeholders.

From the library community perspective three different viewpoints need to be reconciled:

---

[6] 11 Registration Agencies and 17 members (half of these are libraries)

- the needs and expectations of **end-users**,
- the legitimate economic and financial imperatives of **OCLC**,
- the services and investments built by data providers, including **libraries and archives**.

With regard to ISNI these viewpoints need to be reconciled with the wider strategic goals and user constituencies represented by the interests of the ISNI-IA Board.

Reviewing all of those with a constructive and creative mind could help interested stakeholders to **elaborate a sustainable strategy for authority metadata**. We suggest seizing the opportunity of the **next WLIC conference in Columbus** to gather interested parties and start the discussion. The following step could be an **Authority Metadata summit in London** under the auspices of BL and CENL.

Contacts:

Bibliothèque nationale de France : Emmanuelle Bermès (emmanuelle.bermes@bnf.fr), Vincent Boulet (vincent.boulet@bnf.fr), Isabelle Nyffenegger (isabelle.nyffenegger@bnf.fr)

British Library : Caroline Brazier (Caroline.Brazier@bl.uk), Andrew MacEwan (AndrewMacEwan@bl.uk)