

# How to build an «Identifiers’ policy»: the BnF use case

Vincent Boulet<sup>(a)</sup>

a) Bibliothèque nationale de France

**Contact:** Vincent Boulet, [vincent.boulet@bnf.fr](mailto:vincent.boulet@bnf.fr)

**Received:** 15 July 2021; **Accepted:** 12 September 2021; **First Published:** 15 January 2022

## ABSTRACT

Identifiers are at the crossroads of two interconnected, major evolutions which heavily impact national libraries: the massification of dataflow, redrawing the place libraries occupy within the global and national data ecosystem in a shared environment, and the strategic shift towards entity management underlying behind the new professional practices and standards. Based on the experience and maturation libraries are gaining in this field, the time maybe has come to formalize them and to highlight the impressive strike force libraries could have in a highly competitive landscape. This is the aim the Bibliothèque nationale de France is trying to reach by publishing an identifiers’ policy. It comes as the last part of a triptych after the new cataloguing policy (2016, including the indexing policy published in 2017) and the quality policy (2019). This identifiers’ policy is intended to clarify why and on what grounds a national library could, more or less, get involved in a given identifier, taking into account the diversity of scope, governance structure and business model of identifiers, be they international (for instance: ISNI, ISSN, ARK) or local (for instance: the BnF proper identifiers). Therefore, the identifiers’ policy highlights why it is necessary to use permanent, trustworthy identifiers and to what extent they are helpful in the daily working and quality control processes led by cataloguers. This is why the identifiers’ policy is not limited to principles, but has a very concrete dimension, both for internal and external issues.

## KEYWORDS

Identifiers; ISNI; BnF.

## Why an “identifiers’ policy” now ?

Libraries’ web presence now makes them familiar with identifiers and their uses. This presence poses for them several major and well-known challenges. We can summarize them as follows:

1. The adaptation of their system and data model to the requirements of research and “findability” of their resources *in* the Web. This global framework implies and fuels a needed, major shift of the data structuring, from a world where libraries used to standardize records for making them exchangeable into a world where libraries, along with other players, have to structure data for making them sharable. This issue is at the heart of the crucial problematic of the future of the bibliographic control and has many, crucial implications. For instance, the division of the bibliographic world into bibliographic records and authority records is now close to an end. Therefore the emerging international standards go with the flow, be it the IFLA-LRM data model published by IFLA in 2017 or the new version RDA reshaped by the “3-R project” which became the official version of the RDA international cataloguing code last December. Both have the same underlying principle, namely an entity/relations-based overall model. It means, from the authority control point of view, to switch to logic based on entity management.
2. Resources in a digital world are increasingly more agile and more scalable, due to changes in research and uses’ practices. Have we to describe serials or articles published on several platforms? Have to describe coherent set of musical works or a given piece of music diffused by various platforms under various formats? That issue has major implications on legal deposit for digital sound, books and movies. This complex reality challenges the new, above-mentioned library models and cataloguing codes, as they have to take into account changing resources which do not necessary feel part of any idealistic pyramidal model. What is recorded now should not be considered as permanent.
3. The data flows are becoming more and more massive, as the metadata accompanying them. This is also a challenge both for the bibliographic control and for the consistency of library databases. It actually raises the question of how applicable cataloguing rules are for the whole data set libraries deal with. Here is the issue of quality control processes and quality policy, because quality processes can be applied differently according to different data sources and subsets. This makes the question of sourcing data crucial, both for data flows reused by libraries, and for data flows libraries disseminate to end-users.
4. The technical and legal opening of datasets and catalogues is one of the points to be considered for having really sharable data. It may also be a political, strategic commitment taken by public administration towards citizens. As far as the legal opening is concerned, it may also put on the table the issue of mentioning the source of the data and keeping it associated with the metadata produced by a given player.

All these challenges are well-known for the future of the bibliographic control and we have to draw consequences from them. The entity management is unthinkable and impossible without any identifier management and identifiers’ policy. The shift from labels (different forms of a name for a person for instance) to identifiers provides less ambiguous data and a kind of stability. Identifiers allow access points or labels to be treated as entities being differently usable according to

context and needs : this is the key-principle of the “nomen” entity in IFLA-LRM. This shift also improves interoperability of data in regards with different contexts<sup>1</sup>.

Beyond these principles and opportunities, libraries have nowadays to deal with a very scattered landscape, due to the wide variety in nature offered by identifiers they are using or have intention to use. We can distinguish:

- The global identifiers supported by an ISO standard, which ISO signs an agreement with an international agency about. They correspond to a specific business model and global governance, whose libraries are a part of, along with other players, like music and cultural industry or copyright management firms. Libraries take part in a global business and scientific framework deciding on attribution and possible uses of a given identifier, and they can act as basic members, or a registration center for a given community or a specific field. This is, for instance, the case for ISNI (ISO standard 27729:2012), ISSN (ISO standard 3297), and ISAN (ISO standard 15706-2). For instance, BnF hosts the French national ISSN center and has an official, nationwide responsibility on this identifier. BnF is, furthermore, an ISNI registration agency since 2014 for a specific dataset, corresponding to the scope of the French legal deposit and national bibliography. But BnF has no special responsibility on ISBN.
- The global identifiers which could be assimilated to a de facto standard, or are being engaged in a standardizing process, and which are supported by an users' community. For instance, ARK (*Archival Resource Key*), an identifier created by *California Digital Library* (CDL), intending for identifying all resources, both physical or digital, records from catalogues or even immaterial resources as concepts. ARK is based on some key-principles and on a community of players engaged to maintain them (“naming authorities”, being able to attribute ARK to their resources, and “addressing authorities”, being able to resolve the identifier in order to give through it access to resources, by applying a policy of permanence). Moreover, the ARK identifiers have an explicit structure, which make them a de facto standard. So, about ARK, BnF respects an engagement framework with an users' community.
- The specific identifiers BnF has itself set up and is maintaining for internal uses and management of its databases, as for instance internal numbers of bibliographic and authority records (for instance: FRBNF identifiers). But external players can reuse them when reusing these records. So, even if these identifiers have been designed for internal uses at the time of catalogues' automatization, they are also de facto external. BnF keeps the complete control on their maintenance.

So, this short review shows that, over time, successive projects and needs, identifiers have been piled up one on another. In the same time, we have been gaining gradually more maturity and more experience on the overall identifiers' issue.

Managing identifiers doesn't fall from the Jabal Musa as Ten Commandments, but is highly de-

---

<sup>1</sup> Gordon Dunsire and Mirna Wilner, “Authority versus authenticity: the shift from labels to identifiers”. In: *Authority, provenance, authenticity, evidence: selected papers from the conference and school Authority, provenance, authenticity, evidence*, Zadar, Croatia, October 2016. Edited by Mirna Willer, Anne J. Gilliland and Marijana Tomić. Zadar : Sveučilište u Zadru, 2018. p. 87-113.

pending on human and IT resources, on transparency in how these identifiers are managed and on what libraries intend to do with them. Nevertheless, the way of dealing with identifiers as a whole, of choosing them, of handling with them must be consistent with best practices given from a global perspective. We have already framework documents for them, endorsed by W3C<sup>2</sup>, by IFLA<sup>3</sup> or by other international authoritative bodies. But, the question, for a given library, could be raised from another perspective. From the point of view of a given institution, to what extent using and disseminating identifiers can be helpful for addressing its own role and tasks? What criteria can be used strategically to justify the commitment of the library in one or more identifiers, and, possibly, its non-involvement? Here is the aim of an identifiers' policy.

## Identifiers: a commitment story

Using identifiers highly depends on how committed or engaged libraries want to be. We can easily assume an activist aspect for conceiving and implementing policies. In its identifiers' policy, BnF defines the idea of « engagement » as following:

- For a given and explicit dataset, BnF integrates identifiers in its dataflow and in its development policy regarding metadata (for instance : ARK for every resource, and ISNI for “agent” entities). This is why the identifiers' policy is a follow-up of the BnF quality policy. Identifiers are a tool to delineate specific data subset on which a specific quality control can be applied. It is also helpful to automatize some data processing, by helping interconnections of data. For instance, one of the projects BnF ISNI registration agency is developing is to propose alignments between EAN and ISNI so as to help cataloguers to create links between bibliographic and authority records (and, tomorrow, between manifestations, works and agents).
- BnF ensures, through identifiers, persistence of accessibility to its resources, in a broader meaning of the word: physical resources, digital resources (both digital version of physical documents, and natively digital resources), metadata describing and identifying resources. Identifiers ensure how trustworthy resources are identified for end-users.
- BnF builds up for end-users specific services and transactions thanks to identifiers, being based on its status of national bibliographic agency. For example, the BnF ISNI registration agency has built some transactions with the French book supply chain to register and disseminate ISNIs for their authors.
- BnF disseminates identifiers and resources for free, thanks to legal and technical opening. From this regard, the identifiers' policy is a follow-up of the open data policy BnF has set up as early as 2011 for data.bnf.fr and as 2014 for every resource.

In other words, the identifiers' policy ensures: to have easily disseminated resources, for the broadest communities, to have traceable, linkable, visible and discoverable resources.

This is the reason why the identifiers' policy is at the crossroads of the strategic shift made by BnF

---

<sup>2</sup> Data on the Web Best Practices, W3C recommendation, 31<sup>st</sup> January 2017 (<https://www.w3.org/TR/2017/REC-dwbp-20170131/>)

<sup>3</sup> Best Practice for National Bibliographic Agencies in a Digital Age, <https://www.ifla.org/FR/node/8786>

under the name of « bibliographic transition »<sup>4</sup>, and embodied by several strategic documents : the statement on open data (2014), the cataloguing policy (2017)<sup>5</sup>, the indexing policy (2018)<sup>6</sup> and the quality policy (2019)<sup>7</sup>. A global metadata policy is under preparation and should be published this year.

## Negotiating tensions

An identifiers' policy has to deal with three major tensions.

The first tension is the relationship between principles and concrete work and data libraries have to handle with. An identifiers' policy should be intended to give a general framework to action and to concrete involvement on identifiers, both internally, by integrating the identifiers management to the concrete dataflows and cataloguers' work, and externally, for end-users. The question is not to add even more practices for a given identifier, but give practices global framework and direction. In other words, an identifiers' policy finds its role somewhere between a statement of principles on the one hand, and concrete practices and using in the other hand.

The second tension regards the relationship between a common policy and the diversity of identifiers, as said above. It means handling with the diversity of identifiers themselves, and the diversity of how libraries can exercise some responsibility on them. Libraries can only use identifiers in their dataflows, without any significant role ; or they can attribute them ; or they can maintain alignments, or they can build up services for third parties, for instance for the library national community, or the book supply chain.

Here are, for instance, the different role BnF exercises, or intends to exercise on identifiers.

BnF role	International ISO identifiers	Identifiers with an international audience	Local identifiers
Attribution or registration responsibility	ISSN, ISNI	ARK	FRBNF
Identifiers BnF doesn't attribute, but BnF uses and builds services for the community on.	ISBN	EAN	
Identifiers which BnF develops alignments with		LCSH, MESH, GND, datos.bne.es, VIAF, NOMISNA, Geonames, Agrovoc, Wikidata	
Identifiers integrated in dataflows	ISAN	EIDR	

<sup>4</sup> For more details on the « Bibliographic transition » national programme, see : <https://www.transition-bibliographique.fr/enjeux/bibliographic-transition-in-france/>

<sup>5</sup> <https://www.bnf.fr/fr/politique-de-catalogage-dans-bnf-catalogue-general>

<sup>6</sup> <https://www.bnf.fr/fr/politique-dindexation>

<sup>7</sup> <https://www.bnf.fr/fr/politique-de-qualite-des-donnees>

We should distinguish “registration” from “attribution”. “Attribution” means that library attributes directly a given identifier, following international policies and rules. This is the case for ISSN, through ISSN French Centre, which *attributes* ISSN identifiers following rules and policies validated by the ISSN International Centre and ISSN international network. “Registration” means sending data for asking attribution to international authoritative body. For instance, BnF *registers* ISNI by sending authority records for names of persons and the bibliographic records linked to them to the ISNI International Attribution Agency, by getting back ISNIs attributed on its own data by this attribution agency, and by disseminating ISNIs through the book supply chain and the library community in France.

The third tension regards the necessity to keep a two-fold diachronic, dynamic approach. On the one hand, the international landscape of identifiers is moving. On the second one, the responsibility libraries can take on one given identifier can move, too. For instance, BnF is thinking about taking more responsibility on ISAN, ISWC and ISRC, depending on their business model, legal structure, on the one hand, and on resources BnF can invest on them, on the other hand.

Therefore, setting up an identifiers’ policy means to declare principles, on which BnF can commit itself, by taking into accounts these tensions, and concrete conditions allowing such a commitment by a State and non-for-profit institution to be concretely achieved.

## The policy content

The key-principle is permanence. The identifier shall give guarantees on permanence, which concretely implies for it to be based on shared, transparent governance, broad and, if possible, global community, sustainable business model, as for the identifier itself, as for the community using it, and a standardizing process. All these elements create trust in the opportunity of consuming human and financial resources to integrate the identifier in the library dataflows and in the services and engagement the library agrees on with other players.

We have also formulated four main conditions to make these principles concretely applied.

1. The identifiers must benefit from a broad and stable community or inter-community commitment. It implies that the identifier is part of a normative strategy:
  - either because it corresponds to an ISO standard (for example: ISO 27729: 2012 for the ISNI identifier; ISO 15706: 2002 and ISO 15706-2 for the ISAN identifier; ISO 3297 for the ISSN) and undergoes the international consultation process applied to periodically revised ISO standards;
  - or because it is part of a strategic standardizing process (for example: ARK<sup>8</sup>)

The identifier must therefore benefit from support of an international community or of a cross-domain commitment, depending on its scope of use. Its use must also be recognized and promoted by one or more communities. The governance of the identifier, whether at a national or international level, must be based on a written contract and allow the community or communities to be represented in decision-making bodies and to contribute to the technical and strategic orientations of the identifier.

---

<sup>8</sup> See above



The identifier must be based on a negotiated, transparent, stable, contractual and sustainable economic model for a public institution. Business model should enable the BnF to develop a medium and long-term policy of use and services for the communities it serves. It must also be balanced in order to provide guarantees of financial stability in the medium term. This is the case, for example, for the ISNI business model, which allows libraries overall business model of this identifier.

2. The identifier must have a clear and explicit application policy, in other terms, we must clearly know what does identify the identifier. For instance, we know to what entity ISNI is applied for, as described in the ISO corresponding standard, which put forward the concept of “public identity”, more or less similar to the concept of “bibliographic identity” libraries are familiar with.

The identifier must respect the principle of uniqueness. An identifier relates to one and only one resource. When a resource is stable, so is the identifier. When a resource changes to become something else, a new identifier must be assigned. Similarity and duplication issues need to be identified and addressed. For ARK, BnF is developing practices of redirection when merging two duplicates, for instance. The question is more sensible for concepts and remains under discussion for now, because two concepts are never exactly similar.

The identifier data model must be defined, documented and transparent. The attribution policy and the scope of data and resources to which the identifier applies must be stable, unambiguous and explicit. The conditions for attributing the identifier must be clear and explicit so as to control the mechanism and scope of their attribution, as well as their non-reassignment. This is why BnF has made explicit the scope of ARK, and has recently extend it to records for archives and manuscripts, so as to make every BnF resource covered by this identifier, without any regard to the data base describing it.

3. The identifier must be technically sustainable. The ID is built to last.

That means:

- It must be independent from the technical protocols to ensure its attribution and management, as well as of the authority that technically ensures its attribution. The guarantees of technical sustainability must be made explicit in the contractual commitments binding the national or international governance body on the one hand and the BnF on the other. This is the case for ISNI, for instance.
- The link between the identifier and the resource described must be permanent. The existence of the identified resource must be certified. We are developing the scope of the future French National Entity file (FNE), to be published around 2024, in this direction. The entity, and the identifiers associated to this must correspond to a real resource belonging to a member of the FNE network.
- An identifier must be maintained during and beyond the life of the resource that it identifies. If the resource or entity evolves, the persistent identifier must ensure a redirection to the most recent version of the resource or of the description of the entity to which it returns. The user must be informed of any significant change in the identified resource: deletions, replacements, merges, substantial modifications of the scope of the resource. The memory of the assignment of the identifier must thus be preserved.
- An identifier is never and under no circumstances reassigned.

- In accordance with W3C best practices, it is better for an identifier to be expressed as an URI, as allowed by ARK, for instance.
4. The identifier must be open and neutral politically and technically.  
This means :
- The identifier must be administered by an independent body contributing to the neutrality and uniqueness of the Web. It does not depend on exclusive mercantile interests that unilaterally could impose objectives, governance and an economic model incompatible with the requirements of a public institution. Dedicated and trained teams follow the attribution and registration procedures. This is the case with the ISNI governance structure and Quality Team.
  - The BnF favors identifiers that are opaque in their meaning in order to avoid the temptation to modify them if the resource or entity they identify changes and to allow their widest distribution.

## **Conclusion: audience and next steps**

The identifiers' policy is intended to have both an internal and external audience. It aims at explaining cataloguers' and librarians the main directions BnF is implementing, and at committing BnF in its coming discussions with end-users and management bodies of identifiers. The next steps are to concretely develop this policy for the identifiers already used in the workflow.

An identifiers' policy shows how important identifiers are for the future of bibliographic control, by accelerating and making consistent the overall shift of data structure towards entity management. We could say it is both a tool for managing this shift and the aim this shift is supposed to achieve, because it is a tool to redraw the library role and place in the global data ecosystem. It supposes not to have a defensive approach but to elaborate strategic orientations for making libraries not a customer or a victim, but a genuine player in this shift.