ISNI and VIAF – Transforming ways of trustfully consolidating identities

Anila Angjeli
Bibliothèque nationale de France, Département Information bibliographique et numérique, Paris, France.
ISNI International Agency
ISNI: 0000 0004 2755 4724

Andrew Mac Ewan
British Library, Department of Collections, London, UK.
ISNI International Agency
ISNI: 0000 0004 2367 7583

Vincent Boulet
Bibliothèque nationale de France, Département Information bibliographique et numérique, Paris, France.
VIAF Council
ISNI: 0000 0003 5723 4556

Abstract

This article presents ISNI and VIAF as major initiatives leveraging the library Authority control. They are significantly changing the international landscape of bibliographic control addressing the challenge of reliably identifying people in the rapidly emerging global knowledge network. Both systems are built on the principles of UBC. Further, they are revitalising these principles by promoting a global economy in which the library authority data are playing a leading role in the Linked Data chain. The article presents ISNI and VIAF in complementary relation with one another and invites libraries worldwide to become full players in both initiatives.

Keywords: VIAF, ISNI, Authority control, Standard global identifier, Data aggregation, UBC

Introduction

Libraries and allied cultural heritage communities have traditionally focused on collecting and cataloguing the products of human creative activities. Describing the people related to the resources has been a secondary activity serving the primary objective of providing access to the resources themselves. Increasingly, though, the emergence of the Internet as a social phenomenon has raised our awareness that ultimately people are central to our missions, as the producers of the resources in our care, but also as the consumers of these resources, and potentially as future producers of knowledge. We are increasingly realizing that we are fundamentally engaged in a vast, profound, and ongoing humanistic endeavour.

People, individuals as well as organisations, are primary entities that serve to interconnect information across repositories, collections and systems and, more broadly, in the open Web. Libraries' long tradition in creating, curating, and sharing authority data on such entities becomes an invaluable asset in this environment. Accurate identification and disambiguation, wealth of variant forms of names and other identifying information supported by citation of sources, respect of professional standards, and fine granular structure makes authority files ideal candidates for processing with advanced technologies, and for reuse in multiple ways, including uses outside of the international library community.

Devised in the 1970s, the concept of Universal Bibliographic Control (UBC) was both revolutionary and visionary – organise the whole economy of bibliographic data production and quality control on a universal scale by sharing responsibilities and benefits. In the context of Linked Open Data with the ever-increasing data traffic on a global scale, the issues of data quality and confidence have become pivotal and highly strategic. As new technologies enable the interrelation and interplay of data, the necessity of accurate data identifying people becomes urgent in order to ensure the integrity of the network of connections. In this environment, errors are promulgated, dispersed, and undermine the reliability of the network; and the errors, once dispersed, become difficult to discover and repair. The fundamental principles of UBC need to be re-envisioned in light of this environment.

VIAF and ISNI are here presented as major initiatives that are addressing the challenge of reliably identifying people in this rapidly emerging global knowledge network. The article discusses how both systems reaffirm the principles of UBC. Further, it demonstrates how VIAF and ISNI are significantly changing the international landscape of bibliographic control, as well as creating resources that can collectively serve as an axis of control for the vast knowledge network within which the libraries participate. Both are playing a prominent role in the aggregation, enhancement, and dissemination of library data, presenting the library community with the opportunity to serve a leading pivotal role in the global knowledge environment. The resources they are creating are increasingly impacting in the economy of data production through sharing the resource-intensive work of establishing reliable identities. VIAF and ISNI, in complementary relation with one another, are each contributing to this profoundly important emerging landscape.

Authority control and Identification

VIAF - focus on authority control

VIAF is a major source for authority control and is becoming *the* collective reference source at the international level (Bourdon and Boulet, 2013). Focused on the major reference library sources – mainly national libraries and countrywide union catalogues - VIAF currently

gathers the authority files of a large number of institutions.[1] In addition to data from national libraries and union catalogues, VIAF is also open to other sources focused on specific areas. Data from The Getty Institute and other specialized databases managed within the framework of scholarly projects, such as Perseus hosted by Tufts University, and the Syriac Reference Portal from Vanderbilt University, along with information harvested from Wikipedia, is branching VIAF out to the large encyclopaedic domain. Therefore, VIAF is positioning itself at the crossroad of the library data and the broad cultural heritage field, with a special value in increasing the visibility of authority data in the long tail of the Web. In many aspects VIAF is the embodiment of far earlier thinking and planning, long discussed at IFLA since 1978 (Plassard, 2003).

With respect to the types of entities covered by authority files, VIAF has a broad scope. Though it initially started with persons and corporate bodies, in the recent years its scope has continuously expanded. It currently covers, in addition, works, expressions, meetings and geographic names. Focusing on these entities is fundamental for the development of VIAF in the coming years, as this area of interest is directly related to the evolution of library catalogues and their FRBRization process. The loading by OCLC of data on works and expressions into VIAF, based on data mining from bibliographic records in WorldCat, underlines the upcoming role of VIAF in such a strategic domain (Hickey, 2014).
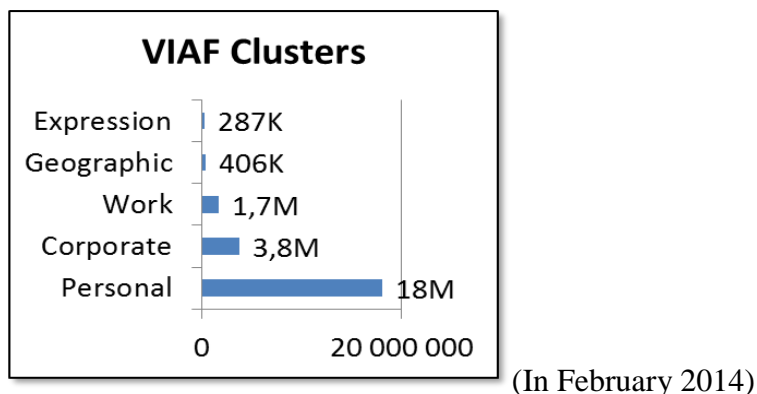

(In February 2014)

Figure 1: VIAF clusters – figures according to types of entities

ISNI - focus on persistent identification

As a separate initiative ISNI - ISO 27729 emerged as a response to the long felt need for a unique, global, cross-domain, standard, persistent identifier for *persons* and *organisations* involved in creative contents. The history of efforts to build such an identifier goes back to the 1970s (Angjeli, 2012). The advent of ISNI as a joint initiative from various information industry players[2] was prompted by recent developments in the global digital economy where standard, persistent, unique identifiers are viewed as the corner stone for supporting the highest level of automation. In addition to the concerns that constitute a common background for all those dealing with information on persons and organisations, as expressed in the

---

[1] In July VIAF includes the authority files from 36 individual major libraries, union catalogues and cultural and research institutions, as well as the whole NACO (Name Authority Cooperative) file, which in and of itself gathers hundreds of member libraries. In addition, 8 other sources are in test (including additional libraries, Wikipedia and research resources).

[2] These include libraries, rights management societies, stakeholders of the book supply chain, aggregators and service suppliers.

introduction section of this paper, other concerns necessitated an operational solution that would be both more focused and more encompassing. More focused in that it needed to address specifically the issue of a standard identifier. More encompassing in that the identifier needed to be global and cross-domain in order to be able to serve in international business interoperations between stakeholders in different domains.

Within the library community the IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) was charged in 1999 to study the issue of a unique number to identify the same person, corporate body (or any other entity likely to be described in an authority record)[3]. In 2008 the Working Group published "*A review of the feasibility of an International Standard Authority Data Number (ISADN)*" (FRANAR, 2008). While recognizing the advantages of a unique number as helping avoid duplication and being language and system independent, the Working Group recommended that IFLA should not pursue the idea of an ISADN itself. Instead it should continue to monitor the progress of the ISNI working group[4] and the VIAF Project and actively seek to influence the ISNI with a view to identifying common purposes with other communities.

In compliance with this recommendation libraries have always been a major driving force within the ISNI initiative, from the development of the ISO 27729 standard through the deployment of the ISNI system and network. The CENL (represented by the British Library and the Bibliothèque nationale de France) and OCLC are members of the ISNI International Agency (ISNI-IA), the ISNI governing body[5]. OCLC runs the ISNI central database[6] and is entrusted with the task of ISNI Assignment Agency. Authority control experts from the BnF and the BL are responsible for the quality of the database; they form the ISNI Quality Team.

To build a system for a cross-domain identifier the ISNI database ingests data from diversified data providers, including library and non-library sources. Their number amounts currently (June 2014) to over 70. In addition to the 41 VIAF data contributors, 30 other organizations have loaded data to ISNI.

Objectives
VIAF

VIAF is designed to be a reference source for the Libraries, Archives and Museums (LAM) worldwide. Its main objectives are the sharing of data; the reducing of cataloguing costs by taking benefit from work done by others; the provision of authority data in any form, language and script that users might need. End-users can reuse VIAF data freely thanks to an

---

[3] The terms of reference of the FRANAR group included two other tasks: to define functional requirements of authority records (a task that led to the development in 2009 of *Functional Requirements for Authority Data (FRAD): a conceptual model*) and to serve as the official IFLA liaison to other interested groups concerning authority files. http://archive.ifla.org/VII/d4/wg-franar.htm

[4] The ISO Working Group for the development of the ISNI standard started its work in 2006. By the time the FRANAR Review on ISADN was published the ISNI standard was under construction.

[5] ISNI-IA is a cross-domain public private partnership of 6 international organizations. The members of the ISNI-IA are: CENL, CISAC (International Confederation of Authors and Composers Societies), IFRRO (International Federation of Reproduction Rights Organisations, OCLC, ProQuest and SCAPR (Societies' Council for the Collective Management of Performers' Rights). ISNI-IA conducts the overall ISNI policy with respect to the development of the ISNI central system and network; it addresses and makes decisions related to data policy, and to technical, business, and strategic issues.

[6] The ISNI central database is accessible through the ISNI official website www.isni.org

Open data license ODC-By. Data are exposed in various output formats (RDF, MARCXML, JSON) that can be used within as well as outside the library community. For its members, VIAF has created an unprecedented opportunity to build a community of interest. Every VIAF member takes part in the VIAF Council. VIAF as a tool of this community gives a new visibility to the data of each of its members at international level. This is especially important for valuing data according to their provenance, by highlighting and enhancing the competence domains of each of the contributors, such as each of the national libraries in their own national domains.

Besides, VIAF allows its members to improve their own data with information related to areas less accurately covered by them, or not falling under their domain of competence. Each VIAF partner benefits from the expertise of the other partners. VIAF is also playing an important role in connecting data with other international projects, such as DBPedia and ISNI.

ISNI

ISNI's primary motivation is to provide an identifier for the interoperable management of information about and in relation to persons and organizations in the digital environment. Its use is critical not only to facilitate research and discovery of resources but also to streamline business transactions. In addition to being language and system independent, ISNI is also designed to be domain and geographical territory independent, in other words, to be a cross-domain and consequently bridging identifier. Actually, the rationale for designing ISNI as a cross-domain, bridge identifier combines several objectives:

- **A unique identifier across domains**. Persons and organizations may engage in a variety of diversified activities across domains, move from one territory to another, be creative in more than one language, be published in countries other than that of their residence or origin, etc. Only one ISNI identifier is assigned to each of the Public Identities (see below) of an individual or organization across all domains of activity, disregarding the roles it plays with respect to creative contents.

- **An 'all-purpose' identifier**. As an international ISO identifier ISNI is designed to address the needs of all information industry stakeholders worldwide, without distinction of activity sector, including LAMs, publishers and other players of the supply chain, rights management societies, musical and visual industry, research organisations, data centres and aggregators, and more.

- **A high level, global identifier *versus* local identifiers**. "ISNI is being established as an interoperable identifier: a core part of its function is to map other standard or proprietary identifiers […]". (LCC, 2014a) ISNI operates at a high interoperability level. It does not necessarily supersede identifiers that operate at local system, special network, or consortium level, whether open, public or proprietary. Such identifiers are necessary for the internal management of the information systems; workflows and business transactions may have been built on them. By ingesting local identifiers from a large variety of data contributors in the process of consolidating a given identity, ISNI links these identifiers to the neutral, high-level ISNI identifier that is assigned to the Public Identity (not to a record) as illustrated in figure 2. The centrally managed database provides the infrastructure that guarantees the consolidation, maintenance, update and diffusion of such links. This is how ISNI plays the role of a bridge identifier.
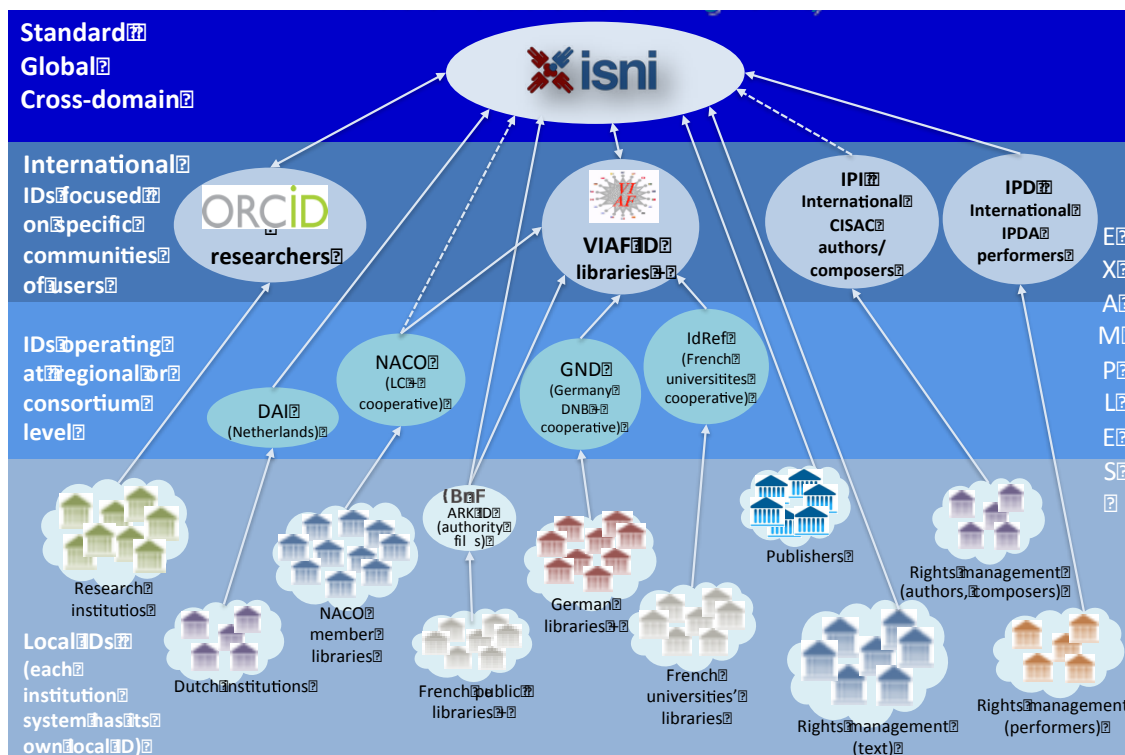
Figure 2: Layers of identifiers related to persons and organizations

The trajectory of ISNI is to position itself as the hub identifier across multiple domains, placing national library authority files centre stage but expanding beyond the limitations of their traditional scope.

Philosophy
VIAF

VIAF enables authority data from a variety of contributors to be brought into relation to one another. This relation creates the condition that data are then corrected and consolidated by each of the partners, enabling trustful associations. VIAF highlights common points and differences amongst data provided by its members. Consequently it also highlights data that need to be checked and fixed. For instance, the Martynas Mažvidas' birthdate in the cluster depicted below has to be defined:

Figure 3: VIAF cluster for Martynas Mažvidas', showing data provided with differences of birthdate

These are the fundamental operating principles and philosophy of VIAF. VIAF does not create any data itself. It only combines and brings together data provided by its members. It is important to point out that VIAF remains neutral towards differences in the cataloguing policy of its partners, and towards their working formats. VIAF deals with various data

6

models, data structure and cataloguing standards. It builds bridges, without imposing either rules or best practices to its partners. Amongst its members, national standards or RDA are used, and data are formatted either in Marc 21, Unimarc or even MADS. VIAF adopts the perspective of data itself; it merely indicates whether data provided by one partner is judged as being the same as (or close enough to) that of another, or not. In case of doubt, it suspends judgment, and creates a new cluster. In a certain sense, VIAF has a Pyrrhonian approach to the data.

These principles underpin VIAF's modus operandi. Specific algorithms are designed to process the data provided by partners, which constitutes the raw material for VIAF. These algorithms take into account the work done by the partners in their respective institutions. Automatic processing is the prevailing method VIAF uses for clustering data. Manual intervention is possible only exceptionally. VIAF itself does not provide any expert work on data quality but fully relies on the quality work performed by its partners. It is only by comparing and clustering high quality data that VIAF is able to create genuine added value in terms of data quality. If incoming data is sufficient and of high quality the result of clustering is trustworthy. Otherwise, if data is of poor quality, sparse, or undifferentiated, the clustering process is compromised, and the risk of creating other parallel clusters for the same entity is high, which in itself is an undesired consequence.

Once they are created, what happens to VIAF clusters with respect to data traffic due to updates and newcomers? Are VIAF clusters stable? If no new member data enters into a VIAF cluster or if none of the source authority records that compose a cluster is updated[7] the cluster remains stable. But if data from a new member is added to a given cluster, or if an updated version of an already existing record (corrected or completed with new information) from a given partner is reloaded, the previous balance of the cluster is likely to change. The algorithmic recalculation of the new balance might show that the previous balance was false. This means that the whole data present in the cluster is re-clustered taking into account the new feedback. The re-processing is either fully justified, leading to the improvement to the quality of the cluster; or it might produce some undesirable side effects, such as the unjustified movement of a given record from one cluster to another.

Thus the issue of VIAF clusters' stability cannot be considered en bloc. The question is: what is stable, and what for?

As far as the VIAF ID is concerned, it is assigned to a given VIAF cluster of data, even if this cluster contains data from a single partner. The VIAF ID doesn't refer to the entity (person, corporate body, work…) itself, but to a given cluster of data. However, although identifying a cluster of data managed by VIAF as a system, the VIAF ID does have an international value, because VIAF's scope is international.

ISNI

ISNI's fundamental concepts are largely inspired by library authority control principles, theory, and practice.

The ISNI core concept is that of "Public Identity" defined by the ISO 27729 standard as "[…] the identities used publicly by parties involved throughout the media content industries in the creation, production, management and content distribution chains." Though the term doesn't come from the library terminology, the concept itself designates something close to what corresponds to the object of a library authority record for a *person*, or a *corporate body*.

---

[7] VIAF receives regular updates of its partner data comprising new creations, deletions and modifications (including change of data within a record and merges and splits of records)

Actually it was FRAD that brought light to the library theory by explicitly specifying that *persons* and *corporate bodies* within the library catalogues environment are "bibliographic entities" or "bibliographic identities". "As such, they reflect intellectual constructs or concepts that are integral to the rules used to create library catalogues […]". (IFLA FRANAR 2013) Depending on the cataloguing rules applied within a given catalogue, an authority record for a *person* or a *corporate body* may or may not identify a real life person or corporate body. "[A]uthors may be viewed in certain circumstances as establishing more than one bibliographic identity, and in that case a specific instance of the bibliographic entity *person* may correspond to a persona adopted by an individual rather than to the individual *per se*". (IFLA FRANAR 2013) A typical case is that of pseudonyms. It is also to be noted that RDA recommends establishing separate public identities for each pseudonym of the same person[8].

The ISNI standard however adopts a strict perspective and stresses the distinction between the real life individual or corporate body and its Public Identity or Identities. Only one separate ISNI identifier is assigned to one Public Identity. Associated metadata, a significant part of it coming from VIAF, supports the disambiguating process of the Public Identities in the central assignment system (main names / variant names / dates / country / publisher / titles of associated works, affiliations, etc.) Links are then established between the Public Identities of one real life entity if and only if the relation between such Public Identities is made public. Otherwise, for confidentiality, no links are made.

Figure 4: John Le Carré and David Cornwell are two public identities of the same individual. A separate ISNI is assigned to each.

The ISNI system is built directly on libraries' legacy. Libraries' data are at the heart of the system, with data from VIAF used as source file when the central ISNI database was initiated in 2011. They were matched against other non-VIAF contributors[9]. VIAF data refreshment is part of the ongoing processes in ISNI. Consequently a good portion of identities from VIAF contributors' authority files gets ISNIs identifiers.

---

[8] RDA 9.2.2.8 ("Individuals with More Than One Identity")

[9] These include data on books in print, on theses, on journal articles, on tables of contents; data from rights management societies in the domain of text, music, and performance rights; data from research and professional societies, and from the encyclopaedic sources on the web. For an updated list of ISNI data contributors see http://www.isni.org/content/data-contributors (Accessed 13 July 2014)
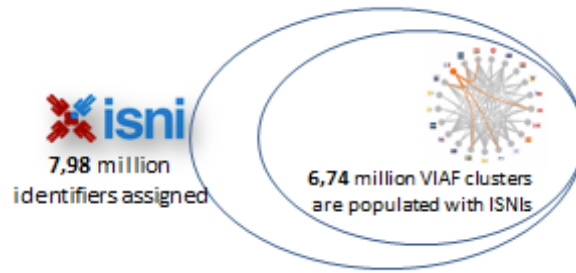
Figure 5: Nearly 85% of the assigned ISNIs contain VIAF as source (July 2014)

Differences between VIAF and ISNI

Although ISNI and VIAF have much in common, differences in philosophy and ways each system is deployed are to be found in the fundamental motivation of each. We will only briefly outline here the main criteria that constitute the fundamental difference between VIAF and ISNI (MacEwan, Gatenby & Angjeli, 2013; ISNI-IA 2013):

- First and foremost ISNI is endowed with its own policymaking body and permanent control infrastructure, independent of the particular practices or cultural background of its data contributors. ISNI's data model policy, and system development plan are discussed at the ISNI-IA level. Control relies both on computational methods and on human expertise. Algorithms, automatic data anomaly detection and correction, data analysis views and reports are developed and refined by OCLC's ISNI team in Leiden, in the role of the ISNI Assignment Agency. Authority experts from the BnF and the BL, from the ISNI Quality Team, that assesses database quality on an ongoing basis. Both the OCLC ISNI team and the BnF-BL Quality Team work hand-in-hand on a day-to day basis

    VIAF, on the other hand, remains very close to its partners. The purpose is to remain open to various standards, practices or cataloguing traditions and make of VIAF a hub for authority files from partners. Thanks to this approach, VIAF highlights the common core of data coming from its partners without excluding any modelling and standardizing approach. The VIAF Council, with representatives from every VIAF contributor, advises OCLC for improving VIAF; it embodies a unique community of discussion on authority data in the world.

- While VIAF remains neutral towards differences in the cataloguing policy of its data contributors[10], ISNI does not; it has a strict policy that ensues from its main objective: consolidate each Public Identity and assign to each separately a unique, reliable ISNI identifier that can then be broadly diffused and referred to with confidence. Although ISNI started using the same matching and merging algorithms as VIAF, very rapidly these algorithms were adapted, and have been iteratively refined and tightened over the past four years to better fit ISNI's purposes. [11]

---

[10] For example, some libraries establish separate authority records for each of the separate identities of the same individual; others establish only one authority record that brings together all the identities of the same real life person, with one of the names declared as preferred authorized form and the other names for the other identities as see references.

[11] Whenever necessary, ISNI splits and merges of data coming from VIAF, and even applies protection to data that has been fixed manually.

Metaphorically, VIAF could be viewed as a common recreation ground for various and different data and ISNI as its supervisor. There is no recreation ground without supervisor and no supervisor without recreation ground.

- For these reasons, the relationship to the source data is different in ISNI and in VIAF.

ISNI maintains Public Identity records and provides online edit facilities, used by the Quality Team and by ISNI members via the member interface. Any ISNI member can ensure their data is correctly clustered with data from others. This implies that the ISNI database can be autonomous *vis-à-vis* its source data in so far as the decision on consolidating a Public Identity is concerned. For the purpose of identification ISNI can decide to split or merge incoming data from a single source, and an ISNI identifier is assigned only if there is enough confidence in incoming data and in matching (more in the section on UBC). "In essence it [ISNI] takes the VIAF concept of an international authority file, but it provides a toolbox for working with it, provides access to wider data sources and gives the possibility of ongoing batch processes for original authority work." (MacEwan, 2013) In other words the ISNI database has the capacity to evolve into an online, shared authority file similar to other networked library authority files like LC/NACO but more global in reach and scope.

VIAF is a *virtual* file. It means that it does not step in its source data itself, but plays a role only at the VIAF cluster level. It can merge or split VIAF clusters but considers the authority records coming from its partners as indivisible packages. Individual records from a partner can join a VIAF cluster, they can move from one cluster to another if there is evidence that such should be done. From this perspective VIAF remains close to the authority work done by its partners and respects this work.

UBC, VIAF and ISNI: from share to control, from control to share

In the new data environment where libraries are facing the challenges of the Semantic Web and of Open Data paradigm, the noble aspiration of the Universal Bibliographic Control is not a paradise lost; on the contrary, it only needs to be reinvested and further developed to adapt to the new context.

The new professional IFLA statement on UBC summarizes the issue as follows:

"In the 1990's it was recognized that having data in a language and script users can understand is extremely important, so respecting the cultural diversity of users around the world should be addressed as well. These are not necessarily mutually exclusive conditions, as has been demonstrated in the subsequent Virtual International Authority File (VIAF) initiative. Web technologies offer new possibilities of sharing data at a global scale and beyond the library domain, but also show a need for authoritative and trusted data. Thus, the concept of UBC is still valid and worth IFLA's continued support and promotion". (IFLA, 2012)

The VIAF and ISNI use cases demonstrate that their achievements and the future vision they are building rely on and strongly promote the UBC principles by applying them in innovative ways in the evolving technological environment.

VIAF, libraries and UBC

The UBC basic principle is to share. Each national player, each national bibliographic agency has a special responsibility for producing authority data related to its specific national field of competence and makes them available for reuse by the other players at international level, even beyond the library community. The IFLA statement on UBC further states:

"NBAs (National Bibliographic Agencies], as a part of the creation of authoritative bibliographic data, also have the responsibility for documenting authorized access points for persons, families, corporate bodies, names of places, and authoritative citations for works related to its own country and for making that authority data available to other NBAs, libraries, and other communities (for instance archives and museums)". (IFLA 2012)

VIAF is the direct offspring of the UBC sharing principle. Each NBA provides its data, immediately recognizable in the VIAF public display of a given cluster by the national flag. VIAF juxtaposes diversified national fields and, therefore, can be considered as a hub for the UBC on authority data, where every NBA member is openly given the possibility to play its authoritative role. With every new coming partner new opportunities are created to highlight specific areas of competence and responsibility, not only for NBAs but also for the other data providers, including research institutions, museums, archives, and more.

Partners provide to VIAF not only data relevant to their national scope but their data set as a whole. Typically, for the purpose of managing access to their collections the coverage of authority files produced by each institution often extends beyond its specific area of competence. This enables VIAF to be not only a basic, mere juxtaposition of different specialised NBA datasets, each in its own authoritative domain, but also a ground for broad data confrontation, comparison and, furthermore, for data complementarity on a "democratic" basis. The principles of UBC are thus applied from a broader perspective; recognized areas of competence are not only juxtaposed in a static way but dynamically compared, and cultural diversity is promoted. VIAF has the advantage of revealing the originality of data provided by each partner. In this respect, while data provided by a partner, recognized as the "authoritative source" in a specific national or other domain, are most likely to constitute the core of a cluster, other data from other partners (another national library, a countrywide union catalogue or a smaller specialized institution) enrich the same data cluster. This is illustrated by the example of Honoré de Balzac.
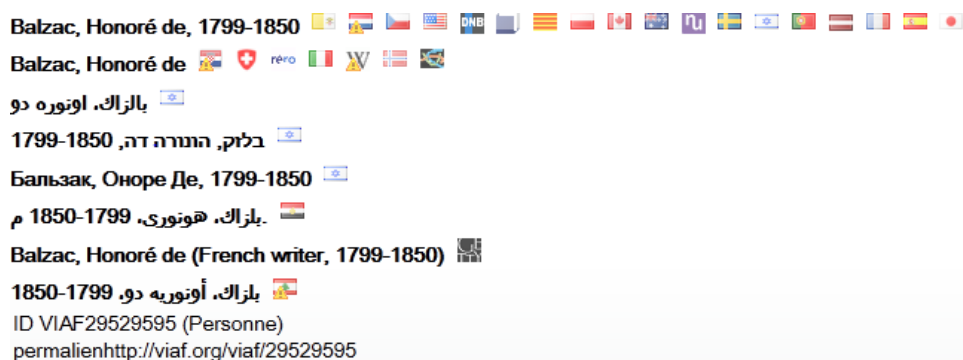
Balzac, Honoré de, 1799-1850
Balzac, Honoré de
بالزاك، اونوره دو
בלזק, הונורה דה, 1799-1850
Бальзак, Оноре Де, 1799-1850
بالزاك، هونوري، 1799-1850 م
Balzac, Honoré de (French writer, 1799-1850)
بالزاك، أونوريه دو، 1799-1850
ID VIAF29529595 (Personne)
permalienhttp://viaf.org/viaf/29529595

Figure 6: "Honoré de Balzac" cluster in VIAF

The French NBA (the Bibliothèque nationale de France and the SUDOC) provides the "core data" for the French author, such as the official, national form of the name, the correct spelling and structure of the authorized form according to national rules, as recommended by the IFLA document "*Names of persons: national usages for entry in catalogues*"[12]. The other VIAF partners and VIAF users are encouraged to use this national form. However, each library and data provider acts in a specific cataloguing environment and has competence in areas that might complement the collective knowledge. For instance, libraries that catalogue in a particular script, conformant to their linguistic environment, e.g.: Hebrew form for the

---

[12] <http://www.ifla.org/files/assets/cataloguing/pubs/names-of-persons_1996.pdf>

Israeli National Library, or Arabic form for the Lebanese National Library and from the Bibliotheca Alexandrina, are likely to supplement the cluster with these script forms of the name. It is evident that the Bibliothèque nationale de France is not likely to provide such script forms for all its national identities. Moreover, as far as the Hebrew and Arabic scripts and languages is concerned, forms provided by these libraries will be recognized as authoritative, as fall within their area of competence. Other partners can also contribute additional elements to the cluster, corresponding to their own rules, such as an additional qualifier ("French writer" for the Getty Institute). In this way, data provided by the NBA that is considered the "authoritative source" for a given identity is consolidated and enriched by other VIAF partners. VIAF enables thus a model where core data from a NBA that is considered as "authoritative source" for a give type of identities becomes a core around which gravitate other partners' data.

Another facet of VIAF is that it reveals differences between partners' data. Either differences are considered as minor and do not prevent data to be clustered together, or they are considered as significant thus avoiding merging into a same cluster. For instance, the German legal expert Moritz August von Bethmann-Hollweg (1795-1877) falls within the scope of the Deutsche Nationalbibliothek, which provides the national and official form of the name, along with the biographical dates.

Figure 7: VIAF cluster for Moritz August von Bethmann-Hollweg (1795-1877) with the DNB form of the name on top of the cluster

Two library sources, Bibliothèque nationale de France and SUDOC have not been included in the above displayed Moritz August von Bethmann-Hollweg's cluster because of an error on birth date[13] which has caused the creation of another, separate, duplicate VIAF cluster for the same identity (see figure 8).

Figure 8: Another, duplicate VIAF cluster for the same identity, due to a difference on birth date

This example illustrates how VIAF allows other members' data to be compared with data provided by the reference institution from the UBC point of view, which in this case is the Deutsche Nationalbibliothek that is the authoritative source for the German people, organizations or works.

Besides, VIAF is a living and dynamic file as it periodically ingests updates and newly created data from the source institutions. It thus becomes a tool for interoperation and dialogue between all VIAF members for improving data in an ongoing process. It also is

---

[13] The SUDOC (French universities documentation system) uses the BnF authority records as base file for creating its own authority data.

showing that the UBC principles are not limited to only providing a reference form of name and reference data. The dynamics of VIAF is an incentive for promoting dialogue among institutions within the library community and beyond. Libraries and culture heritage institutions are strongly encouraged to join VIAF so as to share their data with other contributors, leverage their specific assets in an international environment and get even more value from the work provided by others.

ISNI – further expanding the principles of UBC

While ISNI builds on the UBC principles described for VIAF, we will discuss here how ISNI is expanding the components "Universal" and "Control" inherent to UBC.

As described above, ISNI had to respond to different challenging objectives to VIAF. It needed to provide persistent solutions so that the identifier can be securely and consistently assigned to Public Identities.

ISNI has developed rules and policy in order to ensure the quality levels and accuracy required in fixing identities, wich can be basically resumed as follows:

− An ISNI is assigned only where there is sufficient level of confidence: 1) the data itself is good enough to merit assignment (e.g.: sparse records, and undifferentiated records are not allowed), and 2) the matching score of data is high. "Once assigned the cluster of data sources remains fixed and other data matching to that cluster will either join the cluster or flag as a potential problem if it finds more than one cluster to match." (MacEwan, 2013)

− Data provenance in ISNI is very heterogeneous with library and non-library data being loaded. An ISNI is assigned only when three sources of data match: 1) a VIAF cluster alone is eligible for ISNI assignment only if it combines data from at least three sources; 2) a VIAF cluster with fewer than three sources is likely to get an ISNI only if it matches with one or two other non-VIAF sources;

− In addition to these criteria, ISNI has developed policy and mechanisms for "Unique name assignment" and "Single source assignment". The "Unique name" process ensures assignment of ISNIs to Identities that are unique in the entire database, i.e. do not match to any other source but are deemed confident enough. The "Single source" process is used more cautiously to allow data files from certain authoritative sources to be given assigned status as a base file within the ISNI system on the grounds that the source is a strategic leader for the diffusion of the identifier. In these cases editorial effort is expended to minimise duplications that may arise.

− ISNI has also extended its matching techniques to allow the system to build clusters iteratively based on cross matching identities contributed within records at the bibliographic level. This enables assignment of ISNIs to authors of journal articles, and other contributors to contents, which fall outside the scope of traditional authority files, but featuring in other databases either of libraries [14] or of other non-library data providers. Matching remains based on agreement between at least two separate sources.

---

[14] Over 70,000 ISNIs have been assigned in the ISNI database to authors of theses in the EThOS database (Electronic Theses Online Service) <http://ethos.bl.uk/>.

230,000 ISNIs have been assigned to authors of journal articles from the BL's Electronic Table of Contents <http://zetoc.mimas.ac.uk/>.

Figures 7 and 8 above illustrate that both VIAF clusters related to the German legal expert Moritz August von Bethmann-Hollweg (1795-1877) could not be merged in the ISNI database due to difference on birth date, though obviously the same identity is represented in both. They also illustrate that whilst the cluster depicted in figure 8, with an error on birth date, continues to exist in the source data (and therefore in VIAF), it has been rejected by the ISNI matching criteria as data was deemed as not being confident enough. Consequently, no ISNI identifier has been assigned to the second cluster. The uniqueness of the identifier for the same identity is thus ensured.

Currently (June 2014) 7.98 million ISNI identifiers are assigned as they respond to the confidence criteria. The database contains though more than 9.6 million additional records that still do not match the assignment criteria.
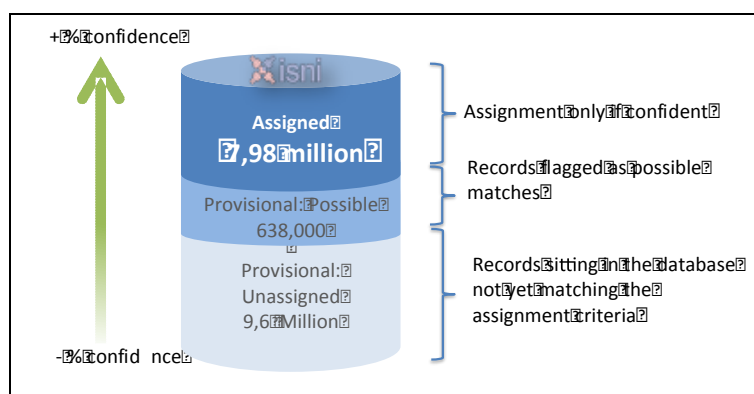

Figure 9: ISNI database – figures – June 2014

ISNI needed to provide persistant solutions so that the identifier can be securely and consistently maintained and diffused thoughout the network of data contributors, as well as openly on the web. For this a systematic notification process is put in place between the ISNI central system and the data contributors. Monthly notifications on assignments, merges, splits and deletions are sent to the contributors. In this way the quality work performed by the ISNI Quality Team directly benefits the data contributors. One of the lessons learned over the past four years of work with the ISNI database and system is that it is impossible to envisage a global economy of sharing without the active participation of the data providers in the quality work . The purpose of notifications is thus not only to diffuse ISNIs and inform contributors about eventual changes regarding their data, but also to request correction of data at the source institution – where data is produced. ISNI Registration Agencies and Members[15], as privileged data contributors, have also access to a cataloguing interface to actively engage with the enrichment and correction of records to ensure maximum assignment. The whole economy of data production and diffusion is thus being transformed with "Control" working both ways. For libraries joining ISNI (directly or through VIAF) means that their traditional authority control efforts are placed in a new context and are gaining more strategic leverage, and opening possibilities to build new partnerships.

Working with a large variety of data providers is also proving each and every day the value of the UBC high level principle according to which work on consolidating identities should be build by sharing responsibility of work according to areas of competence of each of the data providers (e.g. NBA to be entrusted with the quality of identities falling within their respective national area of coverage; domain specific contributors to be entrusted with the

---

[15] For more information on ISNI Registration and ISNI Members see:
<http://www.isni.org/content/isni-community>

quality of the identities of their specific domain, etc.) As a precursor of this approach, from the beginning, ISNI allocates weightings to the data providers according to the provenance of their data, and based on the degree of their likeliness to be in direct contact with the subject of identification. The highest level of confidence is given to data providers that get information directly from the person or organization behind the identity[16]. When, for a given identity, data is provided by libraries, three levels of confidence are defined in increasing order:

1. library authority record agreed by one or two libraries,

2. library authority record agreed by three or more national libraries,

3. library authority record agreed by three or more national library AND country of the providing library for one of the authority records equals nationality of the identity.

This is another demonstration of ISNI's efforts to rely on and carry out the principles of UBC. But to make this fully functional, work should be organized with the active participation of libraries themselves by sharing responsibilities and establishing workflows.

It is worth mentioning that the ISNI control infrastructure described throughout this article, is complemented by a feedback mechanism from the public. The Quality Team monitors end user input, by applying the principle of "monitored crowdsourcing". Feedback is analysed and processed by the Quality Team, and end users are notified about the outcomes of their input. The remarkable feature is that this mechanism is proving to attract either the individuals themselves that have an ISNI record in the database or other specialists of specific domains to help enhancing the quality of the ISNI database.

The representatives of all domains in the ISNI system (Board members and other contributors) unanimously recognize libraries' expertise in identity management. The ISNI Quality Team's reputation is already solid, but team reinforcement is needed. Here again the principles of UBC open new perspectives. From share to control, from control to share, other libraries are invited to share the experience of the ISNI Quality Team and contribute, according to their areas of expertise, in order that libraries continue to play a leading role in identity management worldwide.

VIAF and ISNI interoperation
Interoperation between ISNI and VIAF is of strategic importance for placing library data at the heart of an unprecedented, solid network of information on identities, and for connecting them to other non-library datasets.

Though pursuing different objectives, as already described, both systems have a strong relation from the beginning. But this relation needs to be further developed to make ISNI and VIAF fully interoperable – a requirement for the reliable diffusion of ISNIs through VIAF. Actually, diffusion of ISNIs into library authority files is already implicit in the assignment of ISNIs to VIAF clusters[17]. ISNI appears on top of VIAF clusters as actionable identifier, and is openly diffused through the VIAF RDF dumps. Conversely, diffusion of ISNIs through the

---

[16] Such is the case of rights management societies who are often in direct contact with the parties themselves for the purpose of rights transactions and payments of royalties.

[17] The Bibliothèque nationale de France has already loaded over 1 million ISNIs in its authority file for persons and organisations, made visible on the authority records web as actionable links. The ISNIs loaded are then diffused through the Linked Data service data.bnf.fr, and embedded in the bibliographic products provided regularly to all BnF patrons.

links and URLs embedded in the ISNI database is the next step required to realise the purpose and potential of ISNI as a bridge identifier and as a building block for Linked Data.

Currently regular workflows are already in place between ISNI and VIAF. ISNI uploads VIAF updates every two months, and sends the ISNI identifiers to VIAF every month. In addition, as ISNI continuously refines the matching algorithms to ensure the uniqueness of the identifier, and curates data by performing manual quality control (work done by the Quality Team), it detects anomalies in VIAF clusters and sends regularly notifications to VIAF. Two main categories of errors are reported:

 – VIAF clustering anomalies (including cluster movements, as discussed above);

 – Errors in the source data of VIAF partners.

While clustering anomalies can be handled by VIAF itself, reporting errors found in source data of VIAF partners raise problems related to the efficiency of the notification workflows. At this point, involvement of VIAF partners themselves in the process is needed.

All aspects of VIAF-ISNI interoperation are regularly discussed between the administrative and technical infrastructures of both systems. Within the VIAF Council an ISNI-VIAF Interoperability Taskforce has been created.[18] Issues being addressed include: optimising the dataflow between the two systems; incorporating ISNI notifications into VIAF; engaging participating libraries in the process of data correction; and acknowledging the contributors of high quality data. An outstanding outcome of these joint efforts is the agreement that ISNI sends error notifications directly to VIAF partners whenever errors in source data are detected and processed in the ISNI database by the Quality Team. The process has just started and the first results are encouraging. Libraries are invited to be full players in the process, by correcting their data and closely collaborating with the ISNI Quality Team for the benefit of the entire library community and beyond.

The summary below shows the complementary scope, role and system functionalities of VIAF and ISNI – demonstrating benefits for participating in both.

VIAF scope	ISNI scope
• Persons	• Persons
• Organisations	• Organisations
• Works / uniform titles	
• Expressions	(Excluding sparse)
• Meetings	(Excluding undifferentiated)
• Geographic	

---

18 Discussions were first engaged at the VIAF Council meeting, held at IFLA 2012 Conference in Helsinki, based on a detailed discussion paper "ISNI and VIAF interoperation", presented jointly by OCLC Leiden, in the quality of ISNI Assignment Agency, and by the ISNI Quality Team (BL and BnF). Ever since discussions have been followed up as part of an ongoing process and included in the agenda of all VIAF partners' meetings. In May 2013, the VIAF Council set up the VIAF – ISNI Interoperation Taskforce with the participation of the ISNI-IA representatives. A meeting organized in April 2014 in Paris, with representation from ISNI-IA, VIAF-OCLC, BnF, BL, DNB, BNE and ABES, came up with significant strategic and technical recommendations.

VIAF data nature	**ISNI data nature**
- All public data	- Includes private data

VIAF role	**ISNI role**
- Creates clusters	- Creates an ISO standard identifier
- Ingests authority records from the worlds' major national and research libraries and from other cultural heritage institutions	- Ingests relevant data from VIAF and generates links with non VIAF sources – a platform for cross-domain links
	- Works with non-library data and provides match results (e.g. article data from journals, theses data, etc.)
- Exposes and diffuses	- Only exposes and diffuses the assigned ISNIs (where there is confidence that records are differentiated and de-duplicated)

VIAF system	**ISNI system**
- Harvester	- Batch load
- Clustering mechanism	- Clustering mechanism
	o Clusters are maintained
	- Online request API
- Website (interface in 5 languages)	- Website (interface in English only)
	o More than 16 indexes
	- Online facilities:
	o End user input
	o Edit online facilities for Members and Registration Agencies
	- Quality Team monitors the quality of the database and corrects records
- Notifications to be sent to data providers (under instruction)	- Notifications are sent to data providers
- Display and download in multiple formats	- Display of a record in different formats (XML, RDF, from a content negotiation page)
- Linked Data and SRU	- SRU; soon Linked Data (pending licence issues under discussion)
(Open data (ODC-By licence)	

Conclusion: the UBC principles in a new era

As demonstrated throughout this article, the visions behind both VIAF and ISNI are inspired by the philosophy of the UBC, which itself is inspired by the humanistic ideal of sharing collective knowledge and promoting cultural diversity, and at the same time facilitating efficiencies in the ongoing work of the world's libraries.

For many decades the state of the technology limited the realization of the full potential of the UBC principles. VIAF and ISNI, taking advantage of new and emergent technologies, have reaffirmed these principles with innovative, transformative applications, applications that were unimaginable in the not too distant past. Federating the authority files of libraries from around the world, VIAF is recognized as a brilliant achievement. ISNI further consolidates

identities; and assigns an ISO high-level, cross-domain identifier to the Public Identities of people and organisations; and thereby is becoming a powerful tool to build an efficient Linked Data value chain. By furthering technical and organizational innovation, both systems working together are playing a transformative role in the whole economy of data production, control, diffusion, and reuse. In this respect, ISNI and VIAF are both reviving the UBC principles and opening a new era of collaboratively consolidating identities at an unprecedented, universal scale.

Furthermore, the experience of working in an open, global environment of sharing data and knowledge arguably extends the scope of the UBC principles, as the new technologies present opportunities unforeseen by its creators. At the same time the opportunities present new challenges.

The first crucial challenge is the quality of source data. It is obvious that the higher the quality of source data, the greater are the benefits that VIAF and ISNI can create when aggregating and interrelating them. This raises the issue of how to encourage libraries to devote resources to the quality of the data they share in return for the benefits received from these new resources. An immediate response to this is that the libraries need to actively participate in both systems to see their production placed in a broad environment and fully realize their responsibility in a global economy.

With respect to the responsibilities in the production of authority data, many questions arise. Should the authority of NBAs remain unquestionable? Should they be the only institutions trusted in this area? Or should the value of other specialized institutions be recognized, each in their fields of competence? By gathering data from a large variety of contributors, VIAF and ISNI are demonstrating that new, trustworthy partners in the production of authority data are emerging, partners for whom an exceptional quality of data is essential.

Related to this is the issue of how should the authority work and control at local level be organised? Should the responsibility of authority data be concentrated only in the hands of the NBAs? Local libraries and institutions have special knowledge on identities related to the special local collections they manage; other specialised libraries have experts with good knowledge on identities of their domain. This special knowledge, when shared, can benefit the whole economy of authority data production and control, building together a body of high quality data that none of us could do alone. The role of the NBAs would be precisely to mobilize competences, develop partnerships, share work at national level as a contribution to a shared global network.

References
Angjeli A. (2012) 'ISNI : un identifiant passerelle', *Documentation et Bibliotheques,* Vol 58 No 3, pp 101-108 [Digitized copy available online from the BnF Open Archives Portal 'Hyper Articles en Ligne'] http://hal-bnf.archives-ouvertes.fr/hal-00803772 (Accessed 5 June 2014)

Boulet, V. (2012) 'L'infrastructure de l'information et les besoins des utilisateurs : tout le pouvoir aux données structurées!'. in *World library and information Congres: 78th IFLA general Conference and Assembly*, August 2012, Helsinki, Finland [online] http://conference.ifla.org/sites/default/files/files/papers/wlic2012/80-boulet-fr.pdf (Accessed 13 July 2014)

Bourdon, F. and Boulet, V. (2013) 'VIAF: A hub for a multilingual access to varied collections'. in *World library and information Congres: 78th IFLA general Conference and Assembly*, August 2013, Singapore [online] http://conference.ifla.org/past-wlic/2011/79-bourdon-en.pdf (Accessed 13 July 2014)

IFLA Working Group on FRANAR (2013) *Functional Requirements for Authority Data: a conceptual model (FRAD)* [online] IFLA Working Group on Functional Requirements and Numbering of Authority Records, First edition, 2009 as amended and corrected through July 2013 http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf (Accessed 13 July 2014)

IFLA Working Group on FRANAR (2008). *A Review of the Feasibility of an International Standard Data Authority Number (ISADN)* [online] IFLA Working Group on Functional Requirements and Numbering of Authority Records. Available from the page http://archive.ifla.org/VII/d4/wg-franar.htm Direct link http://archive.ifla.org/VII/d4/franar-numbering-paper.pdf (Accessed 13 July 2014)

IFLA (2013). *IFLA Professional Statement on Universal Bibliographic Control*. Initiated by the Bibliography Section, endorsed by the Cataloging Section and the Classification and Indexing Section and finally by the IFLA Professional Committee. http://www.ifla.org/publications/ifla-professional-statement-on-ubc (Accessed 13 July 2014)

Hickey, Th. (2014). 'FRBRizing WorldCat' *Ongoing* [online] 9 April http://outgoing.typepad.com/outgoing/2014/04/frbrizing-worldcat.html (Accessed 13 July 2014)

International Organisation for Standardization (2012) *ISO 27729: 2012: Information and documentation - International standard name identifier (ISNI)*. Geneva, ISO http://www.iso.org/iso/catalogue_detail?csnumber=44292 (Accessed 13 July 2014)

ISNI-IA (2013) '*ISNI for libraries*' [Brochure, documentation provided on the ISNI official website] http://www.isni.org/filedepot_download/126/346 (Accessed 13 July 2014)

LCC (2014a) '*Principles of identification*' Version 1.1 [online] April 2014. http://media.wix.com/ugd/bff7bc_20e274d80da1462ebee53df9f69ab3db.pdf (Accessed 5 June 2014)

LCC (2014b) '*The LCC Manifesto*' Version 1.0 [online] April 2014 http://media.wix.com/ugd/bff7bc_a39299633ee74d1e82686c79772c95a6.pdf (Accessed 5 June 2014)

MacEwan A., Angjeli A. & Gatenby J. (2013) 'The International Standard Name Identifier (ISNI): The Evolving Future of Name Authority Control', *Cataloging and Classification Quarterly*, Vol. 51, No. 1/3, 2013. [abstract online] http://www.catalogingandclassificationquarterly.com/ccq51nr1-3.html (Accessed 5 June 2014) http://dx.doi.org/10.1080/01639374.2012.730601

MacEwan, A. (2013) '*ISNI, VIAF and NACO and their relationship to ORCID*' [online] Discussion paper for PCC Policy Committee, 4 November. http://www.loc.gov/aba/pcc/documents/ISNI%20PoCo%20discussion%20paper%202013.docx (Accessed 13 July 2014)

Plassard, M.-F. (2003) 'IFLA and Authority Control'. in *Authority control: definition and international experiences: international conference*, 10-12 February 2003. Florence, Italy 2003 http://www.sba.unifi.it/ac/relazioni/plassard_eng.pdf (Accessed 13 July 2014)