Draft Recommendation for BHL Creator Name Authority Management

August 23, 2016
Updated January 6, 2019

The current BHL metadata database supports creator name authority control using the field <AuthorIdentifier> with the subelement of <IdentifierValue> that ultimately supports consistent use of names in metadata that would also support metadata quality and user experience. Since the BHL already has a good supporting data structure, we would like to recommend using the Virtual International Authority File (VIAF, viaf.org) URI as a value for the element <IdentifierValue>.

Why VIAF.org?

Aggregates 40+ name authority files

Initiated by the German National Library and the Library of Congress, and maintained by OCLC, the VIAF provides aggregated name authority files from 40 national libraries and institutions in web actionable URI format. Since the VIAF includes the name authority files of institutions from different countries in different languages, it supplements name entries, especially foreign names, that are not available in the LC Name Authority File. [1]

Provides various name forms and information about the name

Each name entry available in the VIAF includes preferred forms, alternative name forms, and related names when it is available, helping catalogers control different name forms for the same name. The VIAF also includes information about the name, such as gender, nationality, major works, and associated names that could be used to improve the user experience as well as maintain the name entries in the BHL (see figure 1).

Supports actionable URI

Additionally, each name entry is represented as an actionable URI that helps catalogers or system managers manually add or edit information about the name. All information about the name can be updated and maintained by institutions that provide the name authority file, and the changes will be affected simultaneously and directly to any resources (metadata).

Implementation Plan

In Metadata

Many libraries, including the Library of Congress' Program for Cooperative Cataloging (PCC), have a 'Task Group on URIs in MARC'[2] to try to find a better way to practice name authority control by using URIs as values instead of strings. Since MARC, the current library metadata format, does not currently have a specific data field/subelement that allows catalogers to add

[1] http://www.oclc.org/research/activities/viaf.html
[2] https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html

URIs in addition to string values, the PCC's task group experimented using the subfield 0 (Authority record control number or standard number)[3] of any MARC datafields for names (100, 110, 111) as a placeholder for the URIs. For example, if the name has a matching URI available in the VIAF, the VIAF URI can be added in the subfield 0 of the same datafield. This MARC record can be also transformed to MODS with the same information as shown in figure 2 below.

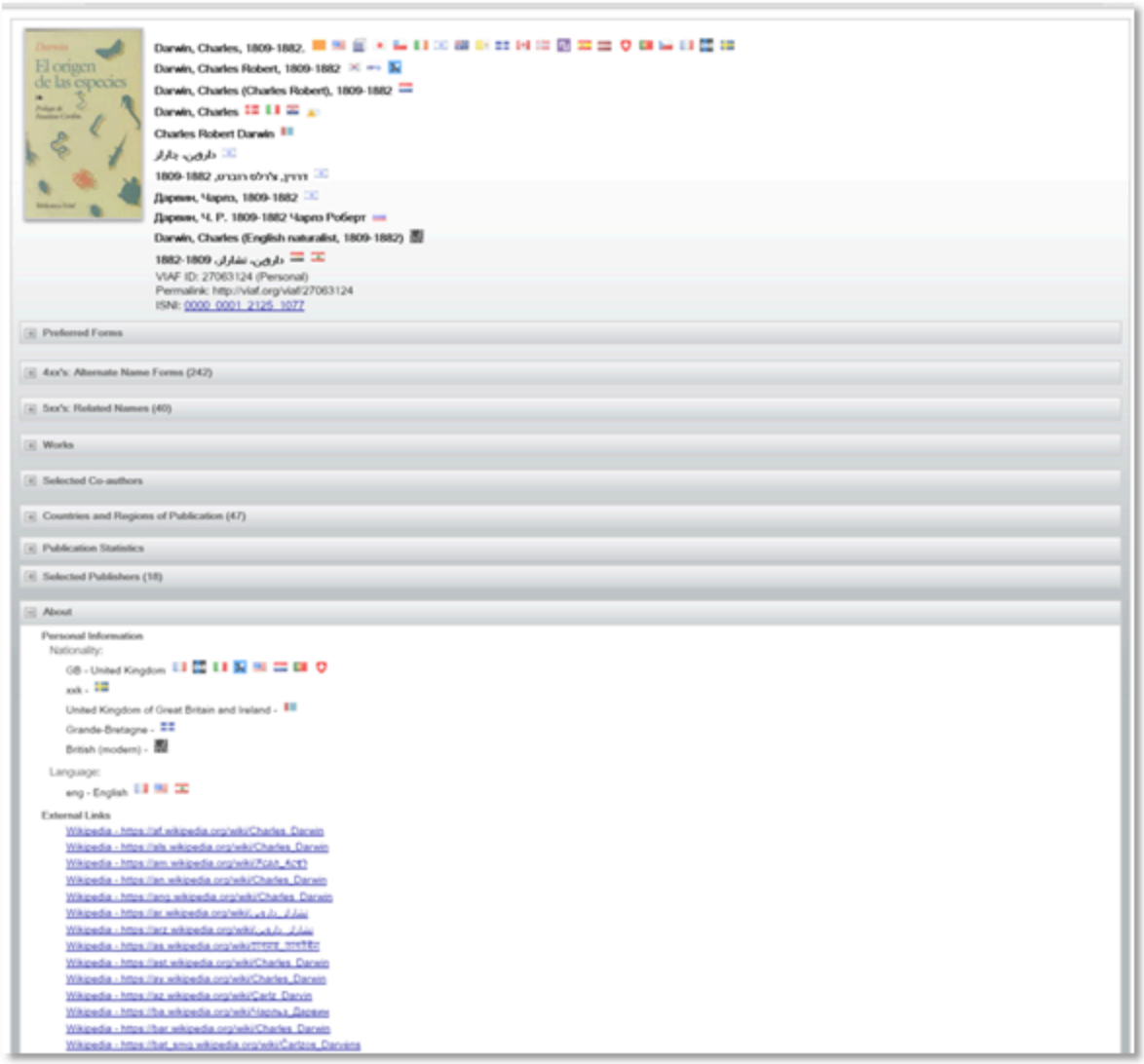

Figure 1: An example of the VIAF entry for Darwin, Charles, 1809-1882 (http://viaf.org/viaf/27063124)

MARC	MODS
<datafield tag="100" ind1="1" ind2=" ">	<name type="personal"

[3] https://www.loc.gov/marc/bibliographic/bd100.html

<subfield code="a">Darwin, Charles,</subfield> <subfield code="d">1809-1882.</subfield> **<subfield code="0">http://viaf.org/viaf/27063124</subfield>** </datafield>	**authorityURI="http://viaf.org/" valueURI="http://viaf.org/viaf/27063124"**> <namePart>Darwin, Charles</namePart> <namePart type="date">1809-1882</namePart> <role> <roleTerm type="text">creator</roleTerm> </role> </name>

Figure 2: MARC and MODS metadata with VIAF URIs

In database

Since the BHL already maintains local authority names, the first step should be reconciliation work with the VIAF. There are two ways to pursue this work:

1. first, the VIAF has an API service[4] that provides URIs for each name string;

2. second, as the WorldCat provides Linked Data Service for all OCLC records, we can utilize the service to have VIAF URIs already identified by OCLC.[5] Not all names will have matching VIAF URIs. (The University of Illinois at Urbana-Champaign Library did the reconciliation work in 2015, and it had 75% matching results, i.e., there were still names without authority records.) The BHL is well prepared for names without matching URIs, since it provides local AuthorIdentifierID and AuthorID for each name. When there is no match, the BHL database will keep using a local ID, or create a temporary URI for each name by using the local ID. This reconciliation work should be performed regularly, and the matching results can be updated to the database (see Figure 3).

Element	Name with VIAF URI	Name without VIAF URI
AuthorIdentifier		
AuthorIdentifierID	1 (PrimaryKey)	2 (PrimaryKey)
AuthorID	BHL1	BHL2
IdentifierID	N-1	N-2
IndetifierValue	**http://viaf.org/viaf/27063124**	**http://bhl.org/authoridentifierID2**

[4] https://www.oclc.org/developer/events/api-workshops/viaf-api-workshop.en.html; https://www.oclc.org/developer/news/2014/take-a-look-updated-viaf-api-documentation.en.html; http://viaf.org/viaf/data/

[5] Joel Richard experimented the work already how to fetch VIAF URIs from WorldCat in October 2018. The script is available in here: https://github.com/cajunjoel/bhl-oclc-creator-viafs

Figure 3: AuthorityIdentifier table example

The BHL Cataloging Group strongly believes that the BHL name reconciliation work will greatly improve not only the controlled vocabulary management but also the discovery services. By using authority files, names that are stored with different forms in the BHL database (see figure 4) will be cleaned up that will ultimately improve user experience.

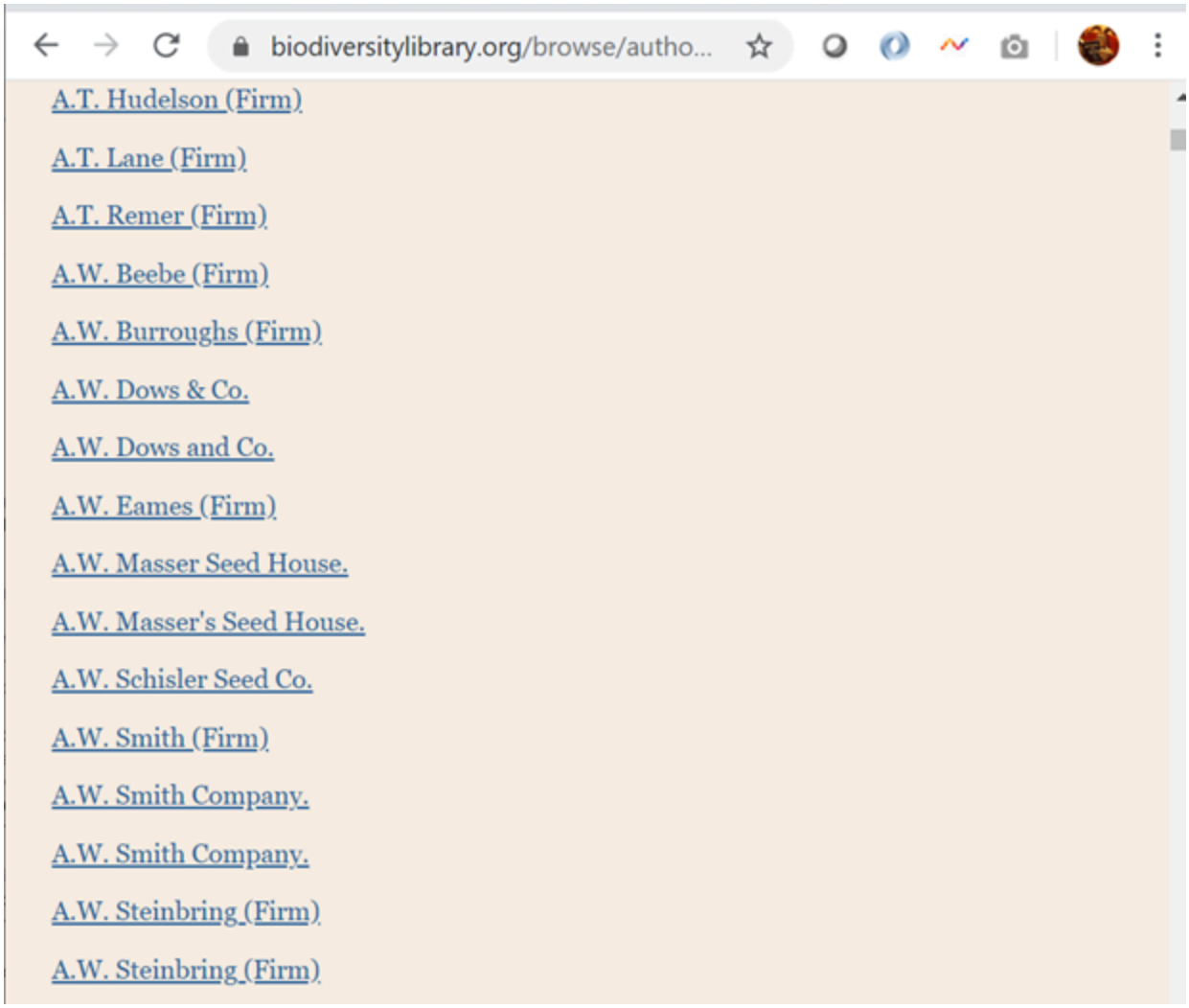

Figure 4: BHL Browse By Authors service uses the Names stored in the database as is. So names like ‘A.W.Smith (Firm), ‘A.W.Smith Company.’ and ‘A.W.Smith Company,’ are listed as different authors.

Procedures

Reconciliation work for the BHL names and VIAF URIs can be done in following procedures.

Phase 1 (3 to 6 months): Setting up Phase

Responsible party - BHL IT and Intern or Graduate Student(s)

1. Testing and finalizing the script: Although the script used for getting URIs from the WorldCat Linked Data Service by using OCLC numbers was developed in 2018, the script needs to be tested and finalized. It is also recommended to have a document that outlines how to run the script.
2. Preparing records with OCLC numbers: Records with OCLC numbers should be extracted for the work in CSV file format. The file should include local record ID, name, and local name ID.
3. Identifying personnel resources who will run the script
4. Identifying and allocating server space used for the work and any additional things needed for the work

Phase 2 (3-6 months): Reconciliation Phase

Responsible party - BHL IT, Intern or Graduate Student(s), Cataloging Group

1. Running the script: As the OCLC allows only a certain number of records for the service, it will be a while to complete the process (this also depends on the number of BHL records that have OCLC numbers)
2. Reviewing the search results: While it is impossible to check every record with results, it is recommended to spot check the result just to see whether the names in the BHL records are the same as the results.
3. Updating BHL names: With the results, start replacing and updating BHL names with the preferred names found in WorldCat service. VIAF URIs will be also added to the BHL database.

Phase 3 (~ 3 months): Wrapping up and planning next steps

Responsible party - BHL IT and Cataloging Group

1. Updating BHL names: With the results, replace and update BHL names with the preferred names found in WorldCat service. VIAF URIs will be also added to the BHL database.
2. Reviewing the BHL interface to see any remaining names with issues
3. Planning the workflow for those BHL records without OCLC numbers: BHL hosts resources from article databases that do not have OCLC numbers. To fully exploit linked data services, we should find ways to identify available linked data sources for authors names for scholarly articles and illustrators.

Phase 4: Improving discovery services with URIs

Responsible party - BHL

1. The other benefits of having URIs of name entities is that it can be used for improving discovery services. The linked data sources often include other links that contain additional contextual information as well as biographical information about names. This can be used for adding a knowledge card or browse service into the BHL resource page. This can be determined by the BHL user group or discovery team.

Resources needed for the Project

Project Coordinator

Project coordinator will monitor and manage the project so the project will run smoothly. Ideally a designated staff person would devote 25-30% of their time to project management.

IT staff

This person will test and finalize the script, oversee the graduate student(s) or any hourly who run the script. The responsibilities will also include finding server space, preparing data for the work, replacing the names and adding URIs to the database. Once all names are updated, this person will also work for improving new user interface with information available in the linked data sources.

Graduate Students(s)/Hourly

This position will mainly be responsible for running the script so names/entities in the records will have URIs. This position should work closely with the BHL IT staff.

BHL Cataloging Group

The BHL cataloging Group needs to work closely with BHL IT staff and the Coordinator to monitor the results and review names found from the reconciliation work. The Group will also identify possible linked data sources for names that do not have VIAF URIs.

Next steps

Enhance search and discovery services using URIs

This could be a long-term work. However, it is worthwhile to start discussing and exploring ways to improve search and discovery services by exploiting VIAF URIs.

Names from different sources and do not have URIs in VIAF

We will still have a lot of names (and records) that do not have linked data URIs. We have to keep looking for ways to find linked data URIs from different sources or creating WikiData links in batch mode.