

Orbis Cascade Alliance's Migration of Faceted Vocabularies to Primo VE

Junghae Lee and Blake L. Galbreath

Abstract

There has been a continued effort among cataloging communities to implement new fields and subfields in MARC 21 to support more structured data that benefit end users. In the current article, we describe how the Orbis Cascade Alliance integrated fields and subfields of this type into Primo VE, in order to provide faceted vocabularies and other facets within the discovery interface that were otherwise unavailable from the vendor. We discuss challenges in migrating the normalization rules behind the facets, the extent to which they are being utilized by member institutions, and possible next steps for promoting their use more extensively.

Keywords: Faceted vocabularies, Primo VE, Normalization rules, Consortium, Discovery platforms / services, Facet

Introduction

For many decades, Library of Congress Subject Headings (LCSH) have included headings and subdivisions that are used to describe what a resource *is* rather than what it is *about*, since headings and subdivisions can describe numerous non-subject aspects. The example LCSH strings below were assigned to a book of poems by African American authors. They consist of non-subject aspects such as language, genre/form, creator characteristics, and time period of creation.

LCSH strings: American poetry--African American authors; American poetry--20th century

Non-subject aspects: Language: English

Original language: English

Genre/Form: Poetry

Creator demographic group: African Americans; Americans

Time period of creation: 20th century

In 2007 the Library of Congress (LC) started developing faceted vocabularies such as Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT), Library of Congress Medium of Performance Thesaurus for Music (LCMPT), and Library of Congress Demographic Group Terms (LCDGT) to describe some of these non-subject aspects.¹ LC created new MACHine Readable Cataloging (MARC) fields and subfields to record Form of Work (380), Medium of Performance (382), Demographic Group Characteristics (385, 386), Place of Origin (370), and Time Period of Creation (046 \$k, l, o, p, 388) to complement existing MARC fields for Genre/Form (655), Language (008/35-37, 041), and Country of Production (257).

In recent years, diverse cataloging communities have focused on incorporating these faceted vocabularies into MARC 21 bibliographic records and discovery systems. One such community working on the implementation of faceted vocabularies is the Orbis Cascade Alliance (hereafter, Alliance), a Pacific Northwest consortium of thirty-eight academic libraries in Oregon, Washington, and Idaho. In 2019, the Alliance published its “Use of Faceted Terms in Cataloging.”² One of the guiding principles of this document is that the inclusion of these data points in bibliographic records will lead to usable facets for the refining of search results within library discovery interfaces.

Within this same timeframe, the Alliance Normalization Rules Standing Group (NRSG) built a host of normalization rules (or norm rules) in Primo Back Office (hereafter, Primo BO) to bring this guiding principle to fruition. Primo normalization rules define how metadata in bibliographic records display in a discovery interface.³ However, the construction of normalization rules is handled differently in Primo VE than in Primo BO. While Primo BO is a

standalone platform where normalization rules are created within web-based templates, Primo VE is built upon the Alma platform where normalization rules are written in a Drools language. Therefore, in order to plan for the Alliance's migration to Primo VE in 2022, the NRSG recreated over 70 normalization rules for display, search, and facet in Primo VE. In the current paper, we discuss the challenges we experienced while migrating these normalization rules to Primo VE, focusing on faceted vocabularies and other facets that were otherwise unavailable from the vendor, the extent to which they are being utilized by member institutions, and possible next steps for promoting their use more extensively.

Literature Review

The music cataloging community has been active among those who are creating faceted vocabularies since its members have long recognized that LC subject headings do not adequately describe materials in their discipline. In their paper, Iseminger et al. describe the rationale for faceted access to music resources, the process of designing and building these new vocabularies, as well as challenges for implementing them in library metadata.⁴ The majority of music headings in LCSH are not subjects, but are rather a combination of genre/form, medium of performance, and other aspects that tend to describe what a resource *is* rather than what it is *about*. The Music Library Association's (MLA) Cataloging and Metadata Committee has collaborated with LC to create new faceted vocabularies for music resources.

McGrath and Lowery outline the Orbis Cascade Alliance's original attempt to incorporate new vocabularies such as these into library metadata.⁵ They discuss the background and implementation of new MARC fields and subfields within Primo BO (e.g., Creator and Audience Demographics, Music: Medium of Performance, Music: Number of Performers, Music: Medium of Performance Statement, Country of Production, Composers, Directors, and

Performers). McGrath and Lowery detail the creation of display, search, and facet rules to meet the evolving needs of Alliance member libraries.

Nahotko analyzes 55 Polish university libraries' Online Public Access Catalogs (OPACs) to investigate practices in faceted navigation.⁶ He suggests that the introduction of faceted navigation deconstructs the traditional knowledge organization structure (KOS) and may require the creation of a new kind of KOS. By comparing information provided in various libraries' *Subject* facets against the structure of their originating MARC records, Nahotko demonstrates how other libraries are also adapting the MARC format to meet their local needs.

Whereas Nahotko focuses on facets in a variety of OPAC systems, Comeaux narrows this focus to the usability of facets within the web-scale discovery system Primo.⁷ His principal research question concerns how well patrons are able to distinguish among different material types, given that the very nature of web-scale discovery systems is to incorporate documents indexed at the article level with all other documents. Many other articles on facet usage in Primo followed Comeaux's seminal study. A number of these studies focus on the naming or appearance of facets within Primo, their overall usage, and their positioning within the list of options.⁸ Kliewer et al. also investigate the combination of facets used within a search session, noting that their participants used an average of 1.68 facets per session.⁹ However, none of these usability studies has incorporated facets that were not provided out-of-the-box by the vendor or focused on the overall quantity of facets presented to the user within the discovery layer.

Background

The NRSB of the Orbis Cascade Alliance has overseen the creation and dissemination of normalization rules since the Alliance first started using the library management system Alma and discovery interface Primo BO. The Alliance is a consortium of thirty-eight Alma Institution

Zones (IZ) and one shared Alma Network Zone (NZ). The NRSB has created and edited a core set of normalization rules over the years, based on best practices and feedback received in periodic surveys of the member institutions' wants and needs. Therefore, this standing group of six members primarily with cataloging and systems backgrounds already had a stable corpus of normalization rules from which to start the conversion to Primo VE.

The NRSB customizes out-of-the-box (OTB) rules and creates new local rules for display, search, and facets in the Alliance production Primo VE for the records linked to the NZ.¹⁰ Primo VE infrastructure allows for the centralized distribution of local rules 51-100. In addition, the NRSB creates local rules 31-50, but since they cannot be centrally “pushed” from the NZ, we host them on our GitHub repository. Since edited OTB rules also cannot be centrally distributed, we make those available on our GitHub repository as well. The terms “rules” and “fields” are sometimes used interchangeably in Primo VE since each field consists of a display rule

	▲ Field	↕ Display label	↕ Updated by
1	creator	Creator	8
2	format	Format	8
3	local_field_100	OCLC Number	Network
4	local_field_25	Local Author	8
5	local_field_47	NonLCSHNonMeSHSubjects	8
6	local_field_49	Dates of Publication	8
7	local_field_50	MESH Genre	8
8	local_field_51	Finding Aid Links	Network
9	local_field_52	General Search Index (not for display)	Network
10	local_field_53	MESH subjects	Network

Figure 1. Manage Display and Local Fields menu under Display Configuration in Primo VE.

and/or search/facet rule. Figure 1 shows the Manage Display and Local Fields menu in Primo VE, where local fields above 50 are updated by Network (i.e., by the Network Zone) and local fields below 51 and OTB fields, such as creator and format, are updated by a staff member in each institution (id starting with “8”).

After new or revised rules are reviewed and evaluated by an NRSB member, they are either pushed from the NZ to all member institutions’ production Primo VE instances or posted to the Alliance GitHub repository from which member libraries can copy them to their production Primo VE. It should be noted that each institution is responsible for configuring display and facet local fields in their Primo VE instance, since views cannot be configured centrally. After migrating to Primo VE in June 2022, the NRSB began receiving feedback from Alliance member libraries via Trello, which includes error reports, enhancement requests, and other suggestions as necessary.¹¹ The NRSB evaluates the items in this feedback, determines their feasibility, and communicates plans for their development.

Primo VE Normalization Rules

In addition to MARC data, Primo VE receives and combines data from “thousands of publishers, aggregators, and repositories” and translates that data into Primo Normalized XML (PNX) records.¹² The PNX record is divided into sections, such as control, display, search, and facets. Each section may contain multiple fields, and those fields may be a combination of both OTB fields and local fields. For example, PNX records for articles typically do not include local fields, while PNX records derived from MARC records can contain both. Local fields are used to add custom display, search, and facet rules within our discovery system. Primo VE transforms incoming data via normalization rules (written in Drools), which are maintained under the Alma Discovery menus.

Normalization rule for local_field_45

```
rule "Primo VE - Lds45"  
  when  
    MARC is "XXX"."a"  
  then  
    create pnx."display"."lds45" with MARC "XXX"."a"  
  end
```

Figure 2. A Primo VE normalization rule stanza provided by Alma.

Each Primo normalization rule consists of four sections: rule name, condition, action, and end statement (Figure 2). The first line contains the name of the rule. The next section begins with a “when” statement and defines what conditions have to be met for the rule to run. This section is used to check if specific MARC fields, subfields, and indicators exist in a record. The following section begins with a “then” statement and defines the processes that will operate on the data, which can include changing case, adding or removing punctuation, adding prefixes or suffixes, and concatenating various strings of data together. Temporary variables store strings of data for later use and regular expression commands allow for powerful data manipulation. This section typically ends by putting the manipulated string into a display field. Finally, the rule stanza concludes with an “end” statement, indicating that the current rule stanza should stop processing and move on to the next rule stanza, if one exists. Some MARC fields require more complicated rules than others and may need multiple rule stanzas to account for the different combinations of data.

Strategies for Implementation of Display and Facet Fields in VE

In order to establish a starting point for its migration planning, the NRSB reviewed the display and facet fields present in each institution’s Primo BO view(s) in April 2021 and prioritized those fields that occurred most frequently for development. This group integrated the following

faceted vocabularies within the Primo VE framework: *Creator Demographic, Audience, Genre, MeSH Genre, Date or Period of Creation, Associated Place, Country of Production, Music: Medium of Performance*, and *Music: Medium of Performance Statement*. Additionally, we incorporated the following other facets: *Performer, Director, Composer, Music: Number of Performers, Numeric Designation of Musical Work, Musical Key, Original Language*, and *NLM Classification*. The list of faceted vocabularies (Figure 3) described in this paper may be viewed within the NRSB Primo production interface (https://orbiscascade-network.primo.exlibrisgroup.com/discovery/search?vid=01ALLIANCE_NETWORK:NRSB), which is used by various Alliance standing groups for development purposes. Strategies for the initial (Primo BO) implementation of faceted vocabularies and other facets were detailed previously in McGrath and Lesley.¹³ In this paper, we focus on those strategies for the implementation of faceted vocabularies and facets that were not covered previously and where significant adaptation was necessary within the Primo VE framework.

Creator Demographics

As in Primo BO, the creation of local field *Creator Demographics* was fairly straightforward. We created similar display and search/facet rules from *Creator/Contributor Characteristics* (386) and did not limit *Creator Demographics* by specific vocabularies since the common practice is to use controlled vocabularies, such as LCDGT and LCSH. However, in Primo VE, we were additionally able to include data from MARC field 386 \$3 (Materials specified) and \$i (Relationship information) for display (not for search/facet) to meet the requests from the Alliance members. Data from \$3 and \$i are added to the beginning of field 386 for all types of records as a prefix (followed by a colon for readability) based on member requests as well as simplicity in rule implementation. When recording creator and audience characteristics for

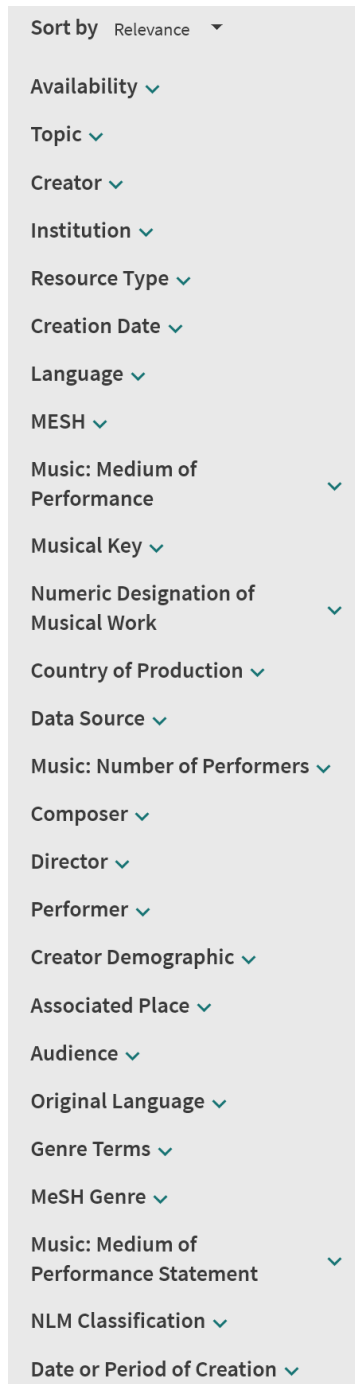


Figure 3. Facets in NRSB Primo VE production interface.

compilations, the presence of \$3 in the fields 385 (Audience Characteristics) and 386 (Creator/Contributor Characteristics) is useful in identifying particular characteristics of a part of the resource. A music sound recording album is a common example of compilation.¹⁴

505 00 \$t Bolero / \$r Richard Trythall -- \$t Quartett für Schlagzeuger / \$r Alfred Schnittke -- \$t X-Pression / \$r Percussion Art Quartett -- \$t Sun Song I/II / \$r Peter Sculthorpe -- \$t Stick attack / \$r Armin Weigert.

Trythall is an American, Schnittke was a Soviet and Russian, Sculthorpe was an Australian, and the Percussion Art Quartett and Weigert are German composers. The 386 fields in this MARC record are as follows.

386 ## \$3 Bolero \$a Americans \$2 lcdgt

386 ## \$3 Quartett für Schlagzeuger \$a Soviets \$a Russians \$2 lcdgt

386 ## \$3 Sun song I/II \$a Australians \$2 lcdgt

386 ## \$3 X-Pression ; Stick attack \$a Germans \$2 lcdgt

If the subfield \$3 is not included in the display rule, the nationality of the composers displays in Primo VE as follows.

Creator Demographic	Americans
	Soviets
	Russians
	Australians
	Germans

After adding the subfield \$3 in the display rule, each work is tied to the appropriate nationality as follows.

Creator Demographic	Bolero: Americans
	Quartett für Schlagzeuger: Soviets; Russians
	Sun Song: Australians
	X-Pression: Germans
	stick attack: Germans

Audience

Local field *Audience* is a conglomeration of data from different MARC fields. For display, the NRSB is combining Audience Characteristics (385), the audience code from the 008 fixed field (008/22), and Target Audience Note (521). In Primo VE, it is possible either to create a table of criteria within the code that serves as a lookup or to use special syntax that refers to a mapping table within Primo VE. For *Audience*, we chose the former approach, since there were only eight codes to account for within the 008/22: {a, b, c, d, e, f, g, or j}. Each code in the 008 fixed field was transformed into Library of Congress Demographic Group Terms (LCDGT) within the normalization rules (Figure 4). For the Target Audience Note field (521), we added prefixes to the display based on the indicator values, where applicable, to describe narrow meanings. The NRSB included Audience Characteristics (385) and the audience code from the 008/22 in the search/facet rules but omitted the Target Audience Note field (521). Additionally, to include any work without the 385 field and the code in the 008 fixed field (008/22), we transformed values “[J]juvenile” and “[Cc]hildren’s” from 6xx fields to “Children” in the search/facet rules.

```
rule "Primo VE - Lds98_008_22_b"  
  when  
    MARC.control."008"(22-23) equals "b" AND NOT  
    MARC.control."LDR"(6-7) matches "e|f|p" AND NOT  
    MARC.control."LDR"(6-8) matches "ab|ai|as" AND NOT  
    MARC is "385"  
  then  
    set TEMP"1" to position("008",22-23)  
    add prefix (TEMP"1","School children,Children")  
    remove string (TEMP"1","b")  
    create pnx."display"."lds98" with TEMP"1"  
end
```

Figure 4. A part of Audience display rules that displays LCDGT from the 008 fixed field (008/22).

Genre

In Primo VE, the OTB *Genre* display field was only mapped to the Form Subdivision of Subject Access fields (6xx \$v) and was not mapped to the Genre/Form field (655 \$a). Rather, the Genre/Form field (655 \$a) was mapped to the *Subjects* display field. To map all the genre data available in the MARC fields and subfields to the local *Genre* display field, we removed the OTB *Genre* display rule from operation (an action we have termed “blanking out,” because we leave a blank rule in place of the OTB rule), and implemented a local *Genre* display field, which includes Genre/Form Term (655), Form of Work (380), and the genre codes from the 008 fixed field. Beginning December 2022, Ex Libris made improvements to its genre terms mapping, updating the display for genre terms to use 655 \$a.¹⁵ However, the NRSB decided to continue to use a local field for *Genre*, since this field is designed to include additional genre data. Therefore, we are using a modified version of the OTB rule for *Genre*.

While codes in the 008 fixed field for audience are unique and therefore can be transformed to constant text (LCDGT) in the normalization rules, codes in the 008 fixed field for genre vary according to type of material. Therefore, it was necessary to map the codes for genre to description (LCGFT when applicable) in the *Local Field Translate* table (Figures 5 and 6).

```
rule "Primo VE - Lds78 008/34"
#biography
  when
    (MARC.control."008"(34-35) equals "a" OR
     MARC.control."008"(34-35) equals "b" OR
     MARC.control."008"(34-35) equals "c") AND
    MARC.control."LDR"(6-8) matches "aa|ac|ad|am"
  then
    set TEMP"1" to position("008",34-35)
    add prefix (TEMP"1";"$Tgenre_008_Biog_")
    add suffix (TEMP"1";"$")
    create pnx."display"."lds78" with TEMP"1"
end
```

Figure 5. A part of *Genre* display rules that shows mapping of the 008 fixed field to the code in the *Local Field Translate* table.

Code	Description
genre_008_Biog_a	Autobiographies
genre_008_Biog_b	Biographies
genre_008_Biog_c	Biographies
genre_008_LitTxtSR_a	Autobiographies
genre_008_LitTxtSR_b	Biographies
genre_008_LitTxtSR_c	Conference papers and proceeding
genre_008_LitTxtSR_d	Drama
genre_008_LitTxtSR_e	Essays

Figure 6. The Local Field Translate table maps code "c" to different descriptions.

Date or Period of Creation

The NRSB created identical display and search/facet rules for *Date or Period of Creation*.

However, a considerable amount of parsing was necessary to account for the various formats found in MARC field 046 (Special Coded Dates). To begin, we used regular expressions to distinguish between dates formatted in pattern yyyy, yyyy-mm, or yyyyymmdd and replaced those with labels according to ISO 8601 standards, via the *Local Field Translate* table (Figure 7). We

Date or Period of Creation	
<input checked="" type="checkbox"/> 1900-1999 (2)	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> 1910s (2)	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> 1800-1899 (1)	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> 1880s (1)	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> 1900-1909 (1)	<input checked="" type="checkbox"/>
<input type="checkbox"/> 1920s (1)	<input checked="" type="checkbox"/>

Figure 7. Date or Period of Creation facet.

also accommodated BCE dates by inserting the string “BCE” after the converted date. When possible, the NRSB also included values from MARC field 388 (Time Period of Creation).

Below is an example MARC record which includes multiple dates (OCLC #26399697).¹⁶

008 920625t19931992nyu b 000 f eng

046 1# \$k 1866 \$2 edtf

046 2# \$k 1992 \$2 edtf

245 10 Crime and punishment : \$b a novel in six parts with epilogue / \$c by Fyodor Dostoevsky ; translated and annotated by Richard Pevear and Larissa Volokhonsky.

264 #1 New York : \$b Vintage Books, \$c 1993.

264 #4 \$c ©1992

The above MARC record includes the following dates: Date of creation of work (1866), Date of creation of expression (the English translation by Pevear and Volokhonsky) (1992), and Date of manifestation (1993).

Based on the previous practices in Primo BO, we excluded the first indicator value “2” (Expression) from both 046 (Special Coded Dates) and 388 (Time Period of Creation) fields in the display and facet rules for *Date or Period of Creation*. Therefore, currently only *Creation Date* and *Publication Date* are shown in the display field.

Publication Date 1993

Creation Date 1866

In the future, Alliance members may need to further discuss whether or not to include the expression date (first indicator value “2”) in both display and facet fields.

Associated Place

For *Associated Place*, the NRSB created similar display and search/facet rules from MARC field

370 (Associated Place): \$g (Place of origin of work or expression), \$c (Associated country), and \$f (Other associated place). We added prefixes to the display based on the subfield codes, where applicable, to describe narrow meanings. Currently, we do not include subfields \$3 (Materials specified) and \$i (Relationship information) in the *Associated Place* display rules. Below is an example MARC record with 370 \$i (OCLC #756585911).¹⁷

370 ## \$i Setting: \$f Seattle (Wash.) \$2 naf

370 ## \$i Setting: \$f Antarctica \$2 lcs

The above MARC record currently does not display \$i in Primo VE as follows.

Associated Place	Seattle (Wash.)
	Antarctica

Since the *Associated Place* display rules are straightforward, adding these subfields would be a request that the NRSB could easily fulfill.

Original Language

Language was one of the most complex fields to implement in both Primo BO and Primo VE. Data for *Original Language* included MARC field 041 (Language Code) \$h (Language code of original) and displayed with prefix *Original Language* under the *Language* field with other language data. The display rule for *Original Language* mapped three-character alphabetic codes used in MARC records to the *Language Codes Labels* table. Below is an example MARC record with multiple original languages (OCLC #1158419826).

041 1# eng \$h fre \$h ger \$h ita

546 ## Text in English, with translations from the German, French and Italian.

The above MARC record displays in the *Language* display field as follows (Figure 8).

Language Text in English, with translations from the German, French and Italian.
Original language(s): French; German; Italian

Figure 8. Original language in Display.

The search/facet rule for *Original Language* was considerably more complicated than the display rule and took several months to bring to production. The NRSB included MARC field 041 (Language Code) \$h (Language code of original) when codes came from ISO 639-3 or MARC language codes, then replaced three-character alphabetic language codes with languages in the *Language Codes Labels* table. Mapping to the MARC language codes is working, however mapping to ISO 639-3 codes is not, because the mapping table lacks the necessary ISO codes. Currently, the search/facet rule for *Original Language* consists of seven rules that can accommodate up to six original languages. Due to system limitations in Primo VE, the NRSB will have to add additional normalization rule stanzas for every additional original language that might exist in a MARC record. We currently have no way to predict what that upper limit might be and have therefore stopped coding at six languages until more information is gathered.

Since the display and facet rules for *Original Language* only include data from MARC field 041 (Language Code) \$h (Language code of original), if there is no 041 \$h in the MARC record, the original language will not display. In order to make the original language surface in display and facet fields, catalogers can code the original language even if the manifestation is not a translation and only includes the original language, as in the following example MARC record (OCLC #1237211183).¹⁸

```
008 210210t19471947fr      000 f fre c
```

```
041 0# $h fre
```

The above MARC record displays in Primo VE as follows.

Language French

Original language(s): French

Results

Tables 1 and 2 summarize the MARC fields and subfields included in each display and search/facet rule. For some rules, the NRSNG included identical display and search/facet rules (e.g., *Creator Demographic*, *Date or Period of Creation*, and *Associated Place*), while for other rules we placed greater restrictions in the search/facet rules so that similar vocabularies displayed under one terminology. For example, while the *Genre* display rule includes all terms in the 380 field, the *Genre* search/facet rule only includes LCGFT in the 380 field. When available, the NRSNG plans to include \$3 (Materials specified) and \$i (Relationship information) in the display rules but is currently still creating a systematic way to incorporate that development.

Table 1. MARC Fields Used for Faceted Vocabularies.

Primo VE Field Label	MARC Fields Used in Display	MARC Fields Used in Search/Facet
Creator Demographic	386 \$a	386 \$a
Audience	385 \$a 008/22 521 \$a \$b	385 \$a 008/22 655 #7 \$a* 6xx #0 \$v juvenile
Genre	655 #0, 655 #4, 655 #7* 380 \$a 694 \$a 008/24-27, 30-31, 33, 34	6xx #0 \$v 655 #7 (\$2 lcgft or \$2 fast) 380 \$a (\$2 lcgft)
MeSH Genre	655 #2	655 #2
Date or Period of Creation	046 \$k \$l 046 \$o \$p 388 \$a	046 \$k \$l 046 \$o \$p 388 \$a
Associated Place	370 \$c \$f \$g	370 \$c \$f \$g
Country of Production	257 \$a	257 \$a (\$2 naf)
Music: Medium of Performance	382 \$a \$b \$d \$e \$n \$p \$r \$s \$t \$v	382 \$a \$b \$d \$p
Music: Medium of Performance Statement	N/A	382 \$a \$b \$e \$n

* 655 #7 \$a is limited to specific vocabularies selected by the Alliance.

Table 2. MARC Fields Used for Other Facets.

Primo VE Field Label	MARC Fields Used in Display	MARC Fields Used in Search/Facet
Performer	N/A*	100/110 \$a \$b \$c \$d \$q** 700/710 \$a \$b \$c \$d \$q**
Director	N/A*	100 \$a \$b \$c \$d \$q** 700 \$a \$b \$c \$d \$q**
Composer	N/A*	100 \$a \$b \$c \$d \$q** 700 \$a \$b \$c \$d \$q**
Music: Number of Performers	N/A	382 \$s \$e 382 \$t
Numeric Designation of Musical Work	383 \$a \$b \$e 383 \$c \$d	383 \$a \$b 383 \$c
Musical Key	384 \$a	384 \$a
Original Language	041 \$h	041 \$h
NLM Classification	N/A	060 \$a Q or R

* Already displays in the Creator or Contributor field (100/110 or 700) as in \$4 or \$e.

** \$4 or \$e of 100/110 or 700 is limited to specific relationship designators indicating performer, director, and composer.

Discussion

Challenges

The NRSB encountered several challenges and learning moments while migrating normalization rules from Primo BO to Primo VE. Namely, we exhausted our rule inventory sooner than expected. This group’s original plan was to create fifty local rules to disseminate to member institutions and leave the remaining fifty local rules for each member institution to use as they wanted. However, because Primo VE does not allow for the same flexibility with OTB rules as under Primo BO, the NRSB was forced to expend extra local rules to accommodate this lack of parity. As one example, the OTB *Genre* field contains a mix of MeSH and non-MeSH genre

```

52 rule "Primo VE - Genre 654"
53     when
54         MARC is "654"."v"
55     then
56         create pnx."display"."genre" with MARC."654"."v"
57     end
58
59 rule "Primo VE - Genre 655"
60     when
61         MARC is "655"."v"
62     then
63         create pnx."display"."genre" with MARC."655"."v"
64     end
65
66 rule "Primo VE - Genre 656"
67     when
68         MARC is "656"."v"
69     then
70         create pnx."display"."genre" with MARC."656"."v"
71     end

```

Figure 9. A portion of the original OTB Genre display rules shows that it only includes subfield \$v (Form subdivision) from 6xx fields and makes no restrictions according to indicator values. As a result, the subfield \$v from both MeSH and non-MeSH is displayed.

terms and does not display genre terms from the 655 \$a Genre/Form index term (although these are included in the OTB search/facet rules) or those derived from the 008 fixed field elements (these are not available for display, search, or facet in the OTB rules) (Figure 9). In response, the NSRG removed the OTB *Genre* field and created a local field to display non-MeSH genre terms in the MARC 655 \$a (and \$v, x, y, and z subdivisions for non-LCGFT terms), 380 (Form of work), 008 positions 24-27, 30-31, 33, 34 as appropriate, plus local genre terms in 694 (Alliance: Local form/genre) (Figure 10). As a result of needing to create extra rules, from the perspective of Primo BO, the NRSRG depleted their allotment and was forced to secure twenty more rules to complete our project.

Secondly, the NRSRG discovered a bug within VE that did not exist under Primo BO. The normalization rules for mapping the genre code from the 008 fixed field are not working

```

rule "Primo VE - Lds78"
#655_7 lcft OCLC 774607605
  when
    MARC:"655" has any "a" AND
    MARC:"655".ind"2" equals "7" AND
    MARC:"655"."2" match "lcft"
  then
    create pnx:"display"."lds78" with MARC "655" subfields "a" delimited by " " remove substring using
    regex "\\.+ $"
  end

rule "Primo VE - Lds78 655 #7 except lcft"
#655_7 fast OCLC 05264690
  when
    MARC:"655" has any "a" AND
    MARC:"655".ind"2" equals "7" AND
    (MARC:"655"."2" match
    "aat|fast|ftamc|gmppc|gsafd|lcac|lcs|lcs|lcs|local|homoit|mesh|ng|olac|vqgt|rbin|rbgen|rbpap|rbpri
    |rbprov|rbpub|rbtyp|mim|migfg")
  then

```

Figure 10. A portion of NRSB's local Genre display rules shows that it includes 655 \$a, but only when the second indicator value is equal to "7." As a result, MeSH genres are excluded from Genre.

properly. That is, the facet rules display codes from the mapping table instead of displaying human understandable descriptions. Figure 11 shows that the genre code from the 008 fixed field displays the code "genre_008_NatureContents_e" instead of displaying "Encyclopedias" while genre terms from 655 \$a and 380 \$a successfully display "Encyclopedias." We have not yet uncovered how to map codes from the 008 fixed field to descriptions in the *Local Field Translate* table and are currently exploring different approaches. Therefore, codes from the 008 fixed field are successfully incorporated into the *Genre* display but are currently excluded from the *Genre* facet.

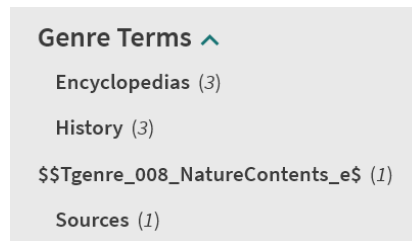


Figure 11. A draft Genre facet shows that search/facet rules are not correctly matched to the Local Field Translate table.

To continue, unlike in Primo BO, data from different fields do not deduplicate when merged together into one field. However, they do continue to deduplicate when coming from multiple instances of the same MARC fields. Using OCLC #71328147 as an example in Figure 12, we see that the genre ‘Programmed instructional materials’ occurs twice. The first instance (1) is mapped from the ‘p’ code of the 008 fixed field, while the second instance (4) is mapped from the 655 fields. The two occurrences are deduplicated between the separate 655 fields (fast and lcft), but not against the 008 field. This new lack of deduplication is an unfortunate characteristic of Primo VE.

```
008 060530s2006  txua  bp  001 0 eng
655 #2 Programmed Instruction $0 (DNLM)D020498
655 #7 self-instructional materials. $2 aat $0 (CStmoGRI)aatgf300026379
655 #7 programmed textbooks. $2 aat $0 (CStmoGRI)aatgf300026387
655 #7 Programmed instructional materials. $2 fast $0 (OCoLC)fst01423785
655 #7 Programmed instructional materials. $2 lcft
655 #7 Matériel d'enseignement programmé. $2 rvmgf $0 (CaQQLa)RVMGF-000001277
```

Genre	1	Programmed instructional materials >
	2	self-instructional materials >
	3	programmed textbooks >
	4	Programmed instructional materials >

Figure 12. Duplicate data in Genre display in Primo VE.

As was to be expected, the NRSB encountered a significant learning curve when needing to translate the Primo BO rules, which were created within a Graphical User Interface (GUI) using Java routines, to the text-based environment of Primo VE using Drools logic. This group spent a considerable amount of time understanding the new options that were available and those that no longer existed. On the one hand, the text-based coding environment was a welcomed

change from the GUI. Group members were easily able to transfer code from text editors to Alma to GitHub via simple copy and paste commands and insert comments directly into the coding, significantly increasing everyone's access to and understanding of the code base. The Drools logic was also very similar to other code many of the NRSB members had already written in Alma. On the other hand, we did lament not having certain Java routines at our disposal. For example, in Primo BO, we were able to use a Split Subfields command within the rule for *Audience* that allowed us to assign multiple translations per code within the 008/22, but no such command exists in Primo VE. We found that the issue can be mitigated for display by adding a paragraph tag (<p>) in place of the separating comma. However, no such workaround is possible for search/facet rules.

Wynne and Hanscom discuss how next-generation catalogs may expose errors and inconsistencies in catalog data that were not prevalent prior to being displayed in the new system.¹⁹ The NRSB also found this to be true and witnessed a few instances of these errors when data did not deduplicate as expected. Although correcting these metadata errors is beyond the scope of the NRSB, this group learned to remove duplicate and/or redundant punctuation from the end of rule stanzas via Replace String by String commands and to use Remove Leading and Trailing Spaces commands to remove errant whitespace in the MARC records for both display and search/facet rules. While it is possible to correct these metadata errors in Alma using Alma normalization rules, it should be noted that all Alliance records linked to the NZ are overlaid from daily OCLC loads. Therefore, in Alma, Alliance members usually do not correct any fields that are not protected from OCLC overlay, but do correct local fields (i.e., local extensions) that are protected from OCLC overlay.

NRSG Facets Adopted in Member Libraries

Despite our best efforts, there is the long-standing problem of incomplete data in certain MARC fields.²⁰ For example, not all records contain the appropriate data encoded for *Audience*, making the facet for this field only useful for certain records. The issue of not having enough data in the records may mislead patrons into thinking they are retrieving all relevant results. Consequently, several Alliance member libraries have chosen not to activate some of these faceted fields until we are able to retrospectively enhance substantially more records. It should be noted here that the Core Subject Analysis Committee (SAC) Subcommittee on Faceted Vocabularies (SSFV) has been developing guidelines and modules (along with human intervention) for retrospective enhancement of bibliographic records with faceted vocabularies.²¹ Therefore, institutions and libraries now have the necessary guidance for filling in gaps in bibliographic records through retrospective cataloging.

To uncover other areas where Alliance members might not be taking advantage of the local fields created by the NRSG, the authors reviewed the facets present on each institution's primary Primo VE view in January 2023 and compared them to those on each institution's primary Primo BO view collected in April 2021 before migration. The data in Tables 3 and 4 show that the number of faceted vocabularies and other facets implemented by member libraries decreased from Primo BO to Primo VE. Many institutions within the Alliance continue to utilize the faceted vocabulary and other facet rules built by the NRSG within their display fields, but not as facets. On the low end, in Primo VE, only eight institutions are displaying *Date or Period of Creation* and *Country of Production*, while almost all institutions are displaying *Genre* (34 out of 38). In terms of facets, twelve institutions are using *Genre*, while there is virtually no utilization of the remaining faceted vocabulary rules. Although *Genre* ranks first in both the display and

Table 3. Changes in the Number of Alliance Institutions Using Faceted Vocabularies from Primo BO to Primo VE (Display/Facet Fields).

Primo Field Label	Changes in the No. of Display Fields Used from Primo BO to Primo VE	Changes in the No. of Facet Fields Used from Primo BO to Primo VE
Creator Demographic	11 → 10	2 → 0
Audience	10 → 14	4 → 0
Genre	20 → 34	9 → 12
MeSH Genre	6 → 4	5 → 1
Date or Period of Creation	12 → 8	3 → 0
Associated Place	9 → 9	1 → 0
Country of Production	13 → 8	4 → 1
Music: Medium of Performance	16 → 13	8 → 1
Music: Medium of Performance Statement	N/A	6 → 1

Table 4. Changes in the Number of Alliance Institutions Using Other Facets from Primo BO to Primo VE (Display/Facet Fields).

Primo Field Label	Changes in the No. of Display Fields Used from Primo BO to Primo VE	Changes in the No. of Facet Fields Used from Primo BO to Primo VE
Performer	N/A	6 → 2
Director	N/A	4 → 2
Composer	N/A	7 → 2
Music: Number of Performers	N/A	5 → 1
Numeric Designation of Musical Work	14 → 10	4 → 1
Musical Key	13 → 12	5 → 1
Original Language	13 → 18	3 → 1
NLM Classification	N/A	4 → 2

facet columns and is used more than in Primo BO, undoubtedly some of its dominance stems from the fact that *Genre* was originally an OTB field in Primo VE that the NRSB converted to a

local field. It should be noted that *Audience* and *Original Language* are the only local fields whose usage increased in display in Primo VE, although their usage decreased in terms of facets.

Given the low adoption rates of NRSB-created facets as post-search filters, should this group continue to provide such an elaborate list of faceted vocabulary and other facet rules for member institutions, or should we consider paring down the list to reduce the overall amount of maintenance on the part of the standing group? We should also consider that the paring down of options would reduce the cognitive load on member institutions, since they have to test and make decisions to include or exclude every newly-introduced facet.

Considerations

For the purposes of this paper, we limited our investigation to the default Primo view at every Alliance institution. It is possible that more specialized views (e.g., Music Library) are taking greater advantage of the facet inventory maintained by the NRSB. This is a potential topic of future research. The inclusion of these facets will also be different for different user types. For example, researchers who primarily focus on articles over books will not see these locally-created facets, because these facets are not included in records coming from the Central Discovery Index (CDI) (Ex Libris' unified index for scholarly and academic material, containing over 5 billion records).²² For this reason, we would encourage vendors to make these fields available in their records for VE so they would be applied to CDI records, thus increasing consistency and the ability to facet among the records.

Researchers have conducted numerous usability tests on Primo (and discovery layers more generally) over the years. However, to the authors' knowledge, none have focused on the quantity of facets displayed to the user within the discovery layer.²³ It is possible that future usability tests will uncover the optimal number and/or arrangement of facets to display to the

user. Meanwhile, we can posit that some institutions and Primo administrators believe there is an upper threshold for the number of facets a user might endure. And, if true, these institutions and administrators would want to limit the number of facets displayed in their results list to whatever they deem manageable. Therefore, another topic of future research would be to survey the Primo administrators at each of the Alliance institutions, asking if they have performed usability testing to justify the exclusion of certain facets or if their institutions came to this conclusion via other means. For example, perhaps they can demonstrate a correlation between the number of facets used and system performance.

We must also consider the educational component needed for each institution to fully understand the host of options available to them. All thirty-eight institutions have been busy with migration and post-migration related tasks. The NRSB has steadily been providing updates to normalization rules since migration concluded and, although we have also provided a master index of the rule inventory,²⁴ some institutions may be falling behind in keeping up with the changes. To get ahead of this potential issue, the NRSB has begun to create and deliver presentations at open call sessions where interested parties can attend. Here, attendees can ask questions and clear up any confusion they might be experiencing related to the normalization rule updates. Although it is beyond scope of the NRSB to recommend which fields member institutions enable in their views, we do hope our educational outreach will at least keep them abreast of the available options.

Conclusions

Although the NRSB did experience challenges in migrating our host of Primo BO normalization rules into the Primo VE framework, ultimately, we were able to accomplish the majority of our task. These challenges included working within a new system that, although more user-friendly,

does not share the same flexibility and parity with Primo BO. This group was forced to expend local rules because search/facet indexes cannot be edited and invent workarounds for coding routines that no longer exist. We also note that mapping tables do not function properly with certain data points and deduplication is not as powerful as under Primo BO. Additionally, in reviewing the degree to which our faceted vocabulary and other facet rules have been adopted across the Alliance, the NRSB must decide whether to continue to maintain the rules supporting these fields going forward. If we do continue to provide our institutions and patrons with such a variety of options for searching and faceting, the standing group should provide more informational sessions to member institutions to apprise them of the existing options.

Notes

¹ Mark McKnight, "Are We There Yet? Toward a Workable Controlled Vocabulary for Music," *Fontes Artis Musicae* 59, no. 3 (2012): 290, <https://www.jstor.org/stable/42765606>.

² Orbis Cascade Alliance, "Use of Faceted Terms in Cataloging" Google Docs, updated December 2020, https://drive.google.com/file/d/1wlGuj8Xy3CLGfsOhHVnU1YOZU3lZI4gJ/view?usp=embed_facebook.

³ "Configuring Normalization Rules for Display and Local Fields," Ex Libris Knowledge Center, last modified November 25, 2021, [https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_\(English\)/050Display_Configuration/Configuring_Normalization_Rules_for_Display_and_Local_Fields](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/050Display_Configuration/Configuring_Normalization_Rules_for_Display_and_Local_Fields).

⁴ Beth Iseminger et al., "Faceted Vocabularies for Music: A New Era in Resource Discovery," *Notes* 73, no. 3 (2017): 409-31, <https://doi.org/10.1353/not.2017.0000>.

-
- ⁵ Kelley McGrath and Lesley Lowery, “Getting More out of MARC with Primo: Strategies for Display, Search and Faceting,” *The Code4lib Journal* 41 (2018), <https://journal.code4lib.org/articles/13600>.
- ⁶ Marek Nahotko, “Knowledge Organization Affordances in a Faceted Online Public Access Catalog (OPAC),” *Cataloging & Classification Quarterly* 60, no. 1 (2022): 86-111, <https://doi.org/10.1080/01639374.2021.2015734>.
- ⁷ David J. Comeaux, “Usability Testing of a Web-Scale Discovery System at an Academic Library,” *College & Undergraduate Libraries* 19, no. 2-4 (2012): 189–206, <https://doi.org/10.1080/10691316.2012.695671>.
- ⁸ Kelsey Brett, Ashley Lierman, and Cherie Turner, “Lessons Learned: A Primo Usability Study.” *Information Technology and Libraries* 35, no. 1 (2016): 7–25, <https://doi.org/10.6017/ital.v35i1.8965>; Blake Lee Galbreath, Corey Johnson, and Erin Hvizdak, “Primo New User Interface: Usability Testing and Local Customizations Implemented in Response,” *Information Technology and Libraries* 37, no. 2 (2018): 10–33, <https://doi.org/10.6017/ital.v37i2.10191>; W. Jacobs, Mike DeMars, and J. M. Kimmitt, “A Multi-Campus Usability Testing Study of the New Primo Interface,” *College & Undergraduate Libraries* 27, no. 1 (2020): 1–16, <https://doi.org/10.1080/10691316.2019.1695161>; Kylie Jarrett, “Findit@Flinders: User Experiences of the Primo Discovery Search Solution,” *Australian Academic and Research Libraries* 43, no. 4 (2012): 278–99, <https://doi.org/10.1080/00048623.2012.10722288>; Greta Kliewer et al., “Using Primo for undergraduate research: a usability study,” *Library Hi Tech* 34, no. 4 (2016): 566–84, <https://doi.org/10.1108/LHT-05-2016-0052>.
- ⁹ Kliewer et al., 572.
- ¹⁰ “Configuring Normalization Rules for Display and Local Fields.”
- ¹¹ Orbis Cascade Alliance, “NRSG Primo VE Feedback,” accessed February 15, 2023, <https://trello.com/b/Ij4Zi9db/nrsg-primo-ve-feedback>.

-
- ¹² “An Overview of CDI,” Ex Libris Knowledge Center, accessed February 1, 2023, [https://knowledge.exlibrisgroup.com/Primo/Content_Corner/Central_Discovery_Index/Documentation_and_Training/Documentation_and_Training_\(English\)/CDI_-_The_Central_Discovery_Index/010An_Overview_of_the_Ex_Libris_Central_Discovery_Index_\(CDI\)](https://knowledge.exlibrisgroup.com/Primo/Content_Corner/Central_Discovery_Index/Documentation_and_Training/Documentation_and_Training_(English)/CDI_-_The_Central_Discovery_Index/010An_Overview_of_the_Ex_Libris_Central_Discovery_Index_(CDI)).
- ¹³ Kelley McGrath and Lesley Lowery, “Getting More out of MARC with Primo.”
- ¹⁴ Adam L. Schiff, email message to author, June 15, 2017.
- ¹⁵ “Primo VE 2022 Release Notes,” Ex Libris Knowledge Center, updated January 20, 2023, https://knowledge.exlibrisgroup.com/Primo/Release_Notes/002Primo_VE/2022/010Primo_VE_2022_Release_Notes.
- ¹⁶ Schiff, Adam L., email message to author, January 31, 2023.
- ¹⁷ Schiff.
- ¹⁸ Schiff.
- ¹⁹ Susan C. Wynne and Martha J. Hanscom, “The Effect of Next-Generation Catalogs on Catalogers and Cataloging Functions in Academic Libraries,” *Cataloging & Classification Quarterly* 49, no. 3 (2011): 179, <https://doi.org/10.1080/01639374.2011.559899>.
- ²⁰ Wynne and Hanscom, 189-190.
- ²¹ Casey Mullin et al., “Retrospective Implementation of Library of Congress Faceted Vocabularies, March 2022,” Report (ALA Core Subject Analysis Committee, March 25, 2022), <https://alair.ala.org/handle/11213/17998>.
- ²² “An Overview of CDI.”
- ²³ Although, not a usability study, and representing older technology than this paper discussed, Hall did collect statistics on the quantity of facets used in academic libraries. The average number of facets displayed in her study was 9.8, with a range of 5 to 18. Catherine E. Hall, “Facet-Based Library

Catalogs: A Survey of the Landscape,” *Proceedings of the American Society for Information Science and Technology* 48, no. 1 (2011): 1–8, <https://doi.org/10.1002/meet.2011.14504801195>.

²⁴ Orbis Cascade Alliance, “A Guide to Display and Local Fields in VE for Alliance Members,” Google Docs, accessed March 18, 2023, https://docs.google.com/document/d/1BkfbHhqcPnfsa0j4BdAXQyqg1ZgQBterU351ZDdG0hI/edit?usp=embed_facebook.