**Context:**

Computer programs have made two sweeps through the LC/NACO Authority File. These sweeps are known as "Phase 1" and "Phase 2" and are part of a project to prepare the LC/NACO Authority File for use under RDA.

* The Phase 1 program marked certain records with a special 667 field. Records so marked were those whose coding or 1XX characteristics indicate that the heading could not be used under RDA without review, and which were also not susceptible to the mechanical changes made during phase 2. The text of this special 667 field begins "THIS 1XX FIELD CAN NOT BE USED UNDER RDA". Reasons for marking an authority record in this manner include a code in 008/32 other than "z" or "c", the presence in the 1XX field of a heading for a conference, and text in subfield $c of 100 fields not present in a restricted list.
* The Phase 2 program found records containing data elements in a non-RDA form that could be manipulated into an RDA-like form. (For example, subfield $d in a personal name reading "b. 1952" could be changed to "1952-".) Some of the changed records also fit the criteria used during Phase 1, and were labeled with the special 667 field; the program re-coded as RDA those changed records that were coded AACR2 in 008/10 and that did not receive the special 667 field.

In each phase, the program created additional RDA-related fields when possible. (For example, the program re-formulated subfield $d in personal name headings into an 046 field.)

These two phases together identified about 500,000 authority records as being not appropriate for use under RDA without review, or suitable for use under RDA as-is. In very round numbers, this left about 7.5 million AACR2 authority records with no indication of RDA compliance. Until now, catalogers have been offered the option of using the 1XX field in these authority records in their RDA bibliographic records, or upgrading the authority records to RDA. Most of the time, the only thing that must be done to a record to mark it as suitable for continued use under RDA is to change the value of 008/10 to "z" and to add subfield $e reading "rda" to the 040 field. (NACO participants are welcome, even encouraged, to make other changes, such as the addition of 046 and 3XX fields, at the same time they re-code an AACR2 record as RDA.)

The proposed Phase 3 of the manipulation of the LC/NACO Authority File for use under RDA will consider primarily the approximately 7.5 million records not yet labeled as acceptable for use under RDA, or marked for review before use under RDA. A new program will either re-code these records as RDA, or (using additional criteria defined since Phase 1 and Phase 2) add to them a 667 field indicating that the 1XX field cannot be used under RDA without review. Because RDA has changed in the time since the completion of Phase 2, the program may also re-evaluate records modified during Phases 1 and 2 to see if further changes are possible. Finally, the program may also consider records coded for RDA, to ensure that their 1XX fields are fully supported by additional authority records (this work may include multi-element access points, such as name/title and subordinately-entered corporate bodies, but the group should also explore access points with parenthetical qualifiers).

**Charge:**

The PCC RDA Authorities Phase 3 Task Group is charged to:

* Develop a plan for making Phase 3 changes to the LC/NACO Authority File, based on the recommendations contained in the [Appendix: Proposal for Phase 3, as narrowly defined](#Appendix).
* Develop a scheme for extending the principles of the [Appendix: Proposal for Phase 3, as narrowly defined](#Appendix) to the remaining records in the LC/NAF.
* Describe a communication plan for informing the community (beyond the PCC, including vendors) of plans and activities.
* Recommend next steps, including the need for any follow-on groups, additional community tasks, etc.
* Develop a mechanism for tracking, updating and testing the automated changes described in the [Appendix: Proposal for Phase 3, as narrowly defined](#Appendix).
* Develop a mechanism for tracking an error rate for the automated changes.
* Develop a method for propagating changes in the authority files once the mechanisms have been tested, addressing additional dependencies that these changes may create.
* The [Appendix: Proposal for Phase 3, as narrowly defined](#Appendix) plans to exclude (but show in a report) records whose 008/32 bears code ‘b’. The TG should consider whether and how best to use this report to mark undifferentiated records as not being in compliance with RDA.
* Given the fact that the majority of records in the LC/NACO Authority File will need to be updated during Phase 3, consider additional but unrelated work that will likely require the updating of all the records in the file, and could be done simultaneous to Phase 3. Specific examples include:
	+ **024 fields:** Primary among the possible projects is the addition of 024 fields to authority records containing one or more types of identifiers (ISNI, VIAF, and/or others). An authority record may be updated to include one or more identifiers even if no other change is made to the record. The TG will investigate possible sources for these identifiers and will recommend ways to integrate them into respective NAF records.
	+ **Subfield $c:** There is a call to re-visit the categorization of texts in subfield $c of personal names that are allowed and not allowed under RDA. Some AACR2 records currently marked with the special 667 field will be re-coded as RDA. The contents of subfield $c in some records will be enclosed within parentheses, even if the record is not re-coded as RDA.
	+ **Elimination of pre-RDA 678 fields:** Before RDA, the 678 field was generally used when an authority card was being converted into the MARC format. Rather than provide full 670 citations, inputters were allowed the option of summarizing the important pieces of information in the 678 field. As originally implemented, this 678 field was not intended for public display. The use of the 678 field has been expanded in recent practice; it can now contain any information, of any length, that the constructor of the authority record wishes to convey; and this field is now available for public display. Whenever possible, pre-RDA 678 fields should be identified, and re-coded as 670 fields (according to a pattern of practice established by the Library of Congress).

There may be yet other similar projects, as yet unknown. Any such work undertaken as part of Phase 3 will call for changes to the project's scope, and the likely increase the number of records to be updated. It may be assumed throughout that any record otherwise changed by the Phase 3 program for any reason will be enhanced with the same fields (046 for personal names, for example) that were created by program during Phase 1 and Phase 2.

**Time Frame:**

The Task Group is encouraged to move quickly and to work interactively with PoCo. Do not wait for the formal deadlines below if you would like to communicate recommendations or questions in order to move ahead with the actual re-coding.

* Deliver progress reports to PoCo in advance of meeting events on the PCC calendar: Operations Meeting, May 2014; ALA Annual, June 2014; Policy Committee, November 2014; ALA Midwinter, January 2015.
* Final report by March 15, 2015. Final report deadline does not mean all the automated work must be completed by then; the report will include a projected timetable for the completion of all tasks.

**Formation of Task Group:** March 2014

**Date of final report to Policy Committee:** March 15, 2015

**Chain of Reporting:** The Task Group’s progress and final reports will be reviewed by the PCC Steering and Policy Committees and then announced to PCCLIST and posted to the PCC web site.

**Task Group Members:**

* Gary Strawn, Chair (Northwestern University), mrsmith@northwestern.edu
* Karen Anderson (Backstage Library Works), kanderson@bslw.com
* Robert Bremer (OCLC), bremerr@oclc.org
* Ana Cristan (Library of Congress), acri@loc.gov
* Paul Frank (Library of Congress), pfrank@loc.gov
* Richard Moore (British Library), Richard.Moore@BL.UK
* David Williamson (Library of Congress), dawi@loc.gov

**Appendix: Proposal for Phase 3, as narrowly defined**

Gary Strawn, Northwestern University Library

February 18, 2014

A program will scan a copy of the LC/NACO Authority File held by the Library of Congress, which uses the Voyager library management system. The program will start with the first record in the file, and examine each record in turn until it has changed a specified number of records. (The number of records that the program must examine to reach its set limit of changes will vary from one run of the program to the next.) The next time the program is started, it will begin where it left off, and continue reading records until it has again changed a specified number of records. The program will be run repeatedly, until it has processed the entire LC/NACO authority file from beginning to end.[[1]](#footnote-1)

*Inclusion criteria*

The criteria for identifying a record of interest given here are preliminary, and are based solely on the needs for Phase 3, as narrowly defined.

* Include only records with an 010 field whose $a text begins "n"[[2]](#footnote-2)
* Exclude records already bearing a 667 field with text beginning "THIS 1XX FIELD"
* Exclude records whose 008/10 contains any code other than "c"[[3]](#footnote-3)
* Exclude (but show in a report) records whose 008/32 bears code "b"

*Examination of the 1XX field*

Each authority record's 1XX field[[4]](#footnote-4) will be divided as appropriate into segments, proceeding from left to right.[[5]](#footnote-5) Each such segment will be compared against 1XX and 4XX fields in other LC/NACO authority records.[[6]](#footnote-6) With some possible exceptions (see below), the program expects that each *segment shorter than the complete heading* will match the 1XX field in an authority record; the program expects that the *complete* heading will either match no other authority field at all, or will (harmlessly) match one or more 5XX fields. In addition, if a segment of a heading matches the 1XX field in another record, that second record must not be marked with the special 667 field,[[7]](#footnote-7) and cannot bear code "b" in 008/32.

This work results in the sorting of authority records into two piles:

1. Authority records with 1XX fields suitable for use under RDA. (The full 1XX does not match anything unacceptable; shorter segments (if any) all match acceptable 1XX fields.)
2. Authority records with 1XX fields not suitable for use under RDA. (The complete 1XX matches something unacceptable; segments shorter than the complete 1XX match nothing at all, or something unacceptable.)

The program will re-code records in the first pile as RDA; the program will add the special 667 field to records in the second pile.[[8]](#footnote-8)

The following are suggested as possible expansions to this general manner of proceeding. These exceptions result in the move of the 1XX field in an authority record from the "unacceptable for use under RDA" pile to the "acceptable for use under RDA" pile.

* If the name portion of a **name/title heading** matches an acceptable authority record, and if the portion of the heading following subfield $t consists only of subfields $k, $l, $s[[9]](#footnote-9) and/or $f (in any combination and in any order), then re-code the authority record as RDA.

*Example:* n 00001030

*Heading:* 100 1# $a Balch, James F., $d 1933- $t Prescription for nutritional healing. $l Spanish

*The name portion matches an AACR2 authority record, the name plus title matches no other authority data, and the full heading matches no other authority data. This record can be re-coded as RDA.*

*Example:* n 00021803

*Heading:* 100 1# $a Bulgakov, Mikhail, $d 1891-1948. $t Works. $l French. $f 1997

*The name portion matches an RDA authority record; none of the remaining segments matches any authority data. This record can be re-coded as RDA.*

*Counter-example:* n 00109385

*Heading*: 100 1# $a Chauvet, Louis Marie. $t Sacrements. $l English

*The name portion of the heading matches an authority record bearing the special 667 field. This name/title record should receive the special 667 field.*

*Counter-example:* n 00063872

*Heading:* 100 1# $a Dressel, Erwin, $d 1909-1972. $t Sonatas, $m saxophone, piano, $n op. 26

*The name portion of the heading matches an RDA authority record; the title portion includes subfields other than $k, $l, $s or $f. This name/title record should receive the special 667 field.*

* If a **title heading** contains only subfields $a, $n, $p, $k, $l, $s and/or $f and if none of the shorter segments matches something unacceptable, then re-code the authority record as RDA.

*Example:* n 00123058

*Heading:* 130 ## $a Minimal-invasive chirurgie. $l English

*Subfield $a does not match any authority data. This record can be re-coded as RDA.*

*Example:* n 94042359

*Heading:* 130 #0 $a Hamburger Beiträge zur Angewandten Mathematik. $n Reihe A, $p Preprint

*There is no authority record, or conflicting authority information, for subfield $a alone, or subfield $a plus subfield $n, and there is no conflicting information for the full heading. This record can be re-coded as RDA.*

*Counter-example:* n 00050183

*Heading*: 130 #0 $a Alcoholics Anonymous. $l Armenian

*Subfield $a is the same as a corporate heading (n 78087105). This record should receive the special 667 field.*

**Expansion of Phase 3 to add one or more 024 fields**

Various projects currently underway produce databases of identities; each identity is assigned a standard identifier. The entities represented in these databases overlap with the entities represented in the LC/NACO Authority File. It may be possible to generate a table that equates (by means of identifiers) entities in the LC/NACO Authority File with entities in one or more of these external databases.[[10]](#footnote-10) If such a table or series of tables can be generated, [[11]](#footnote-11) the standard identifiers used by those projects could be incorporated into the corresponding LC/NACO authority records.[[12]](#footnote-12) This would be a major step toward an environment in which system can hop from one bit of information to another via links.

For this work, the scope of Phase 3 must be expanded to include all records in the LC/NACO authority file.

The program will search the identifier for each LC/NACO record in the table (or tables) of correspondences. If the program finds a match, it will add one or more 024 fields to the LC/NACO authority record for the external identifiers. The program will avoid the addition of duplicate identifiers when already present in the record.

Because the Phase 3 work will occur over an extended period of time, the program will need to take into account changes to the table(s) of correspondences between external databases and the LC/NACO file. However, the program will only apply changes to records that have not been examined; changes to identifiers that occur after they have been added by this program are outside the scope of this project.

**Expansion of Phase 3 to include reconsideration of subfield $c**

During Phases 1 and 2, authority records for personal name headings containing subfield $c were subjected to a test: if the text of subfield $c was not present in a closed list of recognized texts, the authority record was labeled with the special 667 field. In addition, parentheses were added around the text in subfield $c in some cases. Changing views of the nature of subfield $c in the time since the completion of Phase 2 may lead to reconsideration of this work.[[13]](#footnote-13)

If the handling of subfield $c is to be reconsidered, Phase 3 will need to be expanded to include all LC/NACO authority records for personal names with subfield $c.

If the program finds that the text of subfield $c is now acceptable for use under RDA, and if the 1XX field bears no other characteristic that bars the heading from use under RDA until reviewed, the program will remove the 667 field and re-code the record as RDA. Regardless of the current state of the subfield $c text, the program may add, or remove, parentheses around the text in subfield $c.

**Expansion of Phase 3 to include work on the 678 field**

The goal of this work will be the re-tagging of as many pre-RDA 678 fields as possible, leaving the majority of the occurrences of the 678 field containing text that is suitable for public display. If this project is accepted, the scope of Phase 3 will be expanded to include all LC/NACO records with 678 fields. Whenever it can be done reliably, the program will transform unambiguous birth and death dates (or beginning and/or ending dates for corporate bodies) into coded information in the 046 field.

*Example:* n 00001030 *before modification*

100 1# $a Pougy, Liane de, $d 1869-1950

670 ## $a Liane de Pougy, c1994: $b p. 14, etc. (b. July 2, 1869 in La Flèche, Sarthe; d. Dec. 26, 1950; married to Georges Ghika, a Roumanian prince, in 1910; after his death entered the tiers ordre de saint Dominique, and took the name soeur Anne-Marie de la Pénitence)

678 ## $a d. 12/26/50

*After modification:*[[14]](#footnote-14)

046 ## $f 18690702 $g 19501226

100 1# $a Pougy, Liane de, $d 1869-1950

670 ## $a Liane de Pougy, c1994: $b p. 14, etc. (b. July 2, 1869 in La Flèche, Sarthe; d. Dec. 26, 1950; married to Georges Ghika, a Roumanian prince, in 1910; after his death entered the tiers ordre de saint Dominique, and took the name soeur Anne-Marie de la Pénitence)

670 ## $a Info. from 678 field, Feb. 18, 2014 $b (d. 12/26/50)

1. The Phase 2 program changed 30,000 records per day. This figure was arrived at after negotiations among the various NACO nodes. Based on this rate of work and an estimate of 7.5 million records to be changed, the work of Phase 3 as narrowly defined will take about 250 business days. (The program can only be run at LC Monday-Friday.) If Phase 3 is expanded to include additional work, the outside-case scenario involves updating every record in the LC/NACO Authority File; if this is the case, the project will run for approximately 17 additional days. One way or another, with allowance for weekends, holidays, staff vacations and unplanned hiccups, all of this work can be expected to extend over an entire calendar year. [↑](#footnote-ref-1)
2. The Voyager authority file against which the program will be run also contains LCSH records. [↑](#footnote-ref-2)
3. Code "c" means "AACR2". This test excludes RDA records, and all records coded for earlier cataloging rules. All records coded for earlier cataloging rules should already have the special 667 field. It is an interesting question, whether records already coded RDA should be examined by this program, to ensure that their 1XX fields match the same tests as those put to AACR2 records. [↑](#footnote-ref-3)
4. No examination will be made of the 4XX or 5XX fields in any of the authority records. [↑](#footnote-ref-4)
5. For example, given the 110 field "110 1# $a United States. $b Congress. $b Senate" the program will test three headings: the geographic name "$a United States" and the corporate names "$a United States. $b Congress" and "$a United States. $b Congress. $bSenate". This does not mean that every subfield code marks the end of a segment; the personal name heading "$a Eliot, T. S. $q (Thomas Stearns), $d 1888-1965" consists of a single segment. [↑](#footnote-ref-5)
6. The program will draw on a generic code module to do this work. The steps performed by this module are too complicated to describe in this outline document. The program described here simply passes the selected authority record to the module, and then inspects the results. [↑](#footnote-ref-6)
7. In order for this test to behave properly, the program will have to make a secondary examination of each matching record that cannot be used under RDA, to see if changed criteria now allow the 1XX field to be used under RDA. [↑](#footnote-ref-7)
8. Preliminary testing indicates that more than 98% of AACR2 records can be re-coded as RDA. [↑](#footnote-ref-8)
9. Note most carefully the omission of subfields $n and $p from this list. [↑](#footnote-ref-9)
10. Generation of such a table is, happily, outside the scope of this document. [↑](#footnote-ref-10)
11. It is not yet known how many different kinds of identifiers will be available for addition to authority records during this work. Each different kind of identifier will have to be labeled (either by default, if each different kind is available in a separate table, or explicitly if all are contained in a single table) so that the program can supply the appropriate code for 024 subfield $2. [↑](#footnote-ref-11)
12. As long as each identifier from an external system is paired unambiguously with a piece of information present in the LC/NACO Authority File (either the text of 010 subfield $a, or LC's Voyager authority record ID), the program can transfer the information reliably into authority records. [↑](#footnote-ref-12)
13. Exactly what different criteria may be used to evaluate the text of subfield $c remains to be determined. [↑](#footnote-ref-13)
14. The program extracted the complete birth date for the 046 field from the original 670 field; the program extracted the full death date from both the 670 and 678 fields; finding both dates to be identical, it used the full death date. These extractions rely on code already developed for other projects. [↑](#footnote-ref-14)