

To: Joint Steering Committee for Development of RDA
From: Pat Riva, Chair, JSC Capitalization Instructions Working Group
Subject: Capitalization Instructions and RDA

Related

6JSC/Chair/17/2015

6JSC/CCC/12

6JSC/CCC/Discussion/1

Abstract

This discussion paper focuses on options for the presentation of language-specific capitalization instructions within RDA itself or within the RDA Toolkit.

Introduction

The task (#1) for the JSC Capitalization Instructions Working Group for 2015 is to “review the content and coverage of RDA Appendix A and prepare a proposal/discussion paper by Aug. 3, 2015”. In this discussion paper, the WG summarizes its discussion to date and presents some issues for which feedback from JSC and constituencies would be appreciated.

The WG task is divided into three sub-tasks:

“1.1. Review the instructions on capitalization of RDA elements (A.0-A.9) to ensure these are complete.

1.2. Propose possible paths for presenting the language-specific capitalization instructions in a way that is useful for cataloguers working in the translations of RDA as well as those working in English.

1.3. Identify updates needed in specific languages, where the relevant expertise for that language is available.”

The focus of this discussion paper is sub-task #1.2, in addition, brief status reports on the other two sub-tasks are included.

Discussion

Sub-Task 1.1: Instructions on capitalization of RDA elements (A.0-A.9)

In the previous papers that led to the establishment of this WG (particularly 6JSC/CCC/12 and 6JSC/CCC/Discussion/1) several issues in this part of the RDA instructions were identified. Some preliminary further assessment by WG members has highlighted some unclear explanations and additional gaps.

At present RDA A.2-A.9 is partly organized around specific entities (such as A.2 Names of Persons, Families, Corporate Bodies, and Places; A.3 Titles of Works; A.4 Titles of Manifestations), and partly around specific statements in the description (such as A.5 Edition Statement; A.6 Numbering of Serials; A.7 Numbering within Series and Subseries; A.8 Notes; A.9 Details of Elements). This organization is an adaptation of the AACR2 structure that seemed based around the ISBD areas. An idea the WG might explore for this section is to follow the structure of sections, chapters, sub-sections from RDA itself, this would have the result of ensuring that no data element is left out.

Sub-Task 1.3: Evaluation of instructions for specific languages

WG members have made some preliminary assessments of the existing instructions for certain languages and language families, namely, Finnish, French, Swedish, and Slavic languages. In each case issues were identified. These can be characterized as:

- Incorrect examples, i.e. not following the instruction as worded (and either the instruction or the example may be at fault);
- Lack of coverage of a situation for a given language when that language does not in fact follow the same instruction as does English, but due to the current structure of Appendix A this would lead the cataloguer to apply the English convention, leading to an incorrect result;
- Generally dated choice of examples, and emphasis on situations that would now be rarely encountered in a cataloguing situation.

Additionally, the DNB had previously reported that the German rules are not entirely in accordance with the most recent edition of the most authoritative manual.

Also, the WG began to assess the idea of developing a common template for presenting capitalization instructions, at least within groups of related languages. There is definitely room for improvement in the consistency of presentation of the existing instructions, as unmotivated variation in the order and labelling of instructions exists.

All this confirms that the specific instructions for languages other than English require careful review. The extent and nature of this review will depend on preferences for the options for presenting these instructions to the cataloguer.

Sub-Task 1.2: Structure and presentation of language-specific instructions

The WG's assessment for options in the presentation of language-specific instructions are given in the context of considerations of desired or potential coverage and with certain working assumptions in mind.

Language Coverage

Presently 23 languages or language families, plus English, are covered. All but three (Bosnian Croatian, Danish) were included in AACR. In some cases, related languages are grouped by language family (Scandinavian), in other cases this technique is not used, although the languages in the family share many conventions (Slavic and, to a lesser extent, Romance languages). Some languages within families are not mentioned, even though other languages of that family (and with essentially the same conventions) are. Although Arabic and Hebrew (in transliteration) are not covered, there is a specific instruction for names beginning with el- or al- at A.2.2 and A.4.1, placed in the general section. Appendix C (Initial articles) covers 54 languages. There is no policy statement in force guiding the coverage decisions. In the cases where languages in a family share conventions, or languages using the same script share conventions within a cultural area, coverage could be extended very simply. For comparison, ISO 639-1 (2-character language codes) are assigned to 136 languages, while ISO 639-2 (3-character language codes) includes 464 codes (not all of which represent individual languages).

Given that extending coverage to all of the world's extant 6900+ languages¹ is out of the question, and in some writing systems the issue is irrelevant, what criteria for selection can be used?

As whatever level of coverage is attempted will not be complete, an “in case of doubt” provision could be developed to guide a cataloguer when dealing with a language that is not covered, or a situation is not covered. This might be along these lines:

If a cataloguer is working with a language that is not covered by specific instructions, or encounters a situation that seems exceptional in a language that is covered, prefer to follow indications in the resource being catalogued (such as whether the word in question is capitalized within the resource in text that is presented with both upper and lower case letters) rather than to do research in external sources (for example, when an unknown word in a title turns out to be a proper name of a fictional character or imaginary place in a novel).

In any case, an agency may select a style guide for a specific, regularly encountered language that is not covered.

Working Assumptions

1 <http://www.linguisticsociety.org/content/how-many-languages-are-there-world>

The WG has been operating under the following tacit assumptions, and would appreciate feedback as to their appropriateness.

- That internationalization principles need to be applied; in particular that RDA in translation needs to be as usable and as complete as RDA in English.
- That the convenience of the user is served by using language-specific conventions such as capitalization in a way that will feel natural to readers of the language, as it will avoid a potential barrier to the acceptance of and confidence in the data presented in the catalogue.
- That the principles of common usage or practice and of uniformity are both relevant and do not need to conflict.
- That it is worthwhile having guidance on capitalization practice, tailored so as to be relevant to resource description and access points, somewhere within RDA or in ancillary tools. (As an alternative to expecting all agencies to select existing external tools or to independently create their own tools.)
- That cataloguers need to consult such rules for various languages, not just for the language of the catalogue of their institution, since resources in many languages are likely to be included in the database, and the resources may be multilingual.
- That cataloguers need to be alerted to differences in practice between languages.
- That guidance, if not too complex to follow, will actually save cataloguer time by avoiding lengthy research about fine points of grammar.
- That while RDA cannot provide comprehensive coverage of all of the world's languages, it can provide guidance for some languages, reflecting the needs of agencies that apply RDA in English or in translation.

Structure

Currently Appendix A consists of guidelines in two categories:

- How to capitalize in specific RDA data elements, as discussed under sub-task #1.1 above, which are basically language-neutral
- How to capitalize different types of words within data depending on the language of the data

The first type is probably well suited to being part of Appendix A, in the body of RDA, pending review of its structure.

The second type could either be in Appendix A or be presented as a separate resource in the Tools tab. Responses to previous papers from CCC indicated interest in exploring the two options.

The WG considered some of the advantages and disadvantages of these options:

	In Appendix A	On Tools Tab
Pro	Would be a full part of RDA, present in all translations	Would have to determine whether RDA translations are expected to include or not Presently translations are not obligated to translate material on Tools tab (such as RDA-MARC21 mappings)
	Would be published in RDA Toolkit, but also normally included in all print editions of RDA in all languages	Would be published in RDA Toolkit Print editions might or might not need to include all or part of it
Con	A full and balanced treatment for all languages selected for coverage may be lengthy and impractical	Length is less of an issue online Filtering might serve to give focus to the cataloguer's current need
	Would need to have a single linear sort order so that instruction numbers can be assigned	Does not have to have a single sort order from the Toolkit user's point of view (although there would be an order based on internal IDs that would be visible to those editing RDA in the CMS)
	Need to balance linguistic bias (alphabetical order as now) with need to find a specific language	Can provide both groupings by language families and links to specific languages, no one single access path required

In terms of ordering the languages in a language-neutral way, the ISO 639-1 language codes could serve as a key, with the language names given according to the RDA translation language.

On balance the WG feels that a Tools tab resource is more flexible, and definitely more easily expanded for future needs. Considerable cooperation as to technical feasibility will be required, should this option be deemed worth pursuing.