

String Encoding Schemes: Community Choice in RDA

Kathy Glennan

Chair, RDA Steering Committee

Head, Original & Special Collections Cataloging

University of Maryland



R | D | A

Resource Description & Access

What is a string encoding scheme (or SES)?

From the RDA Glossary:

string encoding scheme

A set of string values and an associated set of rules that describe a mapping between that set of strings and a value of an element.

Use for: syntax encoding scheme

See also: element

Set of string values = textual character string(s)

Associated set of rules = instructions

Mapping = matching

Value of an element = a representation of that element

What is a string encoding scheme (or SES)?

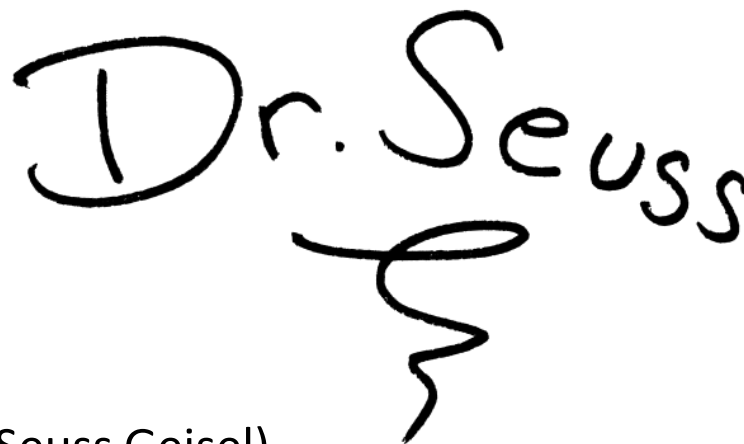
- In more traditional cataloging terms, an SES is about
 - How to construct an access point
 - What information to record
 - In what order
 - With what punctuation
 - How to construct a “statement” about an edition, series, imprint, etc.
- These decisions may vary by cataloging tradition or practice
 - Anglo-American, German, Norwegian, Spanish, etc.

SEs in RDA

- A search in the beta Toolkit on “string encoding scheme” currently returns 124 results, including
 - Edition statement
 - Production statement
 - Publication statement
 - Distribution statement
 - Manufacture statement
 - Series statement
 - Dissertation or thesis information
 - Access point for...
 - Authorized access point for...
 - Variant access point for...
 - and most elements associated with Nomen

SEs May Vary by Agency

- All of these are preferred forms in VIAF, for at least one cataloging agency
 - Dr. Seuss
 - Dr. Seuss, pseud.
 - Dr. Seuss (1904-1991)
 - Доктор Сьюз, псевд., 1904-1991.
 - Seuss, Dr.
 - Seuss, Dr., 1904-1991
 - Seuss, Dr., 1904-1991 (duln. f. Theodor Seuss Geisel)
 - Сьюз 1904-1991 доктор
 - Geisel, Theodor Seuss, 1904-1991

A handwritten signature in black ink that reads "Dr. Seuss". The letters are cursive and somewhat slanted to the right. Below the signature is a decorative, wavy flourish.

Major Question

- As an international standard, what is the best way for RDA to accommodate this kind of community variation?
 - Approach #1: Hard code SESs in the instructions, adding each variant in its own option box
 - Approach #2: Move SES content out of the element pages, providing access through multiple links or expandable boxes instead
 - Approach #3: Move the SES content from the element pages to policy statements
- Each approach requires that communities
 - Determine the appropriate SES for each situation
 - Create policy statements that
 - Identify the designated option (Approaches #1 & #2), or
 - Contain the actual instructions (Approach #3)

Approach #1

- Hard code the SES instructions for each variant – mock-up

STRING ENCODING SCHEME

Record a value that includes, in this order:

1. a value of Agent: access point for agent
2. a value of a base access point for a work

Apply the string punctuation pattern:

"value 1. value 2"

Join each of the values with a full stop followed by a space

STRING ENCODING SCHEME

Record a value that includes, in this order:

...

[Send to Back](#)

LC-PCC

LC/PCC practice: Apply this option.

[Send to Back](#)

LC-PCC

LC/PCC practice: Do not apply.

R | D | A
Resource Description & Access

Approach #2

- Provide information about SEs through links – mock-up

CONDITION

A work is a *single work*.

A work is created by two or more agents in collaboration.

OPTION

String encoding scheme

Example

LC-PCC

[Send to Back](#)

LC/PCC practice: Do not apply.

OPTION

String encoding scheme

Example

LC-PCC

[Send to Back](#)

LC/PCC practice: Do not apply.

OPTION

String encoding scheme

Example

LC-PCC

[Send to Back](#)

LC/PCC practice: Apply this option.

Approach #3

- Move SES instructions to policy statements
(image from 63.83.87.43 – *sample* content & formatting)

Larger place

OPTION

Treat the following as a larger jurisdiction:

- a state, province, territory, etc., of Australia, Canada, the United States, or a country that was a constituent republic of the former U.S.S.R. or of the former Yugoslavia
- England, Northern Ireland, Scotland, and Wales
- an overseas territory, dependency, etc.

LC-PCC

[Send to Back](#)

A place is located within a human settlement.

Apply this string encoding scheme:

Record a value that includes, in this order:

1. a value that is based on Place: **name of place**

RSC Decision – Approach #3

- Policy statement approach minimizes proliferation of options on element pages
 - Clearly presents community SES decision to catalogers in the context of an element page
- Re-use of current RDA instructions supported by the Toolkit's content management system
- Decisions and maintenance left up to communities, but...
 - If preference is to continue practices from the original Toolkit, no new intellectual work needed
 - Use ISBD and Appendix E information
 - If preference is linked data, SESs are irrelevant

Community Vocabularies in RDA

R | D | A
Resource Description & Access

Prompted by Increased Internationalization

- Legacy instructions and guidance in RDA (largely untouched from AACR2) have an Anglo-American focus
 - Abbreviations and symbols
 - Additional instructions on names of persons
 - Capitalization
 - Initial articles
 - Terms of rank
- Some content already relocated from RDA instructions to a separate section under “Resources” (or “Tools” in the original Toolkit)
 - Books of the Bible (in 2016)
 - Terms for medium of performance (in 2018)

Supporting Different Choices

- Different cataloging communities have different preferences
 - Not all conform to Anglo-American practices
 - Communities may wish to support interoperability with legacy data
- Cataloging communities' preferences may change over time
 - As the PCC community's practices evolve, should we still apply the original Toolkit's Appendix B (Abbreviations and Symbols)?
- RDA needs to support development, maintenance, and publication of local vocabularies as part of the Toolkit
 - Identified as a long-term solution to the failed ALA proposal to add more gender terms in 9.7.1.3 (in 2015)

***Under development –
expect further changes through 2020***

Community Vocabularies: Characteristics

- Compliant with RDA (and LRM)
- Associated with
 - Processing of strings in SESs
 - Community-specific vocabulary encoding schemes (VES)
 - These do not necessarily conform to a full VES
 - Concepts may lack definitions, scope notes, IRIs, or notations
 - Coverage may be incomplete
- Used only by specific RDA communities
 - Cover a limited number of languages and scripts
 - Maintained by experts in specific languages, scripts, and/or cultures
 - If continuing existing practices, no new intellectual work needed
 - Do not require translation in every Toolkit language

A Transition in Process

Resources Tab Beta Toolkit – Feb. 2020

RESOURCES
Glossary
Vocabulary Encoding Schemes
Relationship Matrix
Abbreviations and symbols
Additional instructions on names of persons
Capitalization
Initial articles
Terms of rank
Revision History
Books of the Bible
Terms for medium of performance
AACR2

Resources Tab Beta Toolkit – April 2020

RESOURCES
Glossary
Vocabulary Encoding Schemes
Relationship Matrix
Additional instructions on names of persons
Capitalization
Initial articles
Revision History
Community vocabularies
AACR2

Resources Tab: New Pop-Outs

GUIDANCE	POLICIES	RESOURCES
		Glossary
		Vocabulary Encoding Schemes
		Relationship Matrix
		Additional instructions on names of persons
		Capitalization
		Initial articles
		Revision History
		Community vocabularies
		AACR2
Names in the Arabic alphabet		
Burmese and Karen names		
Chinese names containing a non-Chinese given name		
Icelandic names		
Indic names		
Indonesian names		
Malay names		
Roman names		
Romanian names containing a patronymic		
Thai names		
Recording surnames that include an article and/or preposition		

Resources Tab: New Pop-Outs

GUIDANCE	POLICIES	RESOURCES
		Glossary
		Vocabulary Encoding Schemes
		Relationship Matrix
		Additional instructions on names of persons
		Capitalization
		Initial articles
		Revision History
		Community vocabularies
		AACR2
General Guideline for Capitalization		
General guideline for English language capitalization		
Other languages		

Resources Tab: New Pop-Outs

GUIDANCE	POLICIES	RESOURCES
		Glossary
		Vocabulary Encoding Schemes
		Relationship Matrix
		Additional instructions on names of persons
		Capitalization
Initial articles listed by language		Initial articles
Initial articles listed by word or words		Revision History
		Community vocabularies
		AACR2

Resources Tab: New Pop-Outs

GUIDANCE	POLICIES	RESOURCES
		Glossary
		Vocabulary Encoding Schemes
		Relationship Matrix
		Additional instructions on names of persons
		Capitalization
		Initial articles
		Revision History
		Community vocabularies
		AACR2
Abbreviations		
Books of the Bible		
Terms of rank		
Terms for collective titles		
Terms for gender		
Terms for medium of performance		

Community Vocabularies

- **Draft** content, migrated from existing Toolkit instructions

Resources > [Community vocabularies](#) > [Abbreviations](#)

Abbreviations

- [Abbreviations for countries and states](#) →
- [Abbreviations in Latin script](#) →
- [Abbreviations in Cyrillic script](#) →
- [Abbreviations in Greek script](#) →
- [Abbreviations in Hebrew script](#) →

Community Vocabularies

Draft content

- New pop-out for Books of the Bible

	Abbreviations
Books of the Bible: Library of Congress-Program for Cooperative Cataloging	Books of the Bible
Festlegungen für den deutschen Sprachraum zum Erfassen der bevorzugten Titel von biblischen Schriften	Terms of rank
Livres de la Bible : Bibliothèque et Archives Canada-Bibliothèque et Archives nationales du Québec	Terms for collective titles
	Terms for gender
	Terms for medium of performance
	167-3241 Administration About

Community Vocabularies

Draft content

Resources > [Community vocabularies](#) > [Terms of rank](#)

Terms of rank

Terms of rank of nobility, etc. used are:


- [Terms of rank in the Iban language](#) →
- [Terms of rank used in France](#) →
- [Terms of rank used in Indonesia](#) →
- [Terms of rank used in the United Kingdom](#) →

Community Vocabularies

Resources > [Community vocabularies](#) > [Terms for collective titles](#)

Terms for collective titles

Draft content

These conventional collective titles are intended for use as values of Work: [preferred title of work](#) .

OPTION

Record an appropriate term from the following list:


- *correspondence*
- *essays*
- *librettos*
- *lyrics*
- *novels*
- *plays*
- *poems*
- *prose works*
- *short stories*
- *speeches*

Community Vocabularies

Draft content

Resources > [Community vocabularies](#) > [Terms for gender](#)

Terms for gender

These terms are intended for use as values of Person: [gender](#) .

OPTION

Record an appropriate term from the following list:

- *female*
- *male*
- *not known*

Community Vocabularies

Draft content

Resources > [Community vocabularies](#) > [Terms for medium of performance](#)

Terms for medium of performance

These terms for instruments, voices, and ensembles are intended for use as values of

Expression: [medium of performance of musical content](#) 

Instruments

Use the following list of terms as a guide:

cello or violoncello

cor anglais or English horn

...

Community Vocabularies: Still to be Determined

- Final terminology
 - “Community vocabularies” is a working term
- How to identify which community uses which content
 - And present that seamlessly to catalogers using the Toolkit
- How to share content
 - For example, if the British Library follows LC-PCC decisions in a particular area
- Interoperability with policy statements, application profiles, and user-created documentation
- Location and display in the Toolkit

Background Information

- [Community vocabularies in RDA Toolkit](#)
 - Paper discussed at the April 2020 Asynchronous RSC Meeting
 - Available from: <http://rda-rsc.org/node/633>
- [String encoding schemes in RDA Toolkit](#)
 - Paper discussed at the January 2020 Asynchronous RSC Meeting
 - Available from: <http://rda-rsc.org/node/587>
 - Summary of RSC discussion and decision available in the January RSC meeting minutes (Agenda item 209): <http://www.rda-rsc.org/sites/all/files/Minutes%20Public%20January%202020%20meeting.pdf>

Questions?

Ask now, or email me later at:
RSCChair@rdatoolkit.org