String Encoding Schemes: Community Choice in RDA

Kathy Glennan
Chair, RDA Steering Committee
Head, Original & Special Collections Cataloging
University of Maryland
What is a string encoding scheme (or SES)?

From the RDA Glossary:

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
    o What information to record
    o In what order
    o 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.
SESs 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
SESs 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)
  - Seuss, Dr.
  - Seuss, Dr., 1904-1991
  - Seuss, Dr., 1904-1991 (duln. f. Theodor Seuss Geisel)
  - Сьюз 1904-1991 доктор
  - Geisel, Theodor Seuss, 1904-1991
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
    o Identify the designated option (Approaches #1 & #2), or
    o Contain the actual instructions (Approach #3)
### Approach #1

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

<table>
<thead>
<tr>
<th>STRING ENCODING SCHEME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record a value that includes, in this order:</td>
</tr>
<tr>
<td>1. a value of Agent: access point for agent</td>
</tr>
<tr>
<td>2. a value of a base access point for a work</td>
</tr>
<tr>
<td>Apply the string punctuation pattern:</td>
</tr>
<tr>
<td>&quot;value 1. value 2&quot;</td>
</tr>
<tr>
<td>Join each of the values with a full stop followed by a space</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STRING ENCODING SCHEME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record a value that includes, in this order:</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>
Approach #2

- Provide information about SESs 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

**OPTION**  
String encoding scheme  
Example

**OPTION**  
String encoding scheme  
Example

**LC-PCC**  
LC/PCC practice: Do not apply.

**LC-PCC**  
LC/PCC practice: Do not apply.

**LC-PCC**  
LC/PCC practice: Apply this option.
Approach #3

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

- 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

- 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

<table>
<thead>
<tr>
<th>GUIDANCE</th>
<th>POLICIES</th>
<th>RESOURCES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Glossary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vocabulary Encoding Schemes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relationship Matrix</td>
</tr>
<tr>
<td>Names in the Arabic alphabet</td>
<td></td>
<td>Additional instructions on names of persons</td>
</tr>
<tr>
<td>Burmese and Karen names</td>
<td></td>
<td>Capitalization</td>
</tr>
<tr>
<td>Chinese names containing a non-Chinese given name</td>
<td></td>
<td>Initial articles</td>
</tr>
<tr>
<td>Icelandic names</td>
<td></td>
<td>Revision History</td>
</tr>
<tr>
<td>Indic names</td>
<td></td>
<td>Community vocabularies</td>
</tr>
<tr>
<td>Indonesian names</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malay names</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roman names</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romanian names containing a patronymic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thai names</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recording surnames that include an article and/or preposition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AACR2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Resources Tab: New Pop-Outs

<table>
<thead>
<tr>
<th>GUIDANCE</th>
<th>POLICIES</th>
<th>RESOURCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Guideline for Capitalization</td>
<td>Capitalization</td>
<td>Glossary</td>
</tr>
<tr>
<td>General guideline for English language capitalization</td>
<td>Initial articles</td>
<td>Vocabulary Encoding Schemes</td>
</tr>
<tr>
<td>Other languages</td>
<td>Revision History</td>
<td>Relationship Matrix</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Additional instructions on names of persons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Community vocabularies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AACR2</td>
</tr>
</tbody>
</table>
## Resources Tab: New Pop-Outs

<table>
<thead>
<tr>
<th>GUIDANCE</th>
<th>POLICIES</th>
<th>RESOURCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial articles listed by language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial articles listed by word or words</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>GUIDANCE</th>
<th>POLICIES</th>
<th>RESOURCES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Glossary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vocabulary Encoding Schemes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relationship Matrix</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Additional instructions on names of persons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Capitalization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Initial articles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Revision History</td>
</tr>
<tr>
<td></td>
<td>Community vocabularies</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AACR2</td>
</tr>
<tr>
<td>Abbreviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Books of the Bible</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terms of rank</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terms for collective titles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terms for gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terms for medium of performance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

• New pop-out for Books of the Bible

Draft content
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

Terms for collective titles

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](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](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](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